Clocking in on autism:

Time perception and temporal aspects of communication in Autism Spectrum Disorders

Anna Maureen Elisabeth Naomi Lambrechts

Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Department of Psychology

City, University of London

December 2018

Table of Contents

List of illustrations

List of tables

Acknowledgments

I would like to express my heartfelt gratitude to the very special people I have been lucky enough to have around me over the course of my PhD:

To my supervisor Dr Sebastian Gaigg for his invaluable advice, his unconditional support, and for being a constant inspiration throughout my PhD. I am grateful for the trust he placed in me and for having been allowed the freedom to choose my own course of research. I could not have wished for a better guide on this journey.

To my second supervisor Dr Kielan Yarrow for his quiet support and humour, his talent at explaining obscure maths and his humility.

To Professor Dermot Bowler, and to my fellow PhD students at the Autism Research Group, for creating such a special environment to work in and for placing caring and humility at the heart of autism research (as well as for all the cakes).

To my family, for inspiring me to pursue an academic path. To my parents who have always encouraged us to do what we love, to Louis for telling me about his own research and inspiring me to follow his footsteps and Carole for opening the doors of a high security neuroscience lab to an enthusiastic teenager, to Adrien and Paul for always believing in me, and to Philippe for the infectious curiosity he has for the world.

To Simon, for his unwavering support over the past few years and his extraordinary patience. Thank you for being my rock through the ups and the downs.

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Funding

The work presented in this thesis was funded by a 3-year PhD bursary from City, University of London.

Abstract

A recent upsurge in the study of time perception in autism spectrum disorders (ASD) has suggested that atypical temporal processing may contribute to both phenotypical and clinically-defining aspects of ASD. The work presented in this thesis aimed to evaluate if atypical temporal processing does impact behaviour in autism and in particular whether it relates to core features of ASD. In the first part, we sought to establish whether time perception was indeed atypical in autism using a psychophysical short durations comparison task, and found that whilst a number of participants performed the task typically, a high proportion of autistic adults experienced difficulties to perform the task above-chance. In the second part, we turned our attention to temporal aspects of speech and gesture coordination in naturalistic productions. Focusing on a first set of data, we provided a progressive characterisation of successive levels of temporal organisation in communication, finding that autistic and non autistic adults differ mainly in terms of the structure of speech and pauses on the one hand, and gesture and holds on the other hand. Specifically, autistic adults showed similar asynchrony between speech and gesture (absolute delay between an event in speech and an event in gesture), but increased coupling (dependency between the speech and gesture signals) and diminished mutual information (repetition of patterns in the different communication channels), compared to their typically developing counterparts. Importantly, we showed that distinct temporal aspects of communication relate to phenotypical aspects (quality of communication ratings) and clinical severity (scores on the ADOS). Using machine learning algorithms, we found mixed evidence that acoustic and kinematic features of communication can predict a diagnosis of autism with high accuracy. However, an identical analysis on a second set of data failed to replicate the group differences at any level of the temporal structure of speech and gesture. We conclude with some recommendations for the most promising directions to explore in future research.

“I'm not strange, weird, off, nor crazy, my reality is just different from yours.”

Lewis Carroll (Alice in Wonderland)

“Omnia, Lucili, aliena sunt, tempus tantum nostrum est; in hujus rei unius fugacis ac lubricae possessionem natura nos misit, ex qua expellit quicumque vult.”

[“Nothing, Lucilius, is ours, except time. We were entrusted by nature with the ownership of this single thing, so fleeting and slippery that anyone who will can oust us from possession.”]

Seneca, Moral letters to Lucilius, Letter I, On saving time

Chapter 1
General introduction

The overarching aim of the work presented in this thesis is to examine temporal cognition in Autism Spectrum Disorder (ASD) and ask whether differences in temporal processing and in the coordination of thoughts and behaviour may contribute to its core clinical and associated symptoms. The motivation for this work stems from reports that individuals on the autism spectrum experience time differently (Boucher, 2001) and have timing and social timing difficulties (Dawn Wimpory, Nicholas, & Nash, 2002) that could contribute to core clinical features (Allman, 2011).

To provide the necessary context for the work that follows, the literature review set out in this chapter will give an overview of autism as we know it, including its current clinical criteria, associated symptoms, neurobiology and the most influential psychological theories that have shaped our knowledge of autism. The next section will then introduce some notions and paradigms from the timing and time perception literature and the main theories of temporal cognition, following which we will report and discuss evidence in relation to temporal cognition in ASD specifically. A final concluding section will draw together the main points to inform the specific hypotheses that will be examined in the subsequent chapters.

Before proceeding with this review, it is useful to explain the choice of terminology used in this thesis. Research shows that in the UK no term is universally endorsed by members of the autism community (people on the spectrum, their families and professionals of the health and social support network for autism) to refer to autism or individuals on the spectrum (Kenny et al., 2015). Whilst we will avoid using terms which are disliked by a majority in any one of these groups, throughout this thesis we will use interchangeably the terms 'autism', 'autism spectrum disorder' and 'ASD' to refer to the clinical condition, and 'person with autism', 'individual on the spectrum' and 'autistic individual' to refer to the people on the spectrum.

1.1 Autism Spectrum Disorder

1.1.1 Clinical criteria of ASD

The American Psychiatric Association's Diagnostic and Statistical Manual, Fifth Edition (DSM-5) defines Autism Spectrum Disorder (ASD) as a neurodevelopmental disorder characterised by persistent difficulties in social interaction and communication as well as the manifestation of repetitive behaviour and restricted interests (American Psychiatric Association, 2013). Social interaction and communication difficulties include diminished reciprocity in social exchanges such as less initiation and fewer efforts to maintain an interaction, less back-and-forth turns in conversations, and reduced sharing of thoughts, emotions and intentions. They are also reflected in poorer verbal and non-verbal communication skills, such as atypical eye contact and gestures and reduced cross-modal integration of communication (i.e., diminished coordination of eye contact, gesture and speech). Finally, individuals on the autism spectrum generally have a poorer understanding of social relationships and make inappropriate or odd attempts to initiate or foster relationships.

The second clinical criterion for a diagnosis, repetitive behaviour and restricted interests, can be manifested in the motor domain with behaviours like rocking or flapping, unusual use of objects such as lining up toys, or in speech through echolalia or the use of stereotyped phrases. It also includes insisting on following strict routines and having disproportionate difficulties when deviating from a planned schedule. Beyond the motor domain, restricted patterns extend to individual interests and preoccupations. For example, individuals on the spectrum often show intense focus on a specific topic, object or activity to an extent that interferes with other behaviours or social demands. Finally, the current diagnostic criteria also acknowledge sensory atypicalities as a core feature of the autism spectrum, encompassing both under- and oversensitivity to sensory stimuli such as certain sounds and lights, aversion or fascination for some tastes and smells and apparent insensitivity to pain (Ben-Sasson et al., 2009).

ASD is a life-long disorder for which symptoms can usually be first identified and reliably diagnosed around 18-24 months (Johnson & Myers, 2007; Steiner, Goldsmith, Snow, & Chawarska, 2012), although the average age for an ASD diagnosis in childhood in the UK is 55 months (Brett, Warnell, McConachie, & Parr, 2016). Symptoms evolve over time and manifest differently over the course of development as the demands on social-communication skills change (Steiner et al., 2012). Early indications of the disorder include diminished eye contact and social smile between 6 to 12 months (Ozonoff et al., 2010), followed around 12 months by lower responsiveness to name and joint attention (Nadig et al., 2007), atypical object exploration and repetitive behaviours (Kim & Lord, 2010; Ozonoff et al., 2008). Later still, atypicalities in language and non-verbal communication become more and more apparent (Eigsti, De Marchena, Schuh, & Kelley, 2011; Landa & Garrett-Mayer, 2006; Presmanes, Walden, Stone, & Yoder, 2007; Yoder, Stone, Walden, & Malesa, 2009) although in around 45-50% of the ASD population language develops relatively typically with only subtle difficulties in language pragmatics.

The prevalence of autism spectrum disorders (ASD) is estimated at around 60 to 110 in 10,000 children with a gender ratio of approximately 4M:1F (Baird et al., 2006; Fernell & Gillberg, 2010; Fombonne, 2009). There is currently no treatment for ASD because the underlying causes remain elusive and because the disorder is extremely heterogeneous in terms of core clinical features and associated co-morbidities. For instance, Charman et al. (2011) show that around 55% of children on the autism spectrum have additional intellectual disabilities and others have shown that around 70% suffer at least one additional comorbid disorder and 41% suffer two or more (Simonoff et al., 2008). Salazar et al. (2015) recently reported even higher rates of co-morbidity, with 90.1% of children aged 4.5-10 years with a diagnosis of ASD also meeting criteria for at least one psychiatric disorder and 51.4% meeting criteria for 3 or more. According to Salazar and colleagues, the most common additional diagnoses are generalised anxiety disorder (66.5%), specific phobias (52.7%) and attention deficit hyperactivity disorder (59.1%) with some gender and age differences in risk factors. This means that the clinical picture, already complex in ASD, can become extremely complicated and disentangling the causes of the core features of the disorder from those of associated co-morbidities remains a considerable challenge. It also means that to support somebody with ASD, families, clinicians and carers need to take into account multiple factors and prioritise the needs of a particular individual.

1.1.2 Origins of ASD

From the earliest clinical reports, it has been pointed out that close relatives of individuals on the spectrum often present autistic-like traits (Asperger, 1944b; Kanner, 1943). Twin and family studies subsequently demonstrated that autism is highly heritable. Some early studies estimated heritability to be near 90% (e.g., Bailey et al., 1995; Steffenburg et al., 1989) whereas more recent reports suggest a more modest contribution of the genetic risk, estimating heritability to be nearer 50% (e.g., Gaugler et al., 2014; Hallmayer et al., 2011; Klei et al., 2012). Regardless of the precise percentages, evidence of a genetic basis for ASD was crucial in debunking previous notions that the parents (and more specifically the mothers) of autistic individuals were responsible for their children's autism (the myth of the 'refrigerator mother'), which had caused considerable distress particularly in the 1960s and 70s.

Since the early indications of a genetic contribution to the aetiology of ASD, a large number of genetic abnormalities have been implicated in the disorder. For instance, a large-scale study conducted by an international consortium, the Autism Genome Project, found that a number of *de novo* copy number variations (CNVs) were associated with ASD, which consist of a change in the number of gene variants that is not inherited from either parent (either by duplication or deletion of a gene copy). The authors noted that these CNVs implicate particular loci and gene families in ASD such as SHANK2, SYNGAP1, DLGAP2 et PTCHD1 that tend to be involved in the regulation of synaptic transmission or intercellular communication, proliferation, projection and motility, and which play a crucial role in the development of neural pathways (Pinto et al., 2010). Other studies (e.g., Gaugler et al., 2014) have identified mainly common variations (the frequent substitution of nucleotides in the genetic code), with *de novo* mutations accounting for only a small percentage (2-3%) of the variance. Few of these genetic variations, whether CNVs or common variation, are specific to ASD and in particular many loci associated with intellectual disability (ID) were also highlighted in ASD. To date, no single locus or variation has been found to be deterministic of ASD and it is generally agreed that genetic abnormalities represent risk factors that lead to the development of the disorder in complex interactions with environmental factors. Identifying the underlying genetic determinants of autism thus remains a crucial challenge. Whilst at the genetic level causal mechanisms remain unclear, at the neurobiological level there is somewhat more consistency in the findings.

1.1.3 Neurobiology of ASD

Autism has been characterised by neurobiological particularities at several levels. At a global structural level, brain growth (usually measured as brain circumference, brain volume or brain weight) has repeatedly been reported as an early difference between typical infants and infants who will later receive a diagnosis of ASD (e.g., Courchesne, 2004; Shen et al., 2013). Although evidence suggests that brain size is not different at birth for those infants who are later diagnosed with ASD, the two first years of life are marked by abnormal brain overgrowth (peaking around 2-4 years of age, see Allely, Gillberg, & Wilson, 2013, for review), which tends to be particularly marked in cortical, cerebellar, and limbic structures. This period of overgrowth is followed by an atypical period of reduced or interrupted growth (Courchesne, 2004), thus leading to relatively typical brain volumes in later life in ASD. It should be noted that evidence for brain overgrowth in autism is not unanimous and even in studies that report atypical brain growth trajectory the individuals with abnormal brain size are often a subset of the group and not the full sample (see Redcay & Courchesne, 2005, for a meta-analysis). Courchesne and colleagues (2007) proposed the influential idea that brain overgrowth is due to an excessive number of neurons during early infancy which leads to aberrant patterning and networking within the brain. In particular, they hypothesized that this excess leads to overabundant local, short-distance connections and fewer long-distance connections between more distal areas of the brain, which could directly underlie some of the cognitive and behavioural symptoms of autism.

This account is supported by evidence of atypical connectivity in autism, although here again evidence is mixed. Functional connectivity (the correlation between different brain region activities) has been shown to be reduced in autism during various cognitive tasks (e.g., Damarla et al., 2010; Just, Cherkassky, Keller, Kana, & Minshew, 2004; Solomon et al., 2009; see Maximo, Cadena, & Kana, 2014, for review) as well as in the absence of experimental task (so-called ‘resting-state’; e.g., Abrams et al., 2013; Di Martino et al., 2013; Lai et al., 2010; see Maximo et al., 2014, for review). The weaker connectivity reported in these studies concerns primarily connections between the prefrontal cortex and more posterior regions of the brain such as the medial prefrontal cortex, the temporoparietal junction, the superior temporal sulcus and the fusiform gyrus (Schipul, Williams, Keller, Minshew, & Just, 2012), which altogether could jeopardise some of the cognitive functions implicated in autism such as planning, memory, face processing and language. Other networks showing weaker functional connectivity in autism include the amygdala, temporal and frontal regions (Monk et al., 2010), the primary and supplementary motor areas, anterior cerebellum and thalamus (Mostofsky et al., 2009) as well as the visual cortex, thalamus and cerebellum (Villalobos, Mizuno, Dahl, Kemmotsu, & Müller, 2005), although these accounts typically yield less robust data and need replicating. Although a majority of connectivity studies in autism focus on underconnectivity, another set of studies has reported overconnectivity in some areas (Maximo, Cadena, & Kana, 2014, for review). For instance, there is evidence for overconnectivity in the extrastriate cortex, frontal and temporal regions, amygdala and parahippocampal gyri (Murphy, Foss-Feig, Kenworthy, Gaillard, & Vaidya, 2012; Noonan, Haist, & Müller, 2009; Shih et al., 2011, 2010; Uddin et al., 2013; Welchew et al., 2005). Other regions with reported overconnectivity include the posterior cingulate cortex (Monk et al., 2009) and the temporo-thalamic regions (Nair, Treiber, Shukla, Shih, & Müller, 2013). Overconnectivity has generally been interpreted in terms of overspecialised functions in autism rather than increased efficiency in these regions. Shih and colleagues (2011) hypothesise that overconnectivity is due to diminished pruning during early development, which itself has been linked to brain overgrowth in the first years of life. In support of these mixed findings of under- and overconnectivity, several studies find evidence for a mixed pattern of atypical connectivity (which Maximo et al., 2014, refer to as *disrupted connectivity*). Importantly, atypical connectivity correlates with sensory and socio-communicative symptoms (Abbott et al., 2016). Tentative theories propose that due to excessive neural generation and diminished pruning in the first two or three years of life, neural networks become overly connected locally, reducing the communication with more distant brain areas. This could result in a pattern of specialised areas of expertise in low-level perceptual processes (i.e., local over-connectivity) with co-occurring difficulties in global cognitive functions that would rely on longer range neural connections.

At a more cellular level, autopsy studies in the 1980-90s, and from the mid-80s imaging studies, revealed reduced numbers of inhibitory Purkinje cells in the cerebellum in autism (e.g., Bauman & Kemper, 1985; Bauman & Kemper, 1990; Bauman & Kemper, 1986; Courchesne et al., 1994; Courchesne, Hesselink, Jernigan, & Yeung-Courchesne, 1987; Hashimoto et al., 1995). Similarly, the brainstem has been shown to be reduced in size and in the number of cells (e.g., Hashimoto et al., 1995; Rodier, Ingram, Tisdale, Nelson, & Romano, 1996). Bauman and Kemper (1985) also reported unusual neuron-packing in the medial temporal lobe including the hippocampus, entorhinal cortex and amygdala. Evidence also suggests that cortical organisation of neurons in minicolumns is atypical in autism. For instance, Buxhoeveden et al. (2006) and Casanova et al. (2006) reported narrower minicolumns in the frontal and temporal cortex, likely indicating an abnormal increase in the number of neurons and column units during cortical neurogenesis.

Neurochemical systems have also been under investigation as possible correlates of autism. The serotoninergic system was the first to be identified as atypical in autism, and even before serotonin was identified as a neurotransmitter, hyperserotonemia was reported as a potential biomarker for autism (Schain & Freedman, 1961) and remains detectable in between a quarter and a third of the autistic population (Anney, 2013; Gabriele, Sacco, & Persico, 2014). Serotonin is involved in early neurodevelopment in processes such as cell proliferation, migration and differentiation, particularly in sensory regions, and may therefore be implicated in the atypical connectivity patterns seen in ASD. Atypical serotonin function may also play a role in some of the co-morbidities often found in ASD, such as depression, anxiety and OCD symptoms (McCracken et al., 2002; McDougle et al., 1996), as well as gastro-intestinal issues. Other neurotransmitter systems that have been implicated in ASD include the dopaminergic system (Gadow, Pinsonneault, Perlman, & Sadee, 2014; Kriete & Noelle, 2015) and the excitatory/inhibitory balance between the glutamatergic and GABAergic systems (Brondino et al., 2016). The dopaminergic system is commonly associated with the reward-motivation system, and abnormalities may be a source of the 'social-motivation' impairments that some have argued lie at the root of the developmental trajectory of ASD (e.g., Chevallier et al., 2012). Particularly interesting in the context of this thesis is evidence which implicates dopamine in timing and time perception (Cheng, Tipples, Narayanan, & Meck, 2016; Van Rijn, Gu, & Meck, 2014).

One of the current challenges of ASD is to reconcile knowledge of the disorder at different levels: genetic, molecular, neurobiological, psychological and behavioural. Despite the accumulation of evidence and theories in each domain, bridges between the different scales are hard to establish and therefore a full understanding of the spectrum is not yet possible. Because this work focuses on cognitive aspects of ASD, the next section will introduce some of the most influential psychological theories of ASD to date, which will provide a framework for discussing and interpreting the results of the empirical work presented in subsequent chapters.

1.1.4 Influential cognitive theories of ASD

1.1.4.1 Theory of Mind

An early theory, which has attracted great interest for several decades, is the notion that individuals with autism have an impaired 'Theory of Mind (ToM)', in other words, that they show difficulties thinking about their own and other people's thoughts and mental states. The idea stemmed from a study by Wimmer and Perner (1983) who set up a false-belief (FB) task for children of different ages in which a protagonist placed an object in location A. Whilst the protagonist is away, an antagonist moves the object from A to B. Children are asked where they think the protagonist is going to look first for the object when they return. Wimmer and Perner's results showed that until the age of 4-6, children fail to predict that the protagonist, on his return, will look for the object in location A, despite remembering the initial location of the object correctly. Instead, they predict that the protagonist will look in the object's current location B, showing that their own knowledge of the world interferes with understanding someone else's. Based on the earlier work by Premack and Woodruff (1978), who argued that the ability to anticipate the behaviour of others on the basis of false beliefs is proof of a 'theory of mind', Wimmer and Perner concluded that Theory of Mind (ToM) emerges around 4-6 years of age in humans. Using the same paradigm (known from then on as the "Sally-Ann" task), Baron-Cohen, Leslie, and Frith (1985) showed that this ability is substantially compromised in ASD. Whilst typically developing (TD) and Down

Syndrome (DS) children of similar or lower mental age passed the test with 85% success rate, 80% of the autistic children failed to answer the belief question correctly despite knowing both the initial and final location of the marble (the object displaced). This study and many following it propose that individuals on the spectrum do not develop a full theory of mind, and in particular struggle to represent other people's mental states such as knowledge, beliefs and intentions (also referred to as 'mentalising', see Frith & Happé, 1994). Consequently, the process of learning through others is compromised in autism which impacts development throughout life. Early on, joint attention cues are missed because the autistic child might not understand the intention behind complex posture, eye gaze, pointing or verbal behaviour aimed to direct their attention towards a common object (Baird et al., 2000; Boucher, Lewis, & Collis, 2000; Dawson, Toth, et al., 2004a; Dawson, Webb, Carver, Panagiotides, & McPartland, 2004; Rutherford & Rogers, 2003; Schultz, 2005). Language ambiguity is greater because autistic individuals cannot complete the information based on their understanding of the other's knowledge, beliefs and intentions (Happé, 1997; Malle, 2002; Sperber & Wilson, 2002). Failure to understand the world of others would encourage autistic children to turn to the predictable, more straightforward world of objects.

The ToM theory of autism is an elegant, unifying theory that has the merit of proposing a single cognitive function as the critical bottleneck of autistic development: multiple genetic and neuropathological atypicalities converge into a single cognitive deficit in the '*mentalising*' function, which in turn leads to a developmental cascade of varied impairments in behaviour. However, whilst this theory has driven useful research into autism and indeed typical development, and allowed useful predictions to be made, it also has its limitations. Frith and Happé (1994) argue that although ToM accounts well for clinically-defining aspects of autism, it does less well in explaining other features of the disorder such as the presence of stereotypies and a desire for sameness, as well as the "spikey" profile of abilities in autism, which includes strengths in low-level perceptual processing. In addition, it does not explain why 20% of the children tested, even in the first report of impaired ToM, successfully solve the false-belief task. Another issue with the ToM theory of autism is that, in the original comparison sample (and subsequent studies), a proportion of the non-autistic children also fail to pass the false belief task. Later evidence showed that individuals with hearing or visual impairment show delayed ability to pass false-belief tasks (Minter, Hobson, & Bishop, 1998; Russell, Hosie, Gray, Scott, & Hunter, 1998), whilst individuals with intellectual disability persistently fail false-belief tasks (Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). If ToM impairment is a cause of autism, then these individuals should be showing autistic-like communication and social interaction impairment, which is only the case transiently in development but not pervasive as in the case of ASD (Boucher, 2012). Finally, Boucher (2012) picks up the argument that ToM deficits as measured by false belief tasks are not a tenable account because reliable signs of autism (appearing during the first 3 years of life; see Johnson et al., 2007; Steiner, Goldsmith, Snow, & Chawarska, 2012) are evident long before the ability to solve such tasks would typically develop (around 4 years of age). In addition, false belief tasks make substantial demands on verbal ability and executive functions, particularly response selection and inhibition (Leslie, Friedman, & German, 2004), leaving the possibility that results are driven by verbal and cognitive skills other than the ability to understand another person's state of mind. For instance, in order to "pass" a false belief task, participants not only have to access another agent's belief, but also suppress the response corresponding to their own knowledge of the world.

However, a growing body of evidence indicates that another implicit, automatic ToM system can be observed much earlier in development (Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005; Schneider, Slaughter, & Dux, 2017; Sodian & Thoermer, 2008; Träuble, Marinović, & Pauen, 2010). Although the "mentalising" quality of infants' state of mind (Burge, 2018) and the onset of such a system (Schneider et al., 2017) remain controversial, deficits in early implicit ToM skills are a promising candidate to account for early differences in social communication in autism. Accordingly, more recent accounts suggest that whilst some older or more able autistic individuals 'hack out' the solution to explicit false belief tasks (Happé, 1995), their automatic, implicit sense of other people's state of mind remains compromised. In support of this idea, two studies (Schneider, Slaughter, Bayliss, & Dux, 2013; Senju, Southgate, White, & Frith, 2009) found that autistic adults who pass classic false belief tasks failed to show spontaneous, anticipatory eye movements based on false belief attribution.

The possibility of a dual implicit/explicit ToM system has revived the debate about the central role of a ToM deficit in autism. Further exploration of implicit ToM skills in infants at risk or with a diagnosis of ASD and their impact on social skills is needed to put this revised ToM theory of autism to the test.

1.1.4.2 Social Motivation Theory

Another theory which puts the social world at the centre of the aetiology of ASD is the Social Motivation Theory. Although the idea that autistic individuals "relate themselves" less to other people from a very early age dates back to Kanner's first description (Kanner, 1943), the theory developed mainly from the literature on atypical face processing in autism. Reports of atypical scene and face exploration (Dawson, Webb, & McPartland, 2005; Hobson, Ouston, & Lee, 1988) and atypical activation of the fusiform face area and amygdala (e.g., Schultz, 2005; Schultz et al., 2003) in autism lead authors to propose that reduced social interest at an early age is instrumental in the development of autism (Grelotti, Gauthier, & Schultz, 2002). Specifically, infants who were later diagnosed with ASD were shown to spend less time exploring social stimuli in complex scenes, and when exploring faces they spent less time on the eyes area (Dawson et al., 2005). The assumption made is that in typical development, social stimuli (faces, biological movements) are inherently rewarding. The argument that follows is that failure to attend typically to social cues such as the carer's face at an early age means that the infant has less opportunities for scaffolded learning, and that at the functional level neural networks do not specialise in processing social stimuli (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Gaigg, 2012). For instance, reduced exposure to social stimuli are predicted to result in poorer emotion recognition. A strength of the social motivation theory compared to ToM is that it predicts a diverging development path starting early on, but with cascading consequences as the individual goes through successive developmental stages. A recent review (Bottini, 2018) however reported that overall empirical evidence was mixed regarding social motivation in autism and that the jury was still out regarding the question whether social stimuli really are less rewarding in autism than in typical development.

1.1.4.3 Executive Functions theory

Whilst ToM and the social motivation theory provide useful accounts for the core social-communication difficulties of ASD (with the limitations mentioned above), they both fail to explain non-social clinical aspects of the disorder such as unusual sensory experiences as well as repetitive behaviours and restricted interests in individuals on the spectrum. Addressing this caveat, a third, high-level cognitive theory of autism promised to account not only for social and communication difficulties but also for the repetitive and stereotyped behaviours that characterise ASD. Ozonoff, Pennington, and Rogers (1991) were the first to formally propose that a deficit in executive functions could be a primary cause for autism (hereafter EF theory). Primarily based in the prefrontal cortex (PFC), executive functions (EF) encompass the cognitive functions necessary to engage and maintain flexible goal-directed action, including attention, response selection, inhibition, planning and working memory. Difficulties in regulating attention, inhibiting behaviours and/or planning effectively could account for the repetitive behaviours, disproportionate interests for a topic or objects, and reluctance to deviate from a routine observed in autism. Rumsey (1985) and Rumsey and Hamburger (1988) first reported that individuals on the autism spectrum performed worse than comparison groups on the Wisconsin Card Sorting Test (WCST, Grant & Berg, 1948), which measures cognitive flexibility in learning and updating rules. In particular, autistic individuals were found to persevere with the first identified rule despite receiving negative feed-back. Over the following years the performance pattern of individuals on the spectrum on EF tasks has been refined, with some tasks being less affected, such as the Stroop task (Stroop, 1935), whilst others show clear difficulties (e.g., the WCST), and others yet reveal difficulties that appear only when task complexity increases or in combination with other risk factors (i.e., ADHD, learning disability, frontal lobe damage etc., see Hill and Bird, 2006). Of the various psychological theories of autism, the EF account is perhaps the easiest to relate to neurobiological evidence. Disrupted connectivity between the PFC and more posterior regions such as the MTL should impact the integrity of executive functions and the way they act as a top-down modulator of other functions. At the hormonal level, Kriete and Noelle (2015) offered an explanation of how dopamine dysregulation can explain the mixed pattern of preserved and impaired EF performance in autism.

Despite its strengths, however, the EF theory of autism needs to be better defined to help us understand and predict autism further: executive functioning encompass a wide range of processes, and it remains unclear to what extend each of them is compromised in autism, and how they relate to one another. In addition, EF difficulties are not specific to autism but can be observed in other disorders, some of which have a high co-morbidity with autism such as ADHD (Craig et al., 2016; Geurts et al., 2004), which makes the clinical picture more complex. Moreover, the developmental trajectory of EF in autism and how it relates to the core symptoms of the disorder are not well specified (O'Hearn, Lakusta, Schroer, Minshew, & Luna, 2011).

1.1.4.4 Weak Central Coherence and Enhanced Perceptual Functioning theory

Around the same time as the EF theory was proposed, an alternative account was put forward as a candidate for the underlying cognitive basis of autism. Frith (1989) and Frith and Happé (1994), revised in Happé (1999) and Happé and Frith (2006), propose that autistic cognition is more parsimoniously explained by a lower-level, mechanistic difference in cognitive processing. Typical individuals prioritise gestalt perception (seeing the big picture, e.g., the gist of a story, the animal drawn on a page), which allows them to reduce the amount of information extracted whilst retaining relevant and meaningful aspects of an event. Frith and Happé propose that autistic individuals present a reversed processing bias towards details (e.g., the exact wording in a story, the width of a line in a drawing), resulting in apparent difficulties with the integration of small items of information into a higher-order meaningful whole, or "central coherence". This "weak central coherence" (hereafter WCC) leads them to focus on the details whilst being less aware of the bigger picture. Strong support in favour of the theory first came from the prediction that individuals on the spectrum should perform better than TD individuals in tasks that require focusing on details and ignoring the bigger picture. Shah & Frith (1983) tested autistic, intellectually disabled and TD children on the Children's Embedded Figure Test (Witkin, Oltman, Raskin, & Karp, 1971). The task consisted of having to find a simple shape within a complex figure. The authors found that autistic children performed significantly better than both other groups (who performed similarly) on the test. The authors interpreted this result as indicating that the autistic children's bias to focus on detail and ignore the bigger picture worked to their advantage, and the presence of the complex meaningful picture did not interfere with identifying the smaller, simpler shape. Further support for the WCC theory subsequently came from the observation that individuals with ASD showed (in appearance) an inconsistent pattern of cognitive strengths and weaknesses such as their superior performance on the block design task in the Wechsler Intelligence scales (Caron, Mottron, Berthiaume, & Dawson, 2006; Happé, 1994; Shah & Frith, 1993), or intact semantic memory (memory for decontextualized concepts) compared to poor episodic memory (remembering an event in context, Bowler, Gardiner, & Gaigg, 2007; Bowler, Gardiner, Grice, & Saavalainen, 2000; Bowler, Poirier, Martin, & Gaigg, 2016). WCC has also been put forward as an explanation for exceptional skills or "savant" abilities that appear in the autistic population more frequently than in the TD population (Treffert, 1999, 2009) such as instantly identifying the pitch of any sound (Heaton, Pring, & Hermelin, 1999; Heaton & Hermelin, 1998), drawing a scene in exquisite detail from memory (such as autistic artist Stephen Wiltshire, Hermelin, 2001), or quickly performing complex calendrical calculations (Boddaert et al., 2005). Frith and Happé (1994) argue that WCC is a better candidate than ToM deficit as a single-feature deficit for autism since the small percentage of autistic individuals who pass the false-belief task still reveal an advantage for local-processing task and a disadvantage for global-processing tasks in various contexts. However in the latest version of the WCC account (Happé & Frith, 2006) the authors concede that WCC may not be the underlying cause of the social difficulties observed in ASD. They propose that WCC and ToM may represent 'fractionable' causes of autism that can either independently or in combination lead to a diagnosis of ASD (the "fractionable triad", Happé & Ronald, 2008; Happé, Ronald, & Plomin, 2006).

An alternative formulation of the atypical balance between lower level perceptual processing and higher level cognitive processing in ASD was proposed by Mottron and colleagues (Mottron & Burack, 2001; Mottron, Dawson, Soulières, Hubert, & Burack, 2006) in the form of the Enhanced Perceptual Functioning (EPF) theory. EPF postulates that default autistic cognition is dominated by primary sensory perception in a bottom-up fashion, at the cost of top-down prediction-based processes. Whereas WCC proposes that gestalt perception is impaired in autism, EPF argues that gestalt perception only *appears* impaired because of enhanced perceptual functioning which prioritises low-level processing. This processing bias leads to a "purer" perception of the world (less influenced by prior knowledge) with the trade-off that it is more costly and prone to errors (less assumptions about the world means that each situation is assessed based on a "single measurement") (see also Pellicano and Burr, 2012b). EPF also stipulates that performance is highest in low-level sensory perception and lowest when complexity increases, for instance when processing a stimulus in movement as opposed to a stationary stimulus. EPF accounts for the spikey profile of autistic individuals in general ability tests (with maximum performance in tasks where local sensory processing is an advantage), special abilities and, in their more extreme form, savant syndrome. Compared to its close neighbour WCC, EPF is more readily relatable to a neurobiological level of explanation: overconnectivity in local networks and underconnectivity between more distant brain areas translate easily into predicting a preferred sensory-oriented, bottom-up processing style (Just, Keller, Malave, Kana, & Varma, 2012).

1.1.4.5 Bayesian theories

Based on the same observation that sensory processing is atypical and still widely unexplained in ASD, Pellicano and Burr (2012) proposed that it is not primary sensory processing which is different in autism but the interpretation of sensory input. The need for another theory, they argued, comes from the need to explain the entire range of sensory atypicalities, including sensory hyper- and hyposensitivity and sensory-seeking behaviours. Specifically, they suggested that a Bayesian framework could be useful to consider. In this framework, the percept of a single stimulus results from the combination of the sensory information issued from the stimulus itself (providing a "likelihood" distribution of the stimulus), combined with prior knowledge and expectations (the "prior") to result in a "posterior probability distribution" the peak of which determines the final percept. Pellicano and Burr hypothesise that, in the case of autism, priors are broader, or less constraining ("hypo-priors"), in other words a percept relies more strongly on the stimulus sensory properties than on previous knowledge or expectations. The combination of priors and likelihood is a trade-off between accuracy (how close a percept is to the physical reality) and reliability (minimising errors by taking into account knowledge of the world).

The notion of hypo-priors in ASD suggests that autistic individuals have less specific models of the world compared to TD individuals (a broader prior distribution). As a result, the product of the prior and of the likelihood distributions produces a greater range of solutions in the posterior distribution, which leads to several predictions. First, accuracy should be higher in tasks where previous experience or expectations are misleading, for instance in the case of sensory illusions and detail-oriented search. Specifically, some illusions occur when the stimulus presented is unlikely (e.g., the "pac-men" shapes in the Kanizsa triangle), in which case, following Bayesian principles, the prediction is "corrected" with regards to more likely states of the world (the presence of a triangle on top of regular circles) towards a more probable outcome. In line with the hypo-prior account, susceptibility to the Kanizsa triangle and similar illusions (e.g., Ebbinghaus and Poggendorff illusions in the visual modality and the Shepard tone illusion in the auditory modality) is diminished in ASD (Happé, 1996; Mitchell & Ropar, 2004; Shah & Frith, 1983; Stevenson et al., 2014) whilst autistic individuals show superior performance in tasks such as pattern discrimination (Plaisted, O'Riordan, & Baron-Cohen, 1998) and generally detail-focused tasks (Happé & Frith, 2006). A second prediction is that hypo-priors are a hindrance in ambiguous situations, for example for 'noisy' stimuli, in which prior knowledge of plausible states of the world helps disambiguate the sensory input. Again, this is consistent with reports that autistic individuals experience diminished perception of speech in background noise, compared to non autistic individuals (Alcántara, Cope, Cope, & Weisblatt, 2012; Milne et al., 2002). Finally, hypo-priors could explain self-reports of overwhelming sensory stimulation (Grandin, 1992; Grandin & Scariano, 1986), the feeling that each situation is unique, and difficulties in categorising and generalising across situations (Klinger & Dawson, 1995). Although in its initial proposal the hypo-prior theory attempted to account only for non-social aspects of ASD, it has the potential to extend to all aspects of the ASD phenotype (social and non social) and offer a unified neuropsychological account of autism (Brock, 2012; Pellicano & Burr, 2012a). Using the same Bayesian framework, a couple of years later, Van de Cruys et al. (2014) proposed an alternative account for autism. This proposal is based on the predictive-coding theory, in other words the idea that cognition relies on prediction error management. When an event occurs, it is compared to its prediction, resulting in a measurement of prediction error. The model of the world is adjusted accordingly by correcting the prediction for future similar events. Depending on the kind of event, a single instance might, or might not be enough to adjust the model of the world. In addition, if the error is small, it is not always necessary to adjust the model. Van de Cruys and colleagues propose that autistic individuals have an overly inflexible error management system (HIPPEA: High, Inflexible Precision of Prediction Errors in Autism): when a single event produces even a small prediction error, the prediction for future events is adjusted to match that single occurrence precisely, instead of allowing for some variance in the prediction. This system doesn't allow for even a small deviation from the prediction, resulting in judging most events as being in violation with an individual's model of the world. The authors argue that this could affect exploration and learning early in development and explain abnormal sensory sensitivity, and repetitive behaviour in autism, but also difficulties with social interactions and complexities.

1.1.4.6 Summary

Taken together, these influential neuropsychological models of autism have shaped the image of autism and the direction of research into autism over the past few decades. Until a few years ago, these models could generally be categorised into "social" (ToM, social motivation theory) and "non social" (WCC, EPF, EF) models, depending on what aspects of the multifaceted ASD phenotype the theories primarily targeted to explain. Historically, this theoretical divide between accounts of social and non social aspects of the disorder were driven by the seemingly irreconcilable discrepancies between domains of difficulties such as social interaction and communication and domains of either intact or even enhanced performance such as "rote" memory in autism (Frith, 2012). Despite the presumption that various aspects of autism should be accounted for by a common underlying causal factor, the disorder as a whole was most often described by combining multiple theories together (typically, the juxtaposition of social and non social theories), an approach that culminated in Happé's "fractionable triad" (Happé & Ronald, 2008). The recent Bayesian models (hypo-priors, HIPPEA) open a new perspective by postulating that the core difference of autism lies in a more domain-general mechanism, an approach which has yet to account for the fact that some hallmarks of autism are specific to autism (e.g., difficulties with social interaction and communication, differences in sensory processing) whereas others are shared (at least in part) with other disorders (e.g., deficits in executive functions).

Despite the failure to identify a single cause for autism, however, a host of studies have found relationships between different facets of autism that include both social and non social aspects. For instance, performance on EF and central coherence tasks in able 4-7 years old autistic children have been found to predict later proficiency in ToM tasks (Pellicano, 2010), which is perhaps unsurprising giving the reliance of false belief tasks on EF skills. In contrast, a more recent evaluation of the relationship between EF, ToM and ASD symptomatology in adolescents with a wide range of profiles and abilities (Jones et al., 2018) reported that ToM directly predicts not only social communication skills but also restricted and repetitive behaviours as reported by parents, whereas EF do not (except, indirectly, through ToM). First, these results suggest that processes that were previously thought of as domain-specific might not be so compartmentalised. Second, they are a reminder that autism is a developmental disorder, and that interdependencies between social and non social, domain-general and domain-specific processes are likely to be dynamic over time.

1.2 Timing and time perception

The second pillar of this thesis is the literature on time perception and timing. The use of temporal information is ubiquitous in neural and behavioural activities – integrating sensory information across modalities, preparing and executing an action, deciding on the right moment to speak or act, remembering or anticipating events and co-ordinating behaviour *intra-* and *inter*personally all depend on the accurate timing of the various processes involved. Humans and other animals are remarkably sensitive to the timing of events over a wide span of timescales. Whether we time the coordination of muscles to produce a smile or extend an arm, or we come back to the stove in time to prevent the milk from boiling over, we work with a variety of timescales that are crucial for navigating the world efficiently. Figure 1.1 (reproduced with permission from Buhusi & Meck, 2005) helpfully illustrates the impressive span of durations that humans and other animals are sensitive to along with the processes they tap into.

Although the cognitive and neural mechanisms of time perception are still largely unknown, it is generally accepted that distinct mechanisms underlie the processing of different timescales. Timing durations of a few hundred milliseconds to a few seconds (referred to as *interval timing*) involves a distributed neural network, which feeds temporal information into motor behaviour and cognition. Processing longer durations falls under the domain of memory and its dedicated neural substrates (Pouthas & Perbal, 2004). In this thesis, the focus will be on interval timing, because the central question concerns the relation between temporal processing and communication difficulties in autism. Durations which are crucial for everyday communication range from the millisecond-tuned coordination of voice and eye gaze to the split-second-long interval between turn-taking and the few seconds during which we make a communicative gesture, all of which come under the umbrella of interval timing. Before examining what is known about time processing in the context of ASD, the following sections will briefly provide an overview of the types of experimental paradigms that have been used to examine peoples' ability to perceive and produce timed events and of the theories that have been derived from them.

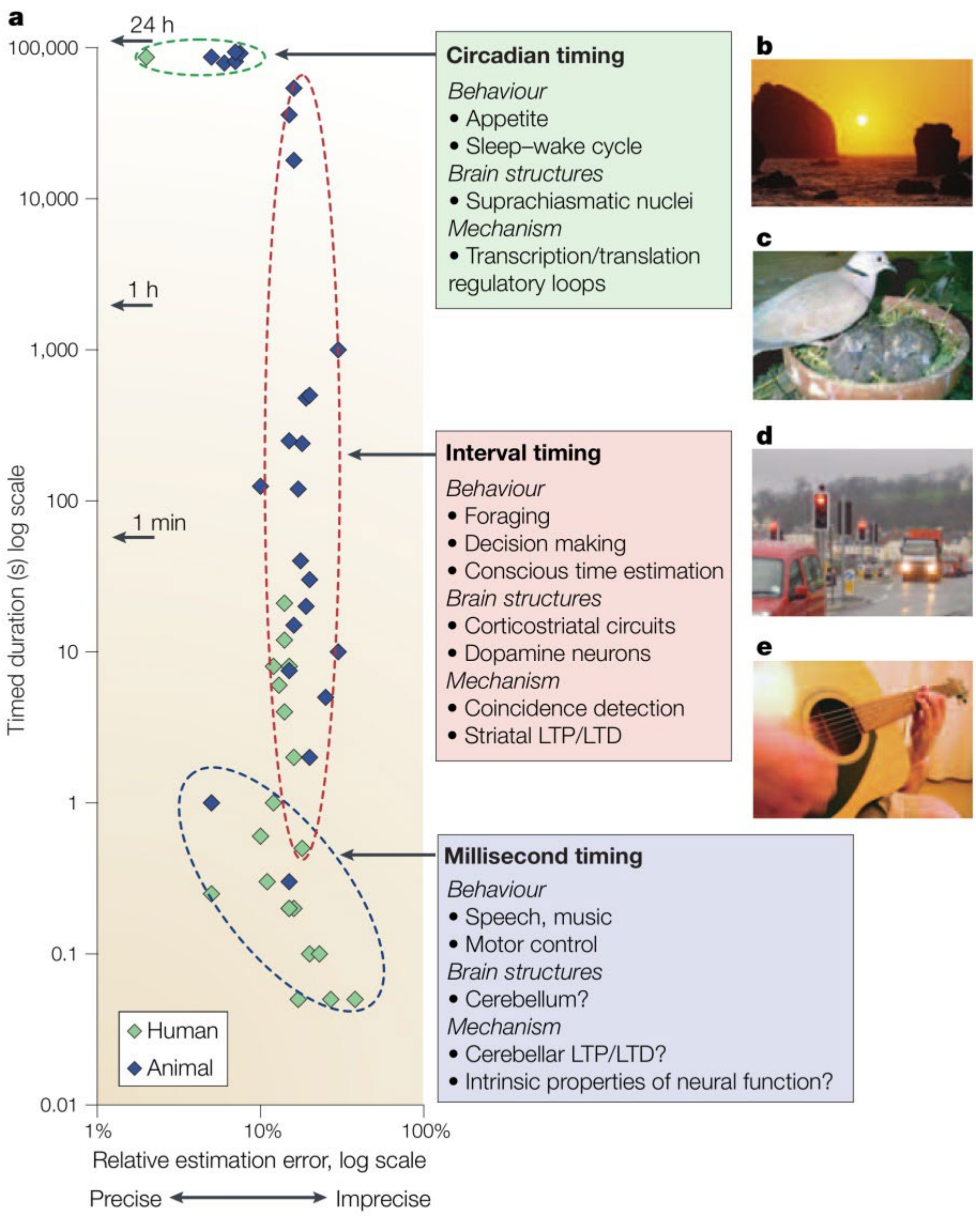

Figure 1.1 The different timescales than human and non humans are able to measure, spanning from milliseconds to days.
Illustration reproduced with permission from Buhusi & Meck (2005). In this work we are concerned with the bottom part of the scale, ranging between a few dozens of milliseconds and a few seconds. (a) A compilation (representative but non exhaustive) of data from various studies indicating the precision of humans and other animals in various timing tasks. (b-e) Circadian rhythms, which are most recognisable in nature (b), but interval and millisecond timing also guide fundamental animal behaviours. For example, although female ring doves use circadian-timing strategies to coordinate egg incubation, males use interval-timing strategies (c) Interval timing is involved in decision making' (d), and millisecond timing is central to the playing of music (e). LTD, long-term depression; LTP, long-term potentiation.

1.2.1 Main paradigms

1.2.1.1 Estimating duration

Several classic paradigms directly assess our ability to estimate the length of an event in time, in other words its duration. In a **temporal generalisation** paradigm

(Church & Gibbon, 1982; Wearden, 1992), the participants are first trained to memorise a standard time interval. Subsequently, they are presented with a variety of intervals and have to decide, for each of them, whether it lasted the same or a different time compared to the standard interval. The expected response profile is curve peaking for the target interval (most intervals equal in duration to the target are judged to be the “same” as the target). The width of the peak indicates how conservative participants are in their judgements (how widely they “accept” a stimulus to be equal to the standard) and the asymmetry of the curve shows biases in judgement. Typically, the response profile shows an asymmetry towards longer durations: for an equal distance to the target interval, a longer interval is judged similar to the target more often that a shorter interval (for instance a 1.2s interval is estimated to be the “same” as 1s more often than an 0.8s interval). The **temporal bisection** paradigm (Allan & Gibbon, 1991) is similar in structure, but participants initially learn two standard intervals: a short and a long interval. Subsequently, they are presented with a variety of intervals and have to classify them as closer to the short or to the long standard interval. A classic analysis of the results is to report a measure of accuracy, the bisection point, which reflects the duration which participants are equally likely to perceive to be similar to the short and long standards, as well as a measure of precision, usually the Weber ratio, which reflects how steep the shift is between a ‘short’ or ‘long’ duration judgement (see Figure 2). A main criticism of these two paradigms is that participants have to memorise the standard comparison intervals, which has been argued to put strain on the cognitive load necessary to perform the task. In addition, the memory for the standard intervals is constantly updated by the presentation of test intervals, which can lead to a shift in performance across trials that reflects learning processes rather than processes of timing. This memory updating effect is a limitation in most studies, although a handful of authors have used it as an experimental manipulation (Filippopoulos, Hallworth, Lee, & Wearden, 2013; Grondin, 2005).

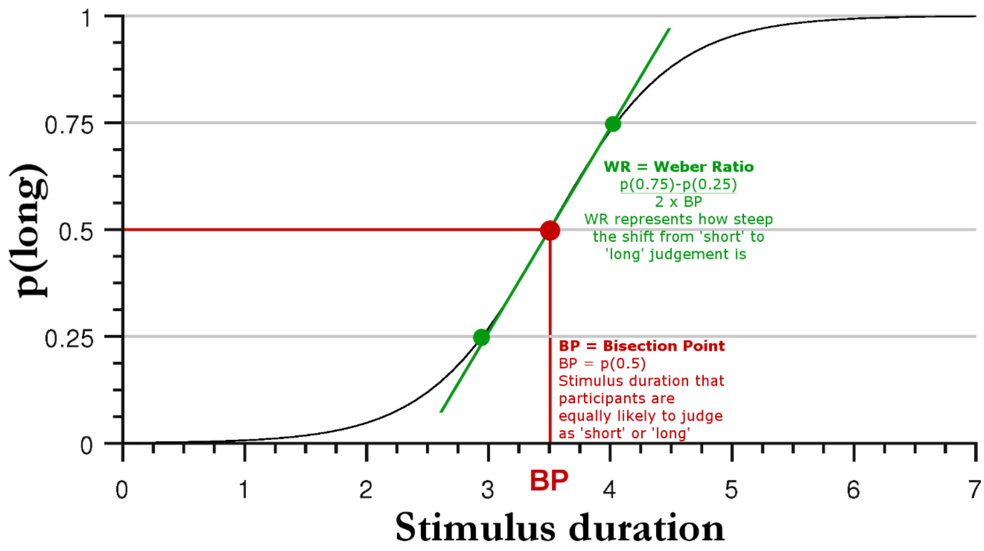

Figure 1.2 Typical psychometric response profile in a bisection task
in which participants have to decide whether the duration of the comparison stimulus (plotted on the x-axis) is closer to the duration of a “short” or a “long” standard stimulus duration (proportion of 'long' responses or p(long), plotted on the y-axis). When the stimulus duration is short, responses show a near-zero proportion of “long” responses. When stimulus duration is long, responses show a near-one proportion of “long” responses. The Bisection Point (BP) is the stimulus duration for which participants are just as likely to give a “short” than a “long” responses (p(long)=0.5) and is an index of accuracy. The Weber Ratio (WR) is computed as the Difference Limen ((p(0.75)-p(0-25))/2) normalised by the Bisection Point and is an index of precision.

The **temporal comparison** paradigm (also called **temporal discrimination** paradigm) (Rammsayer, 1999; Wearden & Lejeune, 2008) tries to address the memory load issue of the generalisation and bisection paradigms by presenting pairs of temporal intervals on each trial and asking participants to decide which of the two intervals lasted longer. One of the intervals is actually a standard interval, which remains constant across trials, therefore allowing for similar analyses to those applied to temporal bisection tasks, providing measures of accuracy and precision (i.e., the bisection point and Weber ratio). A disadvantage of this paradigm is that the task lasts longer, in a setting where the number of trials and the risk of fatigue are usually already high. Another limitation is that response behaviour varies as a function of whether the comparison interval is presented first or second, demonstrating what is known as the time-order error or type B error (Allan, 1977; Ellinghaus, Ulrich, & Bausenhart, 2018; Hellström, 1985).

The three paradigms described above address perceptual aspects of temporal processing, whereas other tasks probe the ability to produce timed behaviours. For example, the **temporal production** task (Fortin & Rousseau, 1987; Wearden &

McShane, 1988) requires participants to produce certain intervals of time (e.g., 1.5 sec) by starting and stopping (or simply stopping) a signal (e.g., a tone). A variant of this task is the **temporal reproduction** paradigm (Kowalski, 1943; Richards, 1964; Woodrow, 1930), in which the participant is first presented with an interval and is then asked to reproduce or copy it by stopping, or starting and stopping a signal. In both paradigms the analyses primarily focus on measures of accuracy. The advantage of reproduction rather than production tasks is that the instructions do not rely on a metric of time (e.g., seconds, minutes etc), although this comes at the cost of making the task longer.

A classic manipulation in the time perception and production literature is to vary the mode of temporal processing. Most experimental designs use **prospective timing** (Hicks, Miller, Gaes, & Beirman, 1977), in which the participant is explicitly instructed to monitor the duration of an upcoming event (experienced duration), which allows for the collection of responses across multiple trials. More resource-intensive studies have addressed the question of **retrospective timing**, in which the participant is prompted to estimate the duration of an event only *after* it has taken place (remembered duration), which can usually only be achieved in single-trial designs. Despite some variability in the evidence reported, the general finding is that prospective judgments tend to produce longer and less variable estimates compared to retrospective judgments (Block & Zakay, 1997). Another robust effect tested using these paradigms is the influence of **sensory modalities** on temporal estimation. A classic result is that auditory intervals are estimated with higher accuracy than visual intervals, and when comparing equal durations auditory intervals are perceived to be longer than visual stimuli (Goldstone & Lhamon, 1974; Wearden, Edwards, Fakhri, & Percival, 1998). Duration estimation also depends on whether the intervals are **'filled'** (continuous beep or image) **or 'empty'** (only delimited by a start and stop signal). A standard effect in humans is that filled intervals are estimated to be longer than empty intervals and that they are also estimated more accurately (Brown, 1931; Goldfarb & Goldstone, 1963; Long & Mo, 1970).

One important property of time perception and production that has been tested repeatedly using the paradigms mentioned above is the fairly constant relative accuracy in temporal judgement over various intervals. Not only the duration of perceived time but also the variability with which we judge time proportionally follows the magnitude of the stimulus (Allan, 1998; Gibbon, 1977; Killeen & Weiss, 1987), because time accumulates in a linear manner (as the interval unfolds, pulses are accumulated regularly). This means that we do not only generally perceive the duration as longer when the stimulus lasts longer, but we are less precise when the duration increases. For instance we might estimate the duration of a 10s-stimulus as being somewhere between 8 and 12 seconds, but will estimate a 30s-stimulus within a wider window of incertitude, perhaps 20 to 40 seconds. In mathematical terms, it is reflected in a constant coefficient of variation (ratio between the standard deviation and the mean of the sample). This property is referred to as the **scalar property or Weber's law**, and mirrors a hallmark characteristic of sensory perception such as hearing and vision (Fraisse, 1984). Several authors however report violations of the scalar property, for instance when systematically testing wide ranges of durations (Bizo, Chu, Sanabria, & Killeen, 2006; Lewis & Miall, 2009; Matthews & Meck, 2014). Despite its shortcomings, the scalar property has been, and still is, instrumental in shaping theories of timing and time perception, in particular the clock-based models (see section 1.2.1).

1.2.1.2 Temporal sequence & rhythm

If duration measures the extent of time, rhythms (cycles in time) and temporal order (chronology or sequence in time) are other features of our experience of time that have been the focus of dedicated experimental paradigms. In a **simultaneity judgement** paradigm (Clark & Geffen, 1990; Engel & Dougherty, 1971; Exner, 1874; Hirsh & Fraisse, 1964; Stone et al., 2001), two events are presented either synchronously or with a short delay between them, and participants are asked to judge whether the events occurred simultaneously or not. A particularly prolific branch of research has investigated simultaneity judgements between sensory modalities, for instance a visual stimulus and a sound (e.g., Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vatakis, Navarra, Soto-Faraco, & Spence, 2008; Zampini, 2005). This research is extremely relevant for everyday life phenomena such as the integration of visual and auditory motion or speech signals. Another approach to investigate the processing of near-synchronous events is to use a **temporal order judgement** paradigm (Bald, Berrien, Price, & Sprague, 1942; Hirsh

& Sherrick, 1961; Rutschmann & Link, 1964), in which two events are presented in different orders and the participants are asked to indicate which event occurred first or last. Interestingly, both types of tasks seem to rely on distinct neural mechanisms (Matthews, Welch, Achtman, Fenton, & FitzGerald, 2016), arguably reflecting that the situations in which we need to make one or the other type of judgment are distinct.

In terms of the production of rhythms, the principal experimental paradigm used is the **tapping** task (Wing & Kristofferson, 1973), which tests a person's ability to produce a given rhythm with some form of motor behaviour (usually tapping a finger). Classic tapping paradigms usually involve both a **synchronisation** task in which a participant is asked to synchronise their behaviour with an ongoing rhythmic stimulus (e.g., a metronome), and a **continuation** task in which the participant is asked to continue a rhythm that is interrupted. The results of tapping tasks are usually analysed to derive a measure of how precise participants are in synchronising with a rhythm, with the presence of immediate feed-back (synchronisation) or without (continuation).

1.2.2 Main theories

1.2.2.1 Dedicated time systems: internal clock models

The longest-enduring and most influential model in time perception is without a doubt the internal clock model, and its most well-known exponent is the Scalar Expectancy Theory (SET; Church, 1984; Gibbon, Church, & Meck, 1984; Meck & Hunter, 1984; Treisman, 1963, 1984). The model, which is illustrated in Figure 1.3, proposes that the processing of time occurs in three stages: a clock stage, a memory stage, and a decision stage. At the clock stage, a pacemaker produces regular 'pulses' that are stored into an accumulator gated by an attentional switch, which opens and closes at the start and end of an attended event (Lejeune, 1998; Zakay & Block, 1996). The number of pulses is then transferred into working memory to be compared to previously experienced durations stored in a temporal reference memory, which thus allows a 'comparator' to compare previously stored (reference memory) with recently experienced (working memory) durations to make a decision about the duration of events or to initiate behaviours of a certain duration.

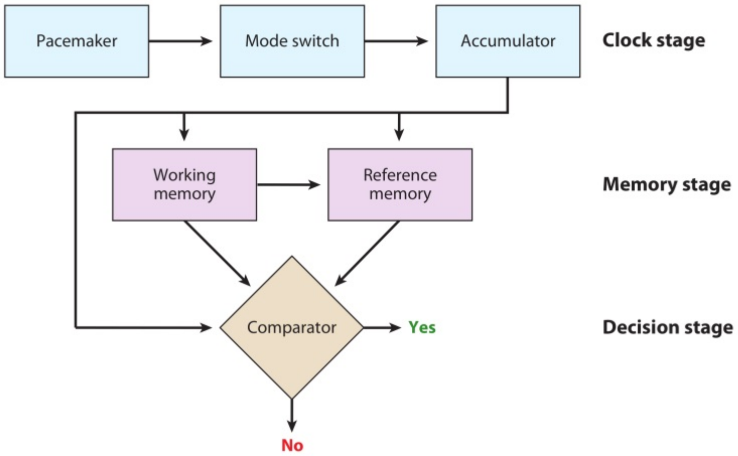

Figure 1.3 Representation of the Scalar Expectancy Theory (SET) model of time perception. Reproduced with permission from Allman, Teki, Griffiths, and Meck (2014).

The original clock model has been adjusted over the years to account for several phenomena. The pacemaker, for instance, is known to be susceptible to arousal and emotional states (Angrilli, Cherubini, Pavese, & Mantredini, 1997; Droit-Volet, Brunot, & Niedenthal, 2004; Droit-Volet & Meck, 2007; Lambrechts, Mella, Pouthas, & Noulhiane, 2011; Noulhiane, Mella, Samson, Ragot, & Pouthas, 2007; Tipples, 2008) and drug administration (Cheng et al., 2016; Meck, 2005), which speed up or slow down the pace at which pulses are produced, thus biasing our experience of time depending on our physiological state. Second, attention modulates the switch which gates the accumulator (Burle & Casini, 2001), so that when attention is not focused on the timed event that is to be processed, pulses emitted by the pacemaker are lost and not transferred into the working memory store.

Other clock models have been proposed that generally fall into one of two categories: those (like SET) with a pacemaker-counter and those with an oscillator system (Grondin, 2010). Whereas pacemaker models suggest that time is counted on an "abacus-like" system which accumulates temporal units, oscillator models work more like a "ruler" which measures time against a set of existing standard scales (oscillating signals of various frequencies and phases). In contrast with pacemaker models, oscillator models are dynamic, non-linear systems (Large, 2008; Schöner, 2002). They rely on physical regularities in the environment (for example heartbeat, music, speech, interlimb coordination) to mark the beginning and end of an event.

The existence and underlying physiological substrates of a dedicated 'clock' have been passionately debated. Proposals of "biological pacemakers" have included dedicated local systems often based in the cerebellum or the basal ganglia, which have the ability to generate an on-going, regular signal of neural firing (Grondin, 2010), whereas oscillator models rely on distributed networks in which time is parsed through oscillations, often the combination of excitatory and inhibitory neuronal populations (Large, 2008). Related to this debate, the possibility of the existence of several parallel clocks has been discussed (Buhusi & Meck, 2009), often to explain the sensory modality effect referred to above whereby the timing of auditory as compared to visual information is processed more accurately. Independent clocks for each sensory modality account for the fact that a same duration is perceived differently in different modalities. However it makes it more difficult to explain how we perceive multisensory events, or how we integrate events perceived in various modalities in a temporally congruent way. The memory stage has also been the topic of debate as to how much it overlaps with the general memory units. The initial proposal did not specify whether temporal working memory and reference memory were similar to short-term or working memory and long-term memory. Subsequent accounts tried to accommodate both systems with some difficulty. The challenge of locating the different parts of the "clock" in the brain is a major critique of the SET and other dedicated clock models. But despite its limitations, SET has proved to be very powerful in describing and predicting timing behaviour in animals and humans.

1.2.2.2 Time as an emergent property of non-specific neural events

At the opposite end of the theoretical field, more recent neurobiological models propose that time can be measured without resorting to a dedicated system, whether a pacemaker- or oscillator-based clock. Instead, they present time perception as an emergent property of the on-going neural activity generated by domain-specific systems, in which temporal information is represented by the spatial spread or the energy of neural activation in distributed networks (Ivry & Schlerf, 2008). These models were born from the need to bring psychological models closer in line with neurobiological data and the failure to identify a neural "clock". Another important rupture with previous models is that, whereas clock-based models often used the scalar property of time as a "litmus test" for validity, the new neurobiological models are less centred on this single property which has repeatedly been shown to be violated.

Buonomano's state-dependent network model is probably the most representative of the neurobiological models (Buonomano, 2000; Karmarkar & Buonomano, 2007). The authors propose that the passage of time is marked in the brain by a succession of neural patterns of activity. In the original paper, Buonomano (2000) showed that a network of excitatory and inhibitory units can be tuned to respond selectively to various temporal intervals (in the range of tens to a few hundred milliseconds) by changing the weights of synaptic connections between units, which is simply normal neural plasticity. As a result, temporal intervals but also sequences of intervals can be coded by the neural activation of successive waves of interval-specific neurons, or state-dependent networks (SDN). They argue, using the principle of parsimony, that temporal processing of short durations does not require a dedicated system since existing cortical networks can be used as they are to code temporal intervals. Instead, they propose that using time-dependent neuronal properties, complex cortical networks are capable of coding short intervals as time-dependent changes in the network's state. As a result, the SDN model supposes that time is encoded non-linearly, in other words temporal information is not accumulated over time but the encoding of a temporal interval depends on previous neural states. Karmarkar and Buonomano (2007), testing their model to fit experimental data, conclude that SDN is a good candidate for short durations up to about 500ms which are encoded non linearly, immune to attentional load, and do not follow the scalar property. These durations are encoded as distinct "temporal objects", rather than the sum of shorter intervals.

1.2.2.3 Time as emotional moments

An unusual but elegant model of time perception was proposed by Craig in 2009. In this model 'awareness of time' is based on interoceptive feelings. Homeostatic feelings of pain, hunger, heat or cold as well as emotional feelings progress from the posterior towards the anterior insula, where they are integrated into an assessment of the body's well-being, which, depending on its salience can reach a certain degree of awareness. For instance, acute pain, hunger or fear might reach a high level of awareness because it is important for the individual to react and act on the situation. In Craig's model, these 'emotional moments' form the matter of time perception. The model proposes that this progression ultimately leads to the integration of salience across all conditions in a unified meta-representation of the 'global emotional moment'. This processing stage thus constitutes an image of the sentient self at the immediate moment of time ('now'). An interesting aspect of this model is the notion that perturbations of interoceptive feelings and emotions should affect time perception. This resonates not only with everyday experience but also with experimental results of how mood, emotion and affective disorders transform our perception of time (Droit-Volet & Meck, 2007; Gil & Droit-Volet, 2009).

1.2.2.4 Time as part of a Magnitude System

A theory of time with a slightly different scope is one proposed by Gallistel and Gelman (2000) and Walsh (2003), which presents time as one of various 'magnitudes' that the brain is able to estimate, produce and compare. According to this theory of magnitude (ATOM; Bueti & Walsh, 2009; Walsh, 2003), continuous abstract quantities that measure physical dimensions such as space, number, intensity and time undergo a common processing stage following the integration of primary sensory information. At that stage, which is proposed to take place in the parietal cortex, any type of quantity (such as time, numerosity, spatial extent etc.) is mapped onto a generic metric or magnitude in order to be estimated or compared. The theory stems from the idea that it is evolutionarily parsimonious to have one mechanism or network to process similar operations (i.e., operations related to 'quantity'). Another intuitive point is that in the environment quantities usually change in the same direction, for instance it takes more time to travel a long distance and more units are likely to cover more surface. A large amount of evidence in favour of this theory comes from interference studies in which one dimension is manipulated to vary in a congruent or incongruent manner with another. These studies show that the estimation of one dimension is affected by the changes in a concomitant dimension. For instance a group of dots are judged to be more numerous or cover more surface when presented for a longer time (Lambrechts, Walsh, & Van Wassenhove, 2013). An interesting feature of this theory is the link between quantities, the way they can affect each other and potentially the way they can combine or correct each other: temporal information is not only provided by the accumulation of time but also by information on other quantities.

1.2.3 Neurobiology of time perception

1.2.3.1 Cerebellum and prefrontal cortex

Contrary to the memory impairments seen in amnesia or the motor impairments seen in Parkinson's disease, there is no neurological disorder that is characterised specifically or primarily by a deficit in time processing. However some focal brain lesions have been found to produce timing dysfunctions. One of the earliest regions identified as important for temporal processes is the cerebellum. Lesions in this area have been shown to affect the accuracy and precision of temporal judgements in both perception and production tasks (Ivry & Keele, 1989), and a number of studies have implicated the cerebellum in motor timing in the sub-second range of durations (Gooch, Wiener, Wencil, & Coslett, 2010; Harrington, Lee, Boyd, Rapcsak, & Knight, 2004; Ivry & Keele, 1989; Spencer, Brien, Girges, Hill, & Johnston, 2007). Moreover, timing difficulties in patients with cerebellar lesions seem unaffected by attentional or working memory load (Coull, Cheng, & Meck, 2010), which is likely due to the fact that the cerebellum is involved in the processing of sub-second durations only, which depend minimally on attentional and mnemonic processes. TMS studies have confirmed that stimulating the cerebellum impairs timing of sub- but not supra-second durations (Fierro et al., 2007; Koch et al., 2007; Lee et al., 2007). The cerebellum has been shown to have a role in predicting the onset of upcoming perceptual events (Bares et al., 2007; Beudel, Renken, Leenders, & de Jong, 2009; Beudel, Galama, Leenders, & de Jong, 2008; O'Reilly, Mesulam, & Nobre, 2008), but it seems not to be needed for timing when temporality is an emergent property of a motor act, such as when we are cycling and incidentally producing regular movements, as opposed to when we are tapping a rhythm and actively trying to produce even durations (Spencer, Zelaznik, Diedrichsen, & Ivry, 2003).

On the other hand, interestingly, TMS stimulation of the right prefrontal cortex (PFC) affects supra- but not sub-second timing (Jones, Rosenkranz, Rothwell, & Jahanshahi, 2004; Koch et al., 2007). Although more research is needed to confirm the specific role of the prefrontal cortex in the duration perception and rule out a role of attention and memory processes, patient studies indicate that lesions in the right prefrontal cortex lead to underestimations of durations in the supra-second range (Danckert et al., 2007; Koch, Oliveri, Carlesimo, & Caltagirone, 2002; Perbal-Hatif, 2012). Whilst the cerebellum allows to establish temporal predictions in the first place, the right prefrontal cortex appears to be crucial to update these temporal predictions based on the flow of time and whether an event has already occurred or not. For instance, if we expect an event to happen, the more we wait, the higher our expectation grows that it is going to happen soon. Patients with lesions in the right prefrontal cortex (but not healthy participants or patients with lesions in the left or medial prefrontal cortex) fail to show faster reaction times to an event occurring after a long compared to a short fore-period (Stuss et al., 2005; Trivino, Correa, Arnedo, & Lupianez, 2010; Vallesi, McIntosh, Shallice, & Stuss, 2009; Vallesi, Shallice, & Walsh, 2007).

1.2.3.2 Basal ganglia

The combination of patient studies and animal studies has informed our knowledge about the neuropharmacology of time perception, involving principally the dopaminergic system as well as the cholinergic system. Most of the literature in this area uses the framework of SET and its three components: the clock, memory and decision stages. The dopaminergic system in particular (in the basal ganglia) has been shown to be involved at the clock stage (Buhusi & Meck, 2002) with dopamine agonists speeding up the clock whilst dopamine antagonists slow down the clock. The cholinergic system, in contrast, affects attentional mechanisms and the memory stage (Meck, 1983a, 1996, 2006), involving the frontal cortex.

Disturbances of the dopaminergic system, for instance in pathologies such as Parkinson's disease and ADHD, are associated with diminished accuracy when estimating short durations (Harrington et al., 2011; Merchant, Luciana, Hooper, Majestic, & Tuite, 2008; Yang et al., 2007). Parkinson's Disease (PD) results from the degeneration of dopamine-producing neurons in the substantia nigra which project to the dorsal striatum of the basal ganglia. PD is therefore a good model of basal ganglia dysfunction and has been studied as such in the context of time perception. Contrary to cerebellar patients, PD patients don't seem to have difficulties with implicit perceptual timing: they can predict the trajectory of a moving object (Bares, Lungu, Husarova, & Gescheidt, 2010; Beudel et al., 2008) and show typical reaction time benefits for temporally predictable targets in simple and choice reaction time tasks (Jahanshahi, Brown, & Marsden, 1992, 1993; Praamstra & Pope, 2007). However PD patients show difficulties when it comes to explicit timing: temporal discrimination has been reported to be impaired both in the sub-second and seconds range, in patients off and on medication (Artieda, Pastor, Lacruz, & Obeso, 1992; Harrington, Haaland, & Hermanowicz, 1998; Smith, Harper, Gittings, & Abernethy, 2007; but see Wearden et al., 2008). In addition the degree of impairment seems to relate to the severity of the disease (Artieda et al., 1992). Finally, the basal ganglia are thought to be involved in determining the clock speed during encoding, affecting the mnemonic process: patients off medication showed exaggerated central tendency for suprasecond durations, which means they overestimated a short duration and underestimated a long duration when presented in the same session (Koch et al., 2008; Malapani et al., 1998), bringing them closer to a central value.

Data from functional neuroimaging further clarify the role of the basal ganglia in timing and time perception, establishing the dorsal striatum as a context-independent "supramodal timer" (Coull et al., 2010). In particular, the putamen and caudate nucleus of the dorsal striatum and the globus pallidus are most often activated during timing tasks (Coull & Nobre, 2008; Meck, Penney, & Pouthas, 2008). Neuroimaging data also contributed to show that basal ganglia are not the only brain region activated during timing tasks, but rather a corticostriatal network of areas, most often the supplementary motor area and prefrontal cortex. Enriching the focal approach of clinical and TMS studies, functional imaging provided a whole-brain approach that revealed a functionally integrated network of timing and time perception. PFC and supplementary motor area (SMA) in particular were consistently found to be activated in various temporal tasks, as well as the cerebellum, premotor cortex and preSMA.

Because of the multiplicity of contexts and processes in which it is required, temporal processing involves many neural structures including the cerebellum, prefrontal cortex, basal ganglia and sensory cortices. Over the last decade or so, Meck and colleagues (e.g., Cheng et al., 2016; Coull, Cheng, & Meck, 2011; Meck & N'Diaye, 2005; Meck, 1983, 1996, 2005) have proposed a unified timing model which combines a core timing network and context-specific structures (see figure 1.4). This system has been interpreted mainly in the framework of SET model and its three stages: the clock, memory and decision stages. The core timing network consists of cortico-thalamic-basal ganglia circuits activated primarily by dopamine and glutamate, which are the closest elements to an “internal clock”. The nigrostriatal dopaminergic pathway in the dorsal striatum of the basal ganglia in particular has the ability to integrate dopaminergic and glutamatergic inputs. Context-specific structures include sensory areas (timing of sensory events), the cerebellum (motor timing), and the prefrontal cortex (working memory).

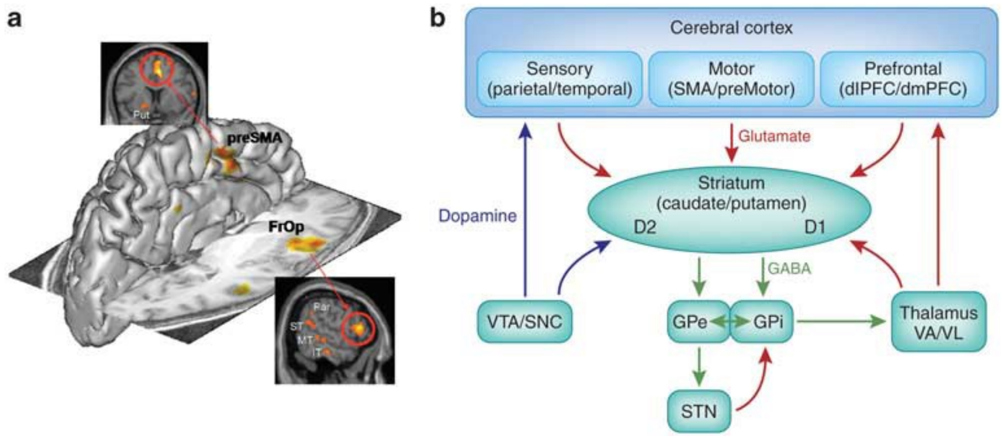

Figure 1.4 Corticostriatal circuits for interval timing.
Reproduced with permission from Coull, Cheng, and Meck (2010). Human functional imaging data (a) showing the corticostriatal circuits (b) implicated in interval timing. Blue lines represent dopaminergic input, green lines represent GABAergic input, and red lines represent glutamatergic input. FrOp, frontal operculum; GPe, globus pallidus external capsule; GPi, globus pallidus internal capsule; preMotor, premotor cortex; dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; Par, inferior parietal cortex; Put, putamen; SMA, supplementary motor area; SNC, substantia nigra pars compacta; STN, subthalamic nucleus; VL, ventral lateral nucleus of the dorsal thalamus; VA, ventral anterior nucleus of the dorsal thalamus. IT/MT/ST, inferior/middle/ superior temporal cortex. Figure a is adapted from Coull, Vidal, Nazarian, and Macar (2004).

1.3 ASD and temporal processing

1.3.1 Suspicion of atypical timing and time perception in ASD

Time, as developed in the previous section, is a ubiquitous part of our perception of the world, and our action upon it. Whether it is used to unify our sensations (sensory integration), coordinate our actions (motor system), colour our perceptions (through duration and rhythm), or give a sense of direction and growth to our experiences (flow of time), time is a lens through which we experience the world. In light of these considerations, atypical temporal processing in autism is a good candidate to explain why making sense of the world and taking optimal advantage of learning opportunities throughout development is difficult for many people with autism. In this sense, being on the autism spectrum could be like seeing the world through a different temporal lens. Some evidence in favour of such a view was highlighted in the previous section concerning the brain areas involved in timing and time perception, which are also precisely the areas that tend to be implicated in the neuropathology underlying autism (the cerebellum, the basal ganglia, the frontal cortex). The current section will focus in more detail on the behavioural characteristics of ASD that also point toward differences in the processing of time.

Early signs of autism in the first developmental stages include reduced joint attention (Bruinsma, Koegel, & Koegel, 2004; Dawson, Toth, et al., 2004b), eye gaze (Falck-Ytter, Bölte, & Gredebäck, 2013), play patterns (Jarrold, 2003), hypotonia or hypertonia and movement difficulties (Green et al., 2009; Ming, Brimacombe, & Wagner, 2007; see Wetherby et al., 2004 for a study of early indicators of autism). It could be said that from an early age, the interface between mind, body and environment is not integrated optimally, and the different channels of information and communication (attention, eye gaze, movement) are not well coordinated in time, at a stage when it is most crucial for learning and developing. For instance, autistic individuals show an extended window of sensory integration (Foss-Feig et al., 2010; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011), which means that events which are further apart in time are integrated as part of the same episode. If this is the case at an early age, one of the predicted consequences is that learning associations would be more difficult. Many events (sensory, cognitive, emotional etc.) happen in a short amount of time, and having an extended window of integration just multiplies the possible candidates for associations and make the world more ambiguous. If integrating meaningfully related information becomes difficult, then this could result in missing important learning cues (for instance through pointing and looking to elicit shared attention). Later on in development, difficulties to integrate but also produce cues in a meaningful and timely fashion would likely impair the delicately tuned multimodal behaviours of social interaction, or as put by Amos (2013) the "dance of relationship".

In addition to the moment-to-moment integration and temporal tuning with the world, individuals on the autism spectrum struggle with seizing a sense of flow and chronology in the succession of events. Bound to the present, they find it difficult to imagine the past and future of a current state (Boucher, Pons, Lind, & Williams, 2007) and various clinical, empirical and anecdotal reports underline a lack of "time sense" (Boucher, 2001; Peeters & Gillberg, 1999; Wing, 1996). Everyday life for a person with ASD is often constrained by a strict structure and highly predictable schedule, which arguably helps them find their way in time, whereas deviation from planned activities are a source of anxiety and sometimes of challenging behaviour. Several authors have proposed that repetitive patterns of behaviour, which are a central feature of the clinical ASD phenotype, might be a compensatory strategy to parse time using the accumulation of highly repetitive units (Allman, DeLeon, & Wearden, 2011; Allman et al., 2014; Jill Boucher, 2001; Lewis & Miall, 2003), although the mechanisms of such a strategy are not clarified by the authors.

Although several authors have suggested that timing and time perception may operate differently in ASD (Boucher, 2001; Wimpory, Chadwick, & Nash, 1995; Wimpory, Nicholas, & Nash, 2002), this idea did not echo much in the autism research community until a few years ago. Since then, a number of studies have been published that examine temporal processes either directly or indirectly. This evidence is reviewed in the following section according to mental concepts of time, which may rely on distinct mechanisms and neural substrates, but which all contribute to the notion of time and temporal processing. The first concept is that of a '**sense of time**', perhaps the most intuitive and arguably the most difficult concept to explore empirically, which refers to the feeling of time passing or flowing. The sense of time refers to the continuous happening of events and the ability to see a direction in their succession: past, present and future. It is necessary for mental "time travelling", which is our ability to mentally remember, anticipate and imagine the past and future, and it is intimately related to our sense of self and our ability to plan ahead (Gardiner, 2001; Tulving, 1983). The concept of time passing is also the closest to memory, particularly episodic memory, which places events in context (spatial, temporal, emotional). The second concept which we will review is '**rhythms**': many external events (seasons, music, language) as well as internal events (sleep, hunger, heart rate, brain oscillations) are organised in fairly predictable cycles characterised by their period. They are strong signals that help us parse time and serve as temporal landmarks to predict events and organise behaviours. The third line of evidence will focus on time as '**chronology**': whereas the sense of time encompasses time as a continuous flow, chronology looks at the discrete succession of events and their order (or synchronicity) in time. Chronology is an efficient way to memorise and plan events as it doesn't require a continuous picture but simply orders events relative to each other. Finally, the most productive line of research has been the one concerned with the notion of '**duration**'. Duration qualifies events according to their "weight" or magnitude in time, as delimited by the beginning and end of an event. It is also a fairly intuitive notion that we use to measure or quantify time, which is particularly crucial for prediction and planning.

1.3.2 Time as a flow: the sense of time

Practitioners Peeters and Gillberg (1999) report that "most people with autism feel lost in a sea of time" (p. 87). Indeed anecdotal and clinical reports and self-reports indicate that the sense of time passing, and the ability to place and imagine events in a temporal structure, is lacking in ASD. Insightful clinician Lorna Wing says about the lack of sense of time in autism:

> *"The problems of time are not related to telling the time by the clock, which some people with autistic disorders are able to do well. The difficulties lie in comprehending the passage of time and linking it with ongoing activities."* (Wing, 1996, p.88)

Jill Boucher further develops this idea and proposes the analogy of disconnected clocks:

> *"This is analogous to an individual's having one clock showing hours, another showing minutes, and another showing seconds, rather than having a single clock representing hours, minutes and seconds simultaneously. With three separate clocks an individual would be able to locate extended events temporally within the hours in which they occurred, and to record their temporal succession. They would also be able to locate shorter, more rapidly successive events within the minutes in which they occurred, and to record their temporal succession; and also to locate very rapidly successive events within the seconds in which they occurred, and to record their temporal succession. However they would not be able to encode the temporal relations between these three sets of events, resulting in difficulty in encoding the briefer events as part of more extended wholes, and a corresponding difficulty in breaking down more*

extended events in terms of constituent parts." (Boucher, 2001, p.114)

A prediction of having difficulties with apprehending the passage of time is that temporally dependent functions and activities will be affected. Episodic memory is one of the domains that critically depend on representations of time, and describes the ability to encode and retrieve autobiographical information and other memories that are rich in spatial and temporal context (i.e., what, when, where, with whom etc. did something happen). Despite preserved rote and semantic memory (i.e., memory for decontextualised facts such as the boiling point of water), growing evidence indicates that ASD individuals have difficulties with episodic memory. A particularly prolific line of research resulted from the 'remember/know' paradigm, in which participants, after learning a list of items, have to not only retrieve them but also indicate whether they 'know' (decontextualized, semantic memory) or 'remember' (contextualised, autobiographical memory) having learnt the item previously. Autistic individuals typically report lower rates of 'remembering' and increased rates of 'knowing' as compared to TD controls (Bowler et al., 2007, 2000; Gaigg, Bowler, & Gardiner, 2013; Meyer, Gardiner, & Bowler, 2014; Souchay, Wojcik, Williams, Crathern, & Clarke, 2013), showing that learnt information is encoded with less contextual cues, including temporal context. Whereas this can confer an advantage for remembering factual information such as the boiling point of water, it compromises autobiographical and episodic memories that specify the when, where and how of our experiences. This gives us an insight about how a lack of sense of time can change the nature of our thoughts and memories, which are the foundation of our identity and sense of self.

Memory for the past is only one half of our ability to mentally time-travel. In fact much of our time is spent thinking about the future rather than dwelling on the past. The ability to imagine the past and future of an event rather than focus only on the present is called diachronic thinking, and has been shown to be a source of difficulty for individuals with autism (Lind & Bowler, 2010). For instance, Boucher et al. (2007) asked children and adolescents with no or mild intellectual disability to perform three tasks: make temporal inferences based on a cartoon (e.g., a man lying on a towel at the beach might have been swimming in the sea), represent the past and future state of an entity (a mature tree), and identify the gist of a story based on its episodes. They found that participants with ASD performed worse than their TD counterparts in all three tasks. Moreover, the differences in performance rates could not be explained by other skills and functions such as the ability to make inferences in general, the ability to draw or a lack of knowledge in the area in question. The difficulties in diachronic thinking observed in this study suggest that individuals on the spectrum live more in the present. Speculatively, living in the present deprives autistic individuals from part of their past experience, as well as the possibility to make decisions based on potential future scenarios, which may contribute to the incertitude and anxiety many people with autism report about unpredictable future events (e.g., Boulter, Freeston, South, & Rodgers, 2014). This could underlie autistic individuals' tendency to prefer repetitive and highly predictable schedules. In fact, several authors posited that repetitive behaviours may be a means to parsing time using highly predictable units of time: instead of being experienced as a flow, time is seized as the succession of moments in time (Allman et al., 2011, 2014; Boucher, 2001; Lewis & Miall, 2003). To reduce their disorientation in time, parents and professionals living alongside autistic people use many tools and strategies involving visual and mechanical representation of time such as planners and timers. They also use the occurrence of familiar events to "time" behaviours (e.g., "when the song finishes it's time to go to bed"). These tools tend to improve well-being and coping strategies in autism and help autistic individuals deal with the uncertainty of the future (Hodgson, Freeston, Honey, & Rodgers, 2017).

1.3.3 Cycles in time: rhythms

Cyclical activity happens in both our external (seasons, daylight pattern) and internal (heart rate, sleep, hunger) environment. They are a crucial element to prediction and planning, as well as adaptation. By knowing cycles and their patterns, we are able to predict not only what might happen, but also when it might happen. This is why we value history, archaeology and geology: learning about the past informs us about our possible futures. Predictions are never certitudes (as meteorology exemplifies so well) but they reduce the incertitude we face. Violation of cycles are also a very strong predictor of perturbation and times of change, and detecting them gives us an adaptive advantage (e.g., detecting change in climatic changes gives us the possibility to combat or adapt to global warming). Internally, abnormal sleep or hunger patterns, irregular heart beat are warning signals which may indicate a state of illness or anxiety.

Circadian rhythms are the cycles determined by the revolution of the Earth around the sun, in other words the succession of night and day and the light and temperature cycles that come with them. Humans are equipped with at least two systems to detect circadian rhythms. A first system is the reduction-oxidation (redox) metabolism which we have in common with the simplest organisms, and which makes sure that cellular division (when DNA material is at its most vulnerable) and oxidative metabolism happen during separate time frames. A second system, which we have in common with other mammals, is our circadian clock. This 'clock' is based on transcription loops of "clock genes" and their protein products. Crucially, the circadian clock is entrained by light, and reset by changes in lighting in the environment (Nicholas, 2015). Redox metabolism and circadian clock can each maintain circadian rhythms but their combination is believed to be evolutionarily more robust and potentially advantageous (Edgar et al., 2012). Genome scans suggest that at least two clock genes might present an atypical genotype in autism: Per1 and Npas2, and additional evidence suggest abnormal methylation patterns on other regions (Nicholas et al., 2007; Wimpory et al., 2002). Differences to the expression of clock genes in autism could relate to sleep disorders often experienced by ASD individuals (Couturier et al., 2005; Krakowiak, Goodlin-Jones, Hertz-Picciotto, Croen, & Hansen, 2008; Malow et al., 2016; Souders et al., 2009) and difficulties with timing.

Embedded in the circadian periods, faster rhythms pace our behaviour throughout the day: music, conversations, heartbeat, running pace, etc. Exploring our ability to perceive and align to such rhythms, Gowen and Miall (2005) and Sheridan and McAuley (1997) asked participants with ASD to either synchronise finger taps with the repeating tones, or to continue a series of tones by producing more tones at the same pace (using inter-tone intervals of 400 to 800ms). In the earlier study, children with ASD were found to show the same tendency to anticipate the tones as TD children, with a similar developmental effect (the asynchrony is larger in younger than in older children). ASD children however showed overall a greater variability in the interval produced, which the authors interpreted in terms of increased clock variance. These results have to be taken with some caution because the groups were not strictly matched in level of abilities. In the later study, Gowen and Miall (2005) showed that while TD adults found the continuation task more difficult than the synchronisation task (with a tendency to anticipate the tones), adults with ASD found both equally difficult and overall more difficult than the TD adults, with a greater tendency to tap too early. Synchronisation is essentially an entrainment task in which the participant must align an endogenous rhythm (the tapping response) to an external rhythm (the stimulus). It is thought to engage a basal ganglia-cortical loop which computes feedback-based error correction (Pfenning et al., 2014; Repp, 2005). In contrast, in the continuation task participants must maintain the pulse without external feed-back, which is thought to rely on a cerebellar and frontostriatal circuit (Cerasa et al., 2006). Given cellular, functional and connectivity atypicalities in cerebellar, limbic and cortical structures in autism (e.g., Eigsti & Shapiro, 2003; Herbert, 2004), it is perhaps not surprising that performance is affected in both tasks. In Parkinson's disease, where sensory-motor timing is compromised by basal ganglia dysfunction, evidence suggests that patients compensate by using a cerebellar and frontostriatal circuit in both synchronisation and continuation tasks (Cerasa et al., 2006). It is possible that, similarly, compensatory pathways are activated in autism in rhythmic tapping tasks. Although the link has yet to be established, increased variability to align with an external rhythm could contribute to difficulties aligning with external rhythms such as conversation and building an efficient rapport with one's interlocutor.

1.3.4 Time as chronology

Whereas rhythms consist of the repetition of the same occurrence in time, chronology is concerned with the order in which events occur. For two events there are two possibilities: they happen synchronously, or asynchronously, in which case one event precedes the other. This defines two types of judgements: synchrony judgements (SJ) and temporal order judgements (TOJ), which are particularly relevant for the domain of sensory integration. Sensory integration in turn impacts on many higher level cognitive functions and behaviours such as learning, communication, motor behaviour and sense of agency, all of which have been shown to be impaired or atypical in autism (Iarocci & McDonald, 2006; Moore & Fletcher, 2012).

Few studies to date have purposefully examined chronology processing in autism. Part of the evidence available comes again from the memory literature. Focusing on short-term memory, Poirier, Martin, Gaigg, and Bowler (2011) found that adults with ASD had more difficulties detecting a change in the order of a memorised sequence of items compared to their TD counterparts. Gaigg, Bowler, and Gardiner (2013) further demonstrate that adults with ASD show difficulties specifically with episodic temporal order, and not with semantic chronological order: autistic participants show no difference in performance compared to TD participants when it comes to ordering historical figures based on their chronological order in history (semantic chronology), but they show poorer performance when it comes to ordering them according to their order of presentation during the task (episodic chronology).

Another branch of the literature interested in chronology detection in autism comes from the sensory integration domain. Foss-Feig et al. (2010) and Kwakye, Foss-Feig, Cascio, Stone, and Wallace (2011) reported an extended window of perceptual integration in children and adolescents with ASD, both in auditory and audiovisual modalities but not in the visual modality. Foss-Feig et al. (2010) used the flash-beep illusion (Shams, Kamitani, & Shimojo, 2000) in which the production of two short auditory tones and one visual flash in a small interval of time often leads to the subjective report of two visual flashes. Crucially, the likelihood of reporting two flashes depends on the timing between the two tones (the visual stimulus being presented simultaneously with one of the two tones). This set-up makes it possible to measure a 'temporal binding window' during which multisensory inputs are likely to be interpreted as one event. Results showed that ASD children have an extended window of integration ([-300; 300 ms]) compared to matched TD children ([-200; 200 ms]). Kwakye and colleagues (2011) confirm and expand these results using unisensory and multisensory temporal order judgement (TOJ) tasks. In the unisensory conditions, children and adolescents with and without ASD had to decide whether an upper or lower visual cue (visual condition) or the tone played in the left or right ear (auditory condition) was presented first. In the multisensory condition, participants performed the visual task with the addition of two tones: the first tone was always synchronous with the first visual stimulus, and the second tone was presented with a variable Stimulus Onset Asynchrony (SOA) centred on the presentation of the second visual cue. Depending on its timing, the auditory cue was expected to facilitate the TOJ performance or not (Hairston, Burdette, Flowers, Wood, & Wallace, 2005; Hairston, Hodges, Burdette, & Wallace, 2006; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Results revealed no differences between groups in the visual task, suggesting that TD and ASD participants processed visual temporal information similarly. In the auditory condition, ASD participants required longer SOAs (often described as higher thresholds) to determine the order of the tones played in the left and right ears. Crucially, in the multisensory condition, participants showed improved accuracy when the additional auditory cue was close to the second visual stimulus and the temporal window for which participants improved was wider in the ASD ([0; 300 ms]) than in the TD group ([50; 150 ms]). The addition of the auditory cue also resulted in faster reaction times in ASD but not in TD participants. Both studies (Foss-Feig et al., 2010; Kwakye et al., 2011) propose that as a result of this wider window of temporal integration of events across modalities, individuals with ASD might be less able to identify the source modality of information, which could contribute to difficulties orienting to a stimulus. In addition, they may not be able to respond to an event in a given sensory modality if another event happens concurrently in another modality.

Several studies have documented how the detection of temporal synchrony in intermodal processing relates to language development. Infants as young as 10-17 weeks of age have been shown to match videos of clashing objects with their soundtrack (Bahrick, 1983), but also dynamic facial movements and speech on the basis of voice-lip temporal synchrony (Dodd, 1979; Spelke & Cortelyou, 1981). In the context of development, having a wider window of multisensory integration would make it more difficult for autistic individuals to disambiguate complex sensory information from various modalities, such as speech and facial expressions, and possibly delay language development and related skills. Using a preferential-looking paradigm, Bebko, Weiss, Demark and Gomez (2006) found that children with autism could detect asynchrony between video and audio tracks for non-linguistic stimuli, but not for speech stimuli (including simple counting), in contrast with children with other developmental disabilities. Although data from Foss-Feig and colleagues (2010) and Kwakye and colleagues (2011) does not pre-empt the discrepancy between speech and non-speech stimuli, having a wider window of integration could account for difficulties distinguishing between synchronous and non-synchronous events. Because speech comprises a particularly complex set of events (lip movements, facial expressions and voice features are temporally coordinated but are not directly caused by one single source) it is possible that individuals with autism struggle more with multimodal integration of linguistic events. For instance some data indicate that adolescents with ASD have difficulties with audiovisual speech integration and lip-reading (Smith & Bennetto, 2007) which could undermine speech comprehension and account for some delays in early language development. The second part of this work (in chapters 3-5) will explore temporal processing in communication behaviour more specifically in ASD, where a more comprehensive review of the evidence of temporal anomalies in speech and other communicative aspects will be provided.

1.3.5 Extent in time: duration

Although the sense of time, rhythms and chronology have been the object of some research, the dimension of time which is probably the most commonly explored empirically is duration. Duration is the extent of an event or “being" in time, an amount of time. It is a measurement along the continuous variable of time elapsing, and allows us to compare, anticipate and plan different events. For instance estimating duration makes it possible for us to know how long to let the phone ring before we assume no one is going to answer, stop the milk from boiling over without watching continuously over it, or know when our interlocutor has finished speaking and we can take a turn in the conversation. One way to test the ability to estimate duration is to use an oddball paradigm, in which a series of standard events of identical durations are interspersed with “deviant” stimuli with a slightly different duration, and probe whether participants detect the change. Using electroencephalography (EEG) recordings in a deviant detection task in two separate studies, Lepistö and colleagues (2005, 2006) found a diminished mismatched negativity (MMN) in response to changes in duration in speech and non-speech stimuli in children with ASD. A subsequent study by the same group (Kujala et al., 2007) revealed enhanced MMN in response to duration deviants in adults with ASD, suggesting that both children and adults show atypical patterns of duration discrimination in the subsecond domain, although it is unclear why the pattern reverses with age.

In terms of behavioural studies, production, bisection, comparison and generalisation tasks, which were introduced in the previous section, are classic paradigms to investigate the ability to process duration and dominate the (small) literature on time perception in autism. Falter and colleagues (2012) used a temporal generalisation task to assess duration estimation in adults with ASD, in auditory, visual and cross-modality conditions. Participants were presented first with a standard duration (600 or 1000ms), followed by a probe duration (300 to 900ms or 500 to 1500ms) and decided whether they were identical or different. In all modalities ASD individuals were found to be less sensitive to small temporal differences compared to TDs, showing a greater tendency to judge non-identical probes equal to the standard (i.e., generate false alarms) without any increase in hit rate for identical probes. Similar results were found in children with autism who also had intellectual impairment (Brodeur, Gordon Green, Flores, & Burack, 2014) using both a bisection task (anchor durations 200 and 800ms) and a generalisation task (standard duration 500ms) with comparison durations ranging from 200 to 800ms and from 125 to 875ms respectively. ASD children were found to show lower sensitivity as compared to their mental-age-matched TD counterparts. However, in the same range of durations but using a comparison task, in which participants judged which of two auditory tones was longer, neither Jones et al. (2009) nor Mostofsky et al. (2006) found any difference between TD and ASD children and adolescents.

A similar picture emerges in studies that have examined multi-second duration ranges. Allman, DeLeon, and Wearden (2011) used a bisection task in the visual modality with children. Participants learned to recognise two anchor durations ("short": 1 or 2s and "long": 4 or 8s) and were then presented with varying probes (1 to 4s in steps of 0.5s or 2 to 8s in steps of 1s) with instructions to judge which anchor they most resembled. Results showed that ASD children had a tendency to overestimate probe durations in comparison to the TD children, especially for durations between 3.5 and 5s. They also showed reduced sensitivity in the 2-to-8s range. A subsequent study by Gil, Chambres, Hyvert, Fanget, and Droit-Volet (2012), however, found no differences between TD and ASD children in a bisection task using durations ranging from 0.5 to 16.62s. In reproduction tasks, the evidence is somewhat more consistent. Szelag, Kowalska, and Galkowski (2004) asked TD and ASD children to reproduce durations in the visual or auditory modality by pressing a key. They found that children with ASD were very imprecise, reproducing every interval between 1 and 5.5s as 3s on average, with large variability in reproduction times. Martin, Poirier, and Bowler (2010) and Brenner et al. (2015) also used a reproduction task in the auditory and visual modalities respectively, but with adults. Both studies found that ASD individuals were both less accurate (i.e., reproduced intervals further from the standard on average) and less precise (i.e., responses were more variable) than TD adults. On the whole, therefore, studies employing psychophysical methods to examine time processing of sub-second and multi-second durations tend to suggest that ASD individuals show somewhat reduced sensitivity.

The observations across studies, however, are far from consistent and the reasons for discrepant findings remain obscure. Some light on the neural underpinnings of reduced sensitivity to temporal information in autism was shed by Lambrechts, Falter-Wagner, and van Wassenhove (2017) who recorded magnetoencephalographic (MEG) responses to two consecutive auditory tones in two discrimination tasks: in one task, participants had to decide whether the two tones (a constant standard and a variable probe) had the same pitch, and in the other they had to decide if they had the same duration. The response evoked by the standard tone in the ASD group showed a delayed offset with a flatter slope as compared to the TD group, indicating that the encoding of a highly predictable tone was less sharp in ASD adults. Difficulties to encode a repeated, predictable duration accurately could account for diminished sensitivity in autism, without jeopardising duration processing altogether. Because it only increases the variability of the duration encoded rather than introduce an encoding error, it would also account for the fact that behavioural performance is found to be inconsistent across studies. Lambrechts and colleagues also found that compared to typical adults who tend to allocate neural resources differently to either task depending on instructions, autistic adults seem to allocate less neural resources to duration processing regardless of the instructions (pitch or duration discrimination). Although this finding needs to be replicated before drawing firm conclusions, this study offers a potential correlate to the diminished sensitivity in temporal processing in autism: lesser allocation of resources to time processing could account for reduced accuracy in duration encoding and as a result lower sensitivity in duration judgements.

1.3.6 Time theories of autism

This review of the literature reveals a fragility to process time across various concepts and dimensions in autism. Although temporal processing is not altogether disrupted, it seems that autistic individuals live in a more uncertain temporal world: their sense of time is not as strong as in typical individuals and they can struggle to mentally travel at will to the past or the future to reminisce or plan ahead. Information is more liberally integrated over time, which results in uncertainty when learning meaningful associations and cause-and-effect phenomena. Finally their sense of duration is fuzzier, which matters particularly in finely-tuned behaviours such as motor coordination, language and communication.

The evidence to date has lead several authors to posit impaired time perception as a central feature of autism with top-down implications for the development of other functions (Allman & DeLeon, 2009; Boucher, 2001; Wimpory et al., 2002; Wimpory, 2015). Allman (2011) even goes so far as to propose impaired temporal processing as a unifying account of ASD: atypical temporal dynamics in social interaction can affect bonding, joint attention and effective communication from an early stage in development. Repetitive behaviours can be reinterpreted as a coping mechanism to parse temporal information. Reviewing evidence of temporal disorientation at various levels, Allman argues that impaired temporal processing could be the root not only of the core symptoms of ASD, but also of other dysfunctions which are at the heart of leading theories of autism: weak central coherence, impaired theory of mind and difficulties with executive functions. Due to abnormal windows of integration for instance, atypical sensory integration may affect perception, learning and memory which would have repercussions for central coherence. Understanding and empathising with others, and the ability to "put ourselves in someone else's shoes" required for theory of mind are likely to depend highly on episodic memory and mental time travel (Perner, 2000; Perner, Kloo, & Gornik, 2007), which are both impaired in autism. Finally difficulties with temporal processing could impact information processing generally with a wide impact on executive functions (working memory, planning etc.).

Narrowing the scope to social timing, Wimpory (2015) provides a more specific and detailed account of how timing deficiency in autism could underlie the atypical development of social functions in autism. She reviews evidence of how social timing is crucial to typical development. From birth, infants show temporal entrainment to parents' speech and behaviour (Condon & Sander, 1974; Malloch, 2000). For instance, new born infants align their limb movements and vocalisation to salient moments in their parents' speech (Condon & Sander, 1974). As they grow, synchronicity and rhythm alignment develops into turn-taking: at 6 weeks old, infant-parent dyads show alternating patterns or proto-conversations (Trevarthen & Aitken, 2001). Parent-infant interaction starts to be better defined around 2-3 months of age, with notably the start of gaze contingency, and by 5 months a pattern of synchronised and turn-taking behaviours can be observed in several modalities (gaze, vocalisations, movement, touch etc., Feldman, 2007b). From 9 months onward, switching pauses appear (silences between turns), which marks the development of *conversational* turns (Jasnow & Feldstein, 1986), and by the end of the first year, proto-conversations incorporate symbolic elements, gesture and spoken language, into a temporally tuned exchange. Importantly, synchronicity and temporal coordination between child and carer predict later cognitive and self-regulatory skills as well as empathy (Feldman, 2007a; Kirsh, Crnic, & Greenberg, 1995), illustrating that the temporal coordination of interpersonal behaviour early in life provides important foundations for developmental maturation across multiple domains.

In autism, twin and sibling studies as well as studies matching autistic and non-autistic participants reveal diminished interactive turn-taking in children with autism as compared with their non-autistic siblings and peers (Kubicek, 1980; Trevarthen & Daniel, 2005; Wimpory, Hobson, Williams, & Nash, 2000). Wimpory (2015) makes the case that reduced ability to time and coordinate social interactions early in development through temporal synchrony and turn-taking has a major impact on the on-going development of the autistic child in the cognitive and social domains. For instance, she highlights how many activities, games and generally interactions with infants are based on fine timing:

> *"[…] it is the temporal element that is the critical fun-factor in the game that delivers salience and amusement for the child" (Wimpory, 2015, p.57)*

Failure to process timing accurately reduces the opportunities for the child to find amusement and engage in the social interaction, and makes it more difficult to distinguish which aspects of the world are most relevant to understand and communicate about with their interlocutor. This in turn delays the development of language, or symbolic communication, which has to be established as a convention with others:

> *"Symbols emerge within an interactive context during positive moments between caregiver and child whereby the synchronicity of experience (affective, communicative) is often crucial" (Wimpory, 2015, p.76)*

This idea of atypical temporal coordination and synchronicity in autism in the context of social interaction will be the subject of chapters 3-5 and will be developed in more detail then. The topic of timing and language learning however has also been touched on in a lesser-known model of autism: the Imbalanced Spectrally Timed Adaptive Resonance Theory (iSTART, Grossberg & Seidman, 2006). This neural model claims to combine three models to provide a better understanding of the brain function (START model) and how it dysfunctions in autism. It offers a mechanistic explanation of how impaired timing can affect learning. One of the components of START is the Adaptive Timing Model (ATM), which involves a neural loop between the cerebellum and the Purkinje cells in the hippocampus, two regions implicated in the neuropathology of ASD (see section 1.3 above). According to the model, ATM allows us to evaluate the likelihood of an event occurring or not occurring *in a certain window of time*. This is particularly relevant when considering the reward system: it is important to be able to predict in which timeframe we expect to receive a reward or punishment, and when (failing to receive it) we have to adapt our behaviour accordingly. Knowing the delay within which a reward is likely to happen is crucial to optimise learning: if the child gets frustrated when not receiving a reward immediately after producing a positive behaviour, they might miss out on useful feed-back. If they wait too long to receive it the context has changed and it is too late to adapt their behaviour. Grossberg and Seidman propose that dysfunction of ATM in autism is at the basis of language delays and difficulties with communication which requires learning to time events adaptively. Specifically, they propose that as a result of hippocampal dysfunction, autistic individuals engage in hyperspecific learning, which is more likely to trigger a mismatched response between an expected event and an observed event. In turn, this mismatch fails to trigger an adaptive learning loop in the hippocampus and cerebellum, reinforcing hyperspecific learning. Whilst iSTART doesn't focus on *how* timing is atypical in autism, it provides an interesting account of how abnormal timing could impoverish learning opportunities.

Another theoretical proposal of how atypical temporal processing has cascading consequences on development is proposed by Boucher (2001) in a less language-centred synthesis of the literature. Boucher makes the argument that time-parsing lies at the heart of temporal difficulties in autism. By registering the succession of happenings in time, the biopsychological time-parsing system is what allows us to encode, organise and bind information in digestible chunks. It allows us to define and individualise 'events' and group information according to their closeness in time, whilst providing a sense of continuity in our experience of the world. Importantly, the time-parsing system is necessary to detect regularities in time, and to create novelty. Boucher postulates that part or all of the time-parsing system is dysfunctional, resulting in difficulties to encode information efficiently (which can be experienced as feeling overwhelmed by the environment), difficulties to construct, bind and retrieve episodic memories meaningfully, difficulties to detect patterns over time and predict the future accurately, and difficulties creating novelty in time. A more mechanistic account of similar ideas was put forward under the term of temporal binding theory by Brock and colleagues (2002) who proposed that a reduction in synchronization of high-frequency gamma activity between local networks prevents autistic brains from linking coherent information and events based on temporal cues. Decoding the timing of neural firing is crucial to encode complex, overlapping information efficiently with a finite number of neurons and to combine pieces of knowledge flexibly. Brock and colleagues argued that local networks function more independently in autism, creating islets of information that are less readily available for combination. Contrary to the other theories mentioned earlier in this section, the temporal binding theory pulls away from the notion that the concept of psychological time has an impact on development and learning in autism, and instead suggests that the impact of time processing on autism is more biomechanical, and concerns the establishment of neural connections between different networks.

1.4 Chapter summary

The empirical evidence and theoretical views reviewed in this chapter have hopefully shown that time is ubiquitous in human behaviour and cognition and that there is wide agreement that the disruption of any temporal processing system early in development can potentially have serious repercussions on the acquisition of many cognitive, emotional and social functions. Evidence has also been reviewed which suggests that processes of time, both at the psychological and neurophysiological level seem to be characterised by atypicalities in ASD, although much of the evidence remains indirect and inconclusive with respect to the nature of the abnormality and its relation to the clinically defining features of ASD.

The work presented in this thesis therefore aims to further explore temporal processing in autism. Chapter 2 will present an assessment of time perception in individuals with ASD using a classic psychophysical task that builds on earlier studies in examining accuracy and precision of temporal judgements across single and multiple sensory modalities. Following a discussion of how intrinsic time perception difficulties may be in autism at this basic psychophysical level, the second part of the empirical work will then focus on a fine-grained analysis of the temporal dynamics and characteristics of real-life communication behaviours to examine whether the timing of communication (including verbal and non-verbal channels) is atypical in autism, and whether this could contribute to clinical features of autism. The final part of the thesis will then evaluate timing in communication as a potential biomarker of autism, and reflect on whether assessing temporal processing could inform or complement a the diagnostic process of ASD.

Chapter 2
Assessment of time perception in Autism Spectrum Disorders

2.1 Introduction

As pointed out in the introduction, individuals with ASD present a complex phenotype of neurocognitive functioning, characterised by an atypical pattern of performance across various cognitive domains. Interestingly, a number of these domains have a temporal dimension and require individuals to build and compute representations of duration, chronology or rhythm. For example, in the domain of memory, evidence points to episodic memory difficulties (e.g., Bowler, Gardiner, Grice, & Saavalainen, 2000) including difficulties remembering the temporal order of events (Gaigg et al., 2013; Poirier et al., 2011). In the domain of perception, evidence shows increased low-level processing of perceptual details in ASD, for example autistic individuals have been found to be less susceptible to global changes in pitch and timing when comparing local features in musical sequences (Foxton et al., 2003). In contrast, they present a diminished tendency to integrate information globally, illustrated by a lower temporal threshold in visual simultaneity judgments (Falter et al., 2012) . In the motor domain, evidence suggests issues with coordination in ASD (Fournier, Hass, Naik, Lodha, & Cauraugh, 2010) and apraxia is commonly associated with the disorder (Mostofsky et al., 2006). Difficulties in motor coordination impinge upon communication abilities: ASD individuals do not coordinate gestures with speech as tightly as in typically developing (TD) individuals (de Marchena & Eigsti, 2010). All these functions require the ability to process time, either by accurately representing interval durations (e.g., to produce a vowel of appropriate length in speech, or evaluate temporal distance between events in memory) or by synchronising events (e.g., integration of information across sensory modalities; interpersonal coordination). Indeed, Allman (2011) proposed abnormal timing as a unifying theory of autism, where the impairment of time processing early in development could, in the long term, result in behavioural atypicalities. Until recently, despite some indications that time processing might be affected in ASD (Boucher, 2001; Wimpory et al., 1995), little research had been conducted to explicitly address this issue. Recently, however, an upsurge of interest has emerged in temporal processing in ASD. The next section will review the literature on time perception in ASD, starting at the sub-second to few-second range of durations, tapping into the domain of interval timing. Studies using psychophysical paradigms are listed in Table 2.1.

2.1.1 Literature review: time perception in ASD

2.1.1.1 Sub-second range of durations

In the sub-second range of durations, Gowen and Miall (2005) asked adults with ASD to either synchronise finger taps with repeating tones, or to continue a series of tones by producing more tones at the same pace (using inter-tone intervals of 400 to 800ms). While TD individuals found the continuation task more difficult than the synchronisation task (with a tendency to anticipate the tones), individuals with ASD found both equally difficult and overall more difficult than the TD group, with a greater tendency to tap too early. In a similar range of durations, Falter et al. (2012) used a temporal generalisation task to assess adults with ASD, in auditory, visual and cross-modality conditions. Participants were presented first with a standard duration (600 or 1000ms), followed by a probe duration (300 to 900ms or 500 to 1500ms) and needed to decide whether the standard and probe durations were identical or different. In all modalities ASD individuals were less sensitive to small temporal differences compared to TDs, showing a greater tendency to judge non-identical probes as equal to the standard (i.e., generate false alarms) without any increase in hit rate for identical probes. Also using a generalisation task with an audiovisual 500ms standard duration and probes varying between 125-875ms, Brodeur, Gordon Green, Flores, and Burack (2014) replicated the finding of decreased sensitivity in a sample of autistic children with complex needs. Moreover, they extended this result to a bisection task in which similar groups of lower-functioning ASD children and mental age-matched TD children compared varying durations to a “short” (200ms) and a “long” (800ms) audiovisual anchor. Karaminis et al. (2016) used a combination of two paradigms as well and found lower accuracy and precision in ASD children in visual reproduction and bisection tasks. Finally, using a comparison task in which participants judged which of two auditory tones was longer than a standard 400ms tone, Kargas, López, Reddy, and Morris (2014) found a higher temporal discrimination threshold in ASD compared to TD adults. Together, these studies suggest that autistic individuals’ temporal perception is characterised by reduced precision in judgements. However, in the same range of durations, neither Jones et al. (2009) or Mostofsky, Goldberg, Landa, & Denckla (2000) using a comparison task, or Jones, Lambrechts, and Gaigg (2017) using a bisection task found any difference between TD and ASD groups. Notably, the comparison task used by Jones et al. and Mostofsky et al. (two independent research groups) should be well suited for revealing basic differences in sensory timing given the straightforward relationship between the slope of the resultant psychometric function and timing precision, and the negligible reliance on long-term memory. Another two studies are worth mentioning in which the authors did not contrast two diagnosis groups, but instead investigated the relation between autistic traits and task performance in the general adult population. Using an auditory temporal comparison task with intervals between 300-600ms, Stewart, Griffiths, and Grube (2015) found a negative correlation between participants' Autism Quotient score (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001b) and their duration comparison thresholds. In other words, participants with a higher AQ (or higher autistic traits) were more proficient in the temporal comparison task. In contrast, using a visual bisection task with intervals between 400-1600ms, Jones et al. (2017) found no correlation between task performance and AQ scores.

Study	Date	N	Diag	Age Mean (SD), range	FIQ Mean (SD), range	Paradigm	Mod	Durations	Group comp	p
Brodeur et al.	2014	n_{ASD} = 15 (3F) n_{TD} = 15 (4F)	Autism (health authorities)	ASD: 10.16 (3.93) - MA: 6.19 (1.28) TD: 6.61 (0.78) - MA: 6.22 (1.40)		Bisection	Aud	200-800ms	pLong: No main effect of group Group x duration	 > .05 < .01
Allman et al.	2011	n_{ASD} = 13 (0F) n_{TD} = 12 (3F)	11 Autism, 2 ASD (ADOS, ADI-R)	ASD: 10.3 (2.4), 7.3-15.2 TD: 10.3 (3.1), 7.3-16.8	ASD: 92.31 (17.13), 72–118 TD: 109.80 (18.14), 78–122 *(avail. for 3 TD only)*	Bisection	Vis	1-4s 2-8s	BPs: ASD < TD $WR_{1\text{-}4s}$: no diff $WR_{2\text{-}8s}$: ASD > TD (ASD less sensitive)	< .02 > .25 = .05
Gil et al.	2012	n_{ASD} = 12 n_{TD} = 12	9 AS, 3HFA/PDD-NOS (health authorities)	ASD: 13 (2.49), 9-17 TD: 13.21 (2.32), 8-16	ASD: 94.37 (22.39) TD: 101.45 (19.49)	Bisection	Vis	0.5-1s 1.25-2.5s 3.13-6.25s 7.81-16.63s	BP, DL, WR: No main effect of group	 > .05
Jones et al.*	2017	n_{TD} = 85 (47F)	Broader phenotype (AQ)	TD: 22:7 (4.94)		Bisection	Vis	400-1600ms	No correlation between AQ score and either PSE or WR	NA
Jones et al.	2017	n_{ASD} = 20 (3F) n_{TD} = 26 (6F)	ASD (ADOS)	ASD: 45:4 (12.6) TD: 44:0 (11.6)	ASD: 114.6 (16.6) TD: 108.1 (15.3)	Bisection	Vis	400-1600ms	PSE, WR: No main effect of group	> .10
Jones et al.	2009	n_{ASD} = 71 (6F) n_{TD} = 48 (2F) (22 SEN)	39 childhood autism, 33 ASD (ADOS, ADI-R)	ASD: 15:6 (5.7m) TD: 15:6 (5.9m)	ASD: 89.33 (21.53) TD: 87.79 (17.32)	Comparison (adaptive)	Aud	Std 640ms Start probe 400ms Increment 8ms	No main effect of group	> .20
Kargas et al.	2015	n_{ASD} = 21 (3F) n_{TD} = 21 (3F)	ASD (ADOS)	ASD: 30:4 (10.4m) TD: 29:4 (11.4m)	ASD: 109.5 (18.3) TD: 115.9 (10.6)	Comparison (adaptive)	Aud	Std 400ms Start probe 600ms Increment 5ms	Threshold: ASD > TD	< .005
Mostofsky et al.	2000	n_{ASD} = 11 (5F) n_{TD} = 17 (11F)	Autism (ADOS, ADI)	ASD: 13.3, 6.8-17.8 TD: 12.5, 8.3-16.7	ASD: 101, 81-132 TD: 105, 80-133 *(not avail. for 5 TD)*	Comparison	Aud	Centred on 550ms	Accuracy score: no main effect of group	> .30

Stewart et al.*	2015	n_{TD} = 24 (12F)	Broader phenotype (AQ)	TD: 22.3 (3.9)		Comparison	Aud	300-600ms	Negative correlation between AQ score and duration comparison threshold	< .05
Karaminis et al.	2016	n_{ASD} = 23 (6F) n_{TD} = 23 (10F)	Autism, AS (ADOS)	ASD: 12:4, 7-14 TD: 11:8, 7-13	ASD: 100.30 (15.720), 7-128 TD: NA	Comparison	Vis	200-1200ms	WF: ASD > TD	< .005
Gowen & Miall	2005	n_{ASD} = 12 (4F) n_{TD} = 12 (4F)	AS	ASD: 27.42 (11.08),18-49 TD: 28.17 (11.70), 18-50	ASD: 104 (22.08), 76-135 TD: 112.42 (15.92), 84-130	Continuation	Aud	400-800ms	No main effect of group Std dev: ASD > TD	> .10 ≤ .05
Sheridan & McAuley	1997	$n_{ASD\ 7\text{-}10}$ = 11 $n_{ASD\ 12\text{-}15}$ = 3 $n_{ASD\ low\ IQ}$ = 4 $n_{TD\ 7\text{-}10}$ = 6 $n_{TD\ 12\text{-}15}$ = 5 $n_{TD\ adults}$= 18	ASD	7-15	$ASD_{7\text{-}10}$: 100.0(14.1) $ASD_{12\text{-}15}$: 96 (22.6) $ASD_{low\ IQ}$: 76.8 (6.3) $TD_{7\text{-}10}$:119.3 (7.9) $TD_{12\text{-}15}$: 113.6 (6.6) NA	Continuation	Aud	600ms	Mean IRI: $ASD_{7\text{-}10}$ > $TD_{7\text{-}10}$	< .05
Wallace & Happe	2008	n_{ASD} = 25 (0F) n_{TD} = 25 (2F)	Autism, AS (health authorities)	ASD: 14.10 (1.94) TD: 13.84 (2.16)	ASD: 96.36 (22.07) TD: 100.08 (16.04)	Estimation	Aud (verbal)	2, 4, 12, 15, 45s	Ratio scores: No main effect of group	> .60
Falter et al.	2012	n_{ASD} = 18 (1F) n_{TD} = 19 (4F)	ASD (ADOS, ADI-R)	ASD: 25:3 (8:1), 16:9-42:11 TD: 26:1 (6:11), 14:10-38:6	ASD: 112 (13), 88-131 TD: 113 (8), 100-133	Generalisation	AA, VV, AV, VA	300-900ms 500-1500ms	Skew: No main effect of group Scale x group Modality x scale x group	> .10 < .05 < .05
Brodeur et al.	2014	n_{ASD} = 15 (4F) n_{TD} = 15 (5F)	Autism (health authorities)	ASD: 10.74 (3.04) - MA: 7.30 (1.64) TD: 6.46 (0.93) - MA: 6.48 (1.34)		Generalisation	AV	125-875ms	pSame: No main effect of group Group x duration	> .05 < .01
Wallace & Happe	2008	n_{ASD} = 25 (0F) n_{TD} = 25 (2F)	Autism, AS (health authorities)	ASD: 14.10 (1.94) TD: 13.84 (2.16)	ASD: 96.36 (22.07) TD: 100.08 (16.04)	Production	Aud (verbal)	2, 4, 12, 15, 45s	Ratio scores: No main effect of group	> .70

Martin et al.	2010	n_{ASD} = 20 (5F) n_{TD} = 20 (7F)	ASD (ADOS)	ASD: 36 (13.4) TD: 35 (10.8)	ASD: 106 (17.3) TD: 108 (16.4)	Reproduction	Aud	0.5-4.1s	Abs diff: ASD > TD MJR: no main effect of group CoV: ASD > TD	< .01 NA < .05
Wallace & Happe	2008	n_{ASD} = 25 (0F) n_{TD} = 25 (2F)	Autism, AS (health authorities)	ASD: 14.10 (1.94) TD: 13.84 (2.16)	ASD: 96.36 (22.07) TD: 100.08 (16.04)	Reproduction	Aud (verbal)	2, 4, 12, 15, 45s	Ratio scores: No main effect of group	 > .40
Szelag et al.	2004	n_{ASD} = 7 (3F) n_{TD} = 7 (3F)	HFA (Checklist for Autistic Children)	ASD: 12:6, 9-16 Matched ±1 year	ASD: 82-102 TD: 95-145	Reproduction	Aud, Vis	1-5.5s 0.5-3s	MDJR: Main effect of group Group x standard MCV: ASD > TD Main effect of group Group x standard Group x standard x modality MCV: ASD > TD	 < .01 <.0001 < 005 < .01 < .001 < .05 < .05
Brenner et al.	2015	n_{ASD} = 27 (4F) n_{TD} = 25 (3F)	ASD (ADOS, ADI-R)	ASD: 12.68 (2.85) TD: 13.41 (2.32)	ASD: 101.31 (11.24) TD: 106.96 (11.46)	Reproduction	Vis	4, 8, 12, 16, 20s	Accuracy: ASD < TD Consistency: ASD < TD	< .05 < .005
Karaminis et al.	2016	n_{ASD} = 23 (6F) n_{TD} = 23 (10F)	Autism, AS (ADOS)	ASD: 12:4, 7-14 TD: 11:8, 7-13	ASD: 100.30 (15.720), 7-128 TD: NA	Reproduction	Vis	1006-1536ms 1270-1800ms	BIAS: ASD > TD CoV: ASD > TD	< .001 < .001
Maister & Plaisted-Grant	2011	n_{ASD} = 21 (1F) n_{TD} = 21 (8F)	Autism (ADI-R)	ASD: 11.3 (1.5) TD: 10.7 (0.8)		Reproduction	Vis	0.5, 1, 2, 4, 10, 30, 45s	Error scores: 0.5, 1, 2, 45s: ASD > TD 4, 10, 30s: no main effect of group	 < .05 > .05
Maister & Plaisted-Grant	2011	n_{ASD} = 15 (0F) n_{TD} = 15 (4F)	Autism (ADI-R)	ASD: 11.8 (1.5) TD: 11.2 (1.2)		Reproduction	Vis	0.5, 1, 2, 4, 10, 30, 45s	Error scores: 0.5, 45s: ASD > TD 1, 2, 4, 10, 30s: no main effect of group	 < .05 > .05

Gowen & Miall	2005	n_{ASD} = 12 (4F) n_{TD} = 12	AS	ASD: 27.42 (11.08),18-49 TD: 28.17 (11.70), 18-50	ASD: 104 (22.08), 76-135 TD: 112.42 (15.92), 84-130	Synchronisation	Aud	400-800ms	Abs error: ASD > TD S-R asynch: ASD > TD Std dev: ASD > TD	< .005 < .05 < .05
Sheridan & McAuley	1997	$n_{ASD\ 7\text{-}10}$ = 11 $n_{ASD\ 12\text{-}15}$ = 3 $n_{ASD\ low\ IQ}$ = 4 $n_{TD\ 7\text{-}10}$ = 6 $n_{TD\ 12\text{-}15}$ = 5 $n_{TD\ adults}$= 18	ASD	7-15	$ASD_{7\text{-}10}$: 100.0(14.1) $ASD_{12\text{-}15}$: 96 (22.6) $ASD_{low\ IQ}$: 76.8 (6.3) $TD_{7\text{-}10}$:119.3 (7.9) $TD_{12\text{-}15}$: 113.6 (6.6) NA	Synchronisation	Aud	600ms	IRI variability: ASD > TD	NA

Table 2.1 Compilation of empirical studies investigating time perception and timing in ASD using psychophysical tasks.
Studies highlighted in light grey found at least one measure of temporal processing that differed significantly between ASD and TD groups. MDJR = Mean Duration Judgement Ratio. MCV = Mean Coefficient of Variation. *Stewart et al. (2015) & Jones et al. (2017): studies do not include a group comparison (ASD vs TD) but look at relation between Autism Quotient score and time perception performance.

2.1.1.2 Multi-second range of durations

A similar picture emerges in studies that have examined multi-second duration ranges. Allman, DeLeon, and Wearden (2011) used a bisection task in the visual modality with children. Participants learned to recognise two anchor durations (“short”: 1 or 2s and “long”: 4 or 8s) and were then presented with varying probes (1 to 4s in step of 0.5s or 2 to 8s in steps of 1s) with instructions to judge which anchor they most resembled. Results showed that ASD children had a tendency to overestimate probe durations in comparison to the TD children, especially for durations between 3.5 and 5s. They also showed reduced sensitivity in the 2-to-8s range. A subsequent study by Gil, Chambres, Hyvert, Fanget, and Droit-Volet (2012), however, found no differences between TD and ASD children in a bisection task using durations ranging from 0.5 to 16.62s. In reproduction tasks, the evidence is somewhat more consistent. Szelag, Kowalska, and Galkowski (2004) asked TD and ASD children to reproduce durations in the visual or auditory modality by pressing a key. They found that children with ASD were very imprecise, reproducing every interval between 1 and 5.5s as 3s on average, with large variability in reproduction times. Martin, Poirier, and Bowler (2010) also used a reproduction task in the auditory modality, but with adults. They found that ASD individuals were both less accurate (i.e., reproduced intervals further from the standard on average) and less precise (i.e., responses were more variable) than TD adults when reproducing intervals between 0.5-4.1s. Maister and Plaisted-Grant (2011) found similar results in an equivalent reproduction study where children when asked to reproduce temporal intervals of 0.5-45s presented visually. They reported that autistic children showed larger error scores for extreme durations (0.5, 1, 2 and 45s) compared to the TD group.

2.1.1.3 Atypical mechanisms of time perception in ASD

On the whole, a majority of the studies described above, which employ psychophysical methods to examine time processing of sub-second and multi-second durations, tend to suggest that autistic individuals show reduced precision in their temporal judgements. In contrast, accuracy data is somewhat less consistent, and the reasons for discrepant findings in this context are unclear. One explanation that has been offered for the difficulties individuals with ASD have with relatively long durations is the atypical involvement of memory and attention. Maister and Plaisted-Grant (2011) provided some evidence that typical time perception of long durations (30-45s) is underpinned by long-term memory, since TD participants' reproduction error scores for long durations correlated with free recall reorganisation scores in their TD group. In contrast, the ASD group showed no correlation between performance on long-duration reproduction tasks and episodic memory measures, suggesting that they did not spontaneously use their episodic memory to perform the reproduction task. Accordingly, interval reproduction studies, which typically require participants to encode and maintain longer intervals in memory, consistently report both reduced accuracy and sensitivity in the autistic group, with the exception of Wallace and Happe (2008) who use manual timing and as such might lack the precision to measure fine group differences. The authors also argued that increased variability in shorter durations (1-10s) in the ASD group is underpinned by attentional difficulties. Indeed they observed that intra-individual variability in reproduction performance is higher in the ASD group, and that this correlated with individual accuracy scores, which is believed to reflect poorer attentional skills (Brown, 1997; Pouthas & Perbal, 2004). In addition, both Allman et al. (2011) and Brenner et al. (2015) reported that participants' accuracy was associated with working memory, further indicating that temporal processing, particularly for short durations, might rely on executive functions. In line with this idea, Gil et al. (2012) showed that when participants are appropriately supported (planning and attention) and motivated, autistic children perform similarly to their typically developing counterparts.

Complementing the roles of attention and memory as possible mechanisms to poorer temporal processing over long durations in ASD, Lambrechts, Falter-Wagner, and van Wassenhove (2017), shed some light on the possible underpinnings of reduced sensitivity to short temporal information in autism. The authors recorded magnetoencephalographic (MEG) responses to two consecutive auditory tones (a constant 600ms standard and a varying 300-900ms probe) in a temporal generalisation task. The MEG response evoked by the standard tone in the ASD group showed a delayed offset with a flatter slope as compared to the TD group, indicating that the encoding of a highly repeated and predictable tone was less sharp in ASD adults. Difficulties to encode a predictable duration accurately could account for diminished sensitivity in autism, without disrupting duration processing altogether. Fuzzier boundaries in the encoding of a temporal stimulus would only increase the variability of the duration encoded rather than systematically affecting the accuracy of the temporal estimate, accounting for the pattern of results across studies outline above, which demonstrates inconsistent group differences in temporal accuracy, but the fairly consensual finding of diminished sensitivity in autism. Lambrechts and colleagues also found that, compared to typical adults who tend to allocate neural resources differently to either task depending on instructions, autistic adults seem to allocate less neural resources to duration processing regardless of the instructions (pitch or duration discrimination), which suggests that autistic individuals engaged limited attentional resources in the task. This resonates with Maister and Plaisted-Grant's suggestion that difficulties in the short range of durations in ASD are underpinned by attentional factors, in this case the ability to direct attention to a specific feature of the stimulus.

Whilst variable task demands on attention, working memory and episodic memory can account for some of the discrepancies observed between studies, notable inconsistencies remain in the literature concerning temporal processing in ASD, particularly for bisection and comparison tasks. One source of such inconsistency might lie in the use of different modalities, given that sensory modality is a known main factor modulating temporal processing in the typical population, and that autistic individuals present with an atypical sensory processing profile. In the literature presented, the effect of modality is partly confounded with the use of different tasks (4/5 bisection tasks use visual stimuli, whilst 4/5 comparison tasks use auditory stimuli). Only two studies (Falter et al., 2012; Szelag et al., 2004), using generalisation and reproduction paradigms respectively, directly compared performance across modalities. Their results point to a pattern of diminished sensitivity in the autistic group when the stimulus is encoded in the auditory as compared to the visual modality, although non significantly so in Szelag and colleagues' work. This pattern contrasts with the results typically reported in TD time perception tasks, that both accuracy and precision are greater in the auditory compared to the visual modality (Goldstone & Lhamon, 1974; Wearden et al., 1998b).

2.1.2 Aim and predictions

The first experiment in this thesis sought to resolve some of the inconsistencies presented by studies of the perception of durations in ASD, particularly at the sub-second range of durations which is thought to be most critical to social-communicative behaviours. For this purpose a comparison paradigm was chosen as this is thought to be most sensitive to the precision of temporal judgments without relying heavily on memory resources. In order to determine whether sensory modality is driving some of the discrepancies in the literature, time perception was examined across auditory, visual and audiovisual modalities, Based on the finding that temporal performance is typically greater in the auditory modality compared to the visual modality, we predicted that autistic participants would show reduced precision in the visual but not in the auditory or audiovisual modalities. Because differences in accuracy are less consistent in the literature and because we worked with a population of relatively able adults, we did not predict a difference in accuracy between groups for any of the sensory modalities.

2.2 Methods

2.2.1 Participants

23 individuals with a diagnosis of ASD (3 female) and 22 typically developing (TD) individuals (5 female) took part in the study. Participants were recruited through the City, University of London's Autism Research Group participant database, and group-matched according to age and verbal,

performance and full-scale intellectual quotient (VIQ, PIQ, FIQ) as measured by the WAIS-R or WAIS-III UK (The Psychological Corporation 2000) (see Table 2.2). Independent-samples t-tests confirmed that there were no differences between groups for age, VIQ, PIQ or FIQ (all $ps>.1$). All ASD participants had received their diagnosis through the national health service in the UK by experienced clinicians according to DSM-IV criteria, and observations during the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000), as well as responses on the Autism Quotient (Simon Baron-Cohen et al., 2001b) confirmed that all participants experienced difficulties in reciprocal social and communicative behaviours that were commensurate with their diagnosis. Participants in the comparison group had no history of neurological or psychiatric disorders and none scored above the recommended cut-off score of 26 on the AQ (Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005). Two of the ASD participants were taking anti-depressant medication, but exclusion of their responses did not affect the pattern of results; their data were therefore retained. All participants had normal or corrected-to-normal sight and hearing (assessed through an informal interview). Participants provided their written consent in accordance with the Declaration of Helsinki (2008) and the ethics committee of City, University of London. They were paid standard university fees for their participation.

	TD (n=22) mean (SD) range	ASD (n=23) mean (SD) range	Cohen's d
Age (years)	**46.0 (12.8)** 20.3-61.3	**41.1 (12.8)** 24.1-61.8	-0.381
VIQ	**110 (12)** 82-128	**109 (15)** 73-143	-0.048
PIQ	**106 (14)** 75-136	**106 (16)** 73-128	0.039
FIQ	**109 (13)** 77-135	**109 (16)** 70-135	-0.002
AQ	**14.7 (6.4)** 4.0-25.0	**34.7 (6.3)** 25.0-45.0	4.044
ADOS	-	**9.0 (3.1)** 5.0-17.0	
BIS	**65.8 (8.6)** 50-83	**64.9 (7.1)** 54-79	0.114
ZTPI	**176.8 (16.0)** 147-197	**183.7 (13.1)** 160-202	-0.472

Table 2.2 Participant characteristics.
VIQ, PIQ, FIQ = Verbal, Performance and Full-scale Intellectual Quotient (The Psychological Corporation, 2000); AQ = Autism spectrum Quotient (Baron-Cohen, 1995); ADOS = Autism Diagnosis Observation Schedule (total score, Lord et al., 1999); BIS = Barratt Impulsiveness Scale (total score, Patton, Stanford, & Barratt, 1995); ZTPI: Zimbardo Temporal Perspective Inventory (total score, Zimbardo & Boyd, 1999).

2.2.2 Materials

Stimulus presentation was controlled by E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The visual stimulus was a 10x10cm light-grey square presented centrally on a black background on a 15" LCD monitor with a refresh rate of 59.9Hz. The auditory stimulus was a 440Hz sinusoidal pure tone presented binaurally through headphones (or bilateral speakers when requested by participants). In the audiovisual condition both stimuli were presented synchronously. Stimulus intensities were well above detection threshold and participants were given the opportunity to set the volume of the auditory stimulus to a comfortable level. There were two standard durations (800 and 1200ms) and probe durations were defined as ±5, 10, 25 and 50% of the standard durations (or as near to these values as achievable, given monitor refresh rate, for visual stimuli). Two ranges of durations were thus defined: a range of *short durations* (400-1200ms) and a range of *long durations* (600-1800ms). Accuracy of stimulus timing was checked in each modality using a 20MHz storage oscilloscope (Gould 214 DSO 1604).

2.2.3 Procedure

Participants were seated comfortably 60cm away from the screen in a dimly lit room. We used a duration comparison task. A trial started with the presentation of a probe, followed by an empty (uniform random) interval of 200 to 600ms, and finally the presentation of a standard. A prompt appeared on the screen until the participant indicated which stimulus lasted longer by pressing one of two keys ("n" or "j" on a qwerty keyboard). The inter-trial interval lasted 400 to 800ms (see Figure 2.1). There were 20 trials for each probe-standard combination, resulting in a total of 960 trials (3 modalities x 2 ranges x 8 probes x 20 measures). Participants first completed a short training phase of 4 trials per sensory modality, which was repeated when necessary until the procedure was clear. The test phase followed, in which trials were blocked, with 80 trials per block, and 4 blocks per sensory modality (12 blocks overall). Participants performed a single block in each modality in pseudo-random order before taking a break, with this sequence repeated four times. The standard was consistently presented in the second position to allow us to quantify time-order error effects but trials from both duration ranges were intermixed within blocks so that participants would not identify the second stimulus as a standard. Participants performed the experiment in one session, except for 2 TD and 4 ASD participants who completed the task over two sessions.

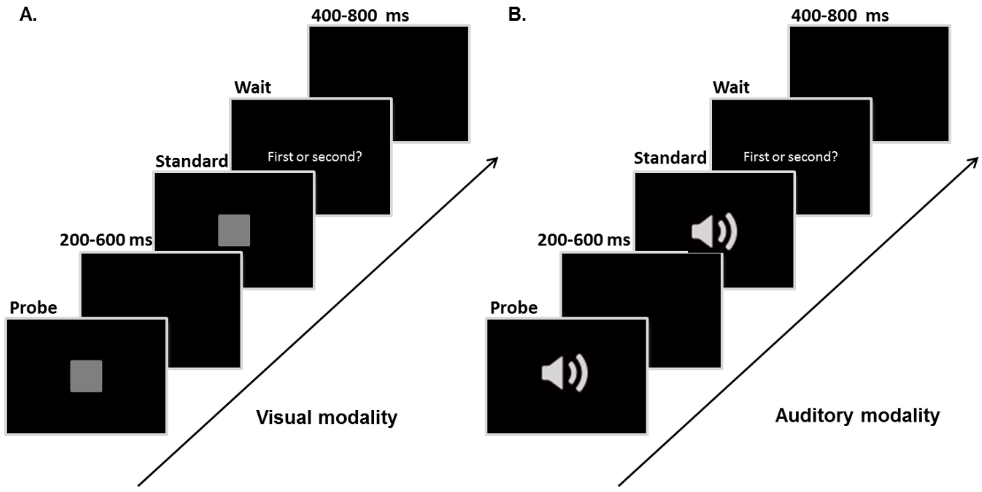

Figure 2.1 Experimental paradigm in the visual (A) and auditory (B) modality.
In the audiovisual modality the design is the same but the visual and the auditory stimulus are presented synchronously. In any modality, a trial started with the presentation of the standard, followed by a 200-600 ms inter-stimulus interval. Participants were then prompted for a non-speeded response about the relative duration of the stimuli. Inter-trial intervals lasted 400-800ms.

2.2.4 Analysis

2.2.4.1 Stage 1: Verification of above-chance performance

Data were first processed individually. For each participant the percentage of response 'probe longer than standard' was computed as a function of probe duration, separately for each standard duration and modality. In this computation, trials for which reaction time was outside ±2 standard deviations from the mean individual reaction time were excluded, as they were likely to reflect some distraction or failure of vigilance. 4.0% and 4.1% of the data were rejected at this stage in the TD and ASD groups respectively. Data were maximum-likelihood fitted against a one-parameter model (a horizontal line) and a two-parameter model (a psychometrical profile, specifically a cumulative Gaussian) using Matlab (The MathWorks, Inc.). The deviance of each model from the data points was computed, and their difference was compared to a chi-square distribution with one degree of freedom. This allowed us to evaluate whether the 2-parameter fit was significantly better than the 1-parameter fit (Wichmann & Hill, 2001b), in other words whether participant performances were following the expected psychometric profile, or were indistinguishable from random guessing. Participants for whom the 2-parameter fit was not significantly better than the 1-parameter fit in at least one condition in a sensory modality had their data excluded for that modality. This resulted in the exclusion of a large number of participants in the ASD group in the visual modality in particular (see results section).

2.2.4.2 Stage 2: Derived measures for above-chance participants

For participants included in the full analysis, the percentages of response 'probe longer than standard' were again fitted to a cumulative Gaussian function, this time using the Psignifit toolbox version 3.0.10 for Matlab (http://bootstrap-software.org/psignifit/) which provides a maximum likelihood fitting procedure and estimates for confidence intervals from bootstrapping techniques based on Wichmann and Hill (2001a, 2001b). All fits were performed using the Matlab *pfit* command with 999 bootstrap runs. This allowed us to compute the Point of Subjective Equality (PSE), i.e., the probe duration perceived as equal to the standard, and the Weber Ratio (WR), which reflects the sensitivity of the temporal judgment (Allan & Gibbon, 1991; Droit-Volet, Meck, & Penney, 2007; Roitman, Brannon, Andrews, & Platt, 2007). The PSE is computed as the probe duration which leads to 50% of 'probe longer than standard' responses. The Weber Ratio (WR) is computed as the Difference Limen (DL) which is half the distance between the probe durations that support 25% and 75% of 'probe longer than standard' responses, normalized by the PSE. WR is a positive, decreasing index of precision: the closer the WR to 0, the greater the sensitivity of the respondent. Shapiro-Wilk tests revealed some violations of the normality assumption for PSE, WR and RT data. Levene's test of homogeneity revealed that variance was not equal between groups in two conditions for PSE data only. To correct this, either a log or inverse transformation was applied to the data where appropriate. Analyses were first conducted with the untransformed data and then a second time with the transformed data. The pattern of findings remained the same in all analyses. Because transformed data make it difficult to interpret results, and because ANOVAs are considered to be fairly robust to deviations from normality (e.g., Schminder, Ziegler, Danay, Beyer, & Bühner, 2010) findings from the original (untransformed) data are presented below. Further detail on the violations of assumptions and the transformations applied are provided in Appendix 1.

2.3 Results

As noted above, the first step of our analysis established whether participants performed above-chance. Table 2.3 reports the number of participants who were excluded in each modality. Descriptive statistics indicate that overall a larger number of participants were excluded in the ASD than in the TD group. In particular, over 40% of the ASD participants (10 out of 23) were excluded in the visual modality as compared to less than 15% in the TD group. In the auditory and audiovisual modalities around 20% of the ASD participants but less than 10% of the TD participants were excluded. Chi-square tests for each modality indicated that the number of excluded participants differed significantly between groups in the visual modality only (χ^2=4.87, p<.05). The absence of a significant effect in the auditory and audiovisual modalities likely reflects lack of power for these comparisons. Exclusion from the analysis indicated an inability to reliably discriminate between the standard and probe durations, including probes lasting ±50% of the standard (i.e., 400 vs 800ms, 800 vs 1200ms, 600 vs 1200ms and 1200 vs 1800ms).

	TD (n=22)	ASD (n=23)
Auditory	**2**	**5**
	9.1%	*21.7%*
Visual	**3**	**10**
	13.6%	*43.5%*
Audiovisual	**1**	**4**
	4.5%	*17.4%*
Total	**4**	**11**
	18.2%	*47.8*

Table 2.3 Number of participants excluded in each modality.

Excluding data in conditions where participants did not perform above-chance altered the extent to which ASD and TD participants were matched (the ASD group had marginally higher FIQ scores than the TD group, t(30,28)=1.869, p=.072). However, since an analysis of closely matched subgroups of 12 TD and 12 ASD individuals yielded the same pattern of results as that reported below, all available data were retained for the analyses and are reported hereafter.

Because participants could be excluded selectively from one sensory modality, we conducted a 2 (short durations, long durations) x 2 (TD, ASD group) mixed-design Analysis of Variance (ANOVA) on PSE and WR measures for each sensory modality separately. Figure 2.2 illustrates the psychometric profiles of responses and Figure 2.3 presents derived PSE and WR values based on individual fits.

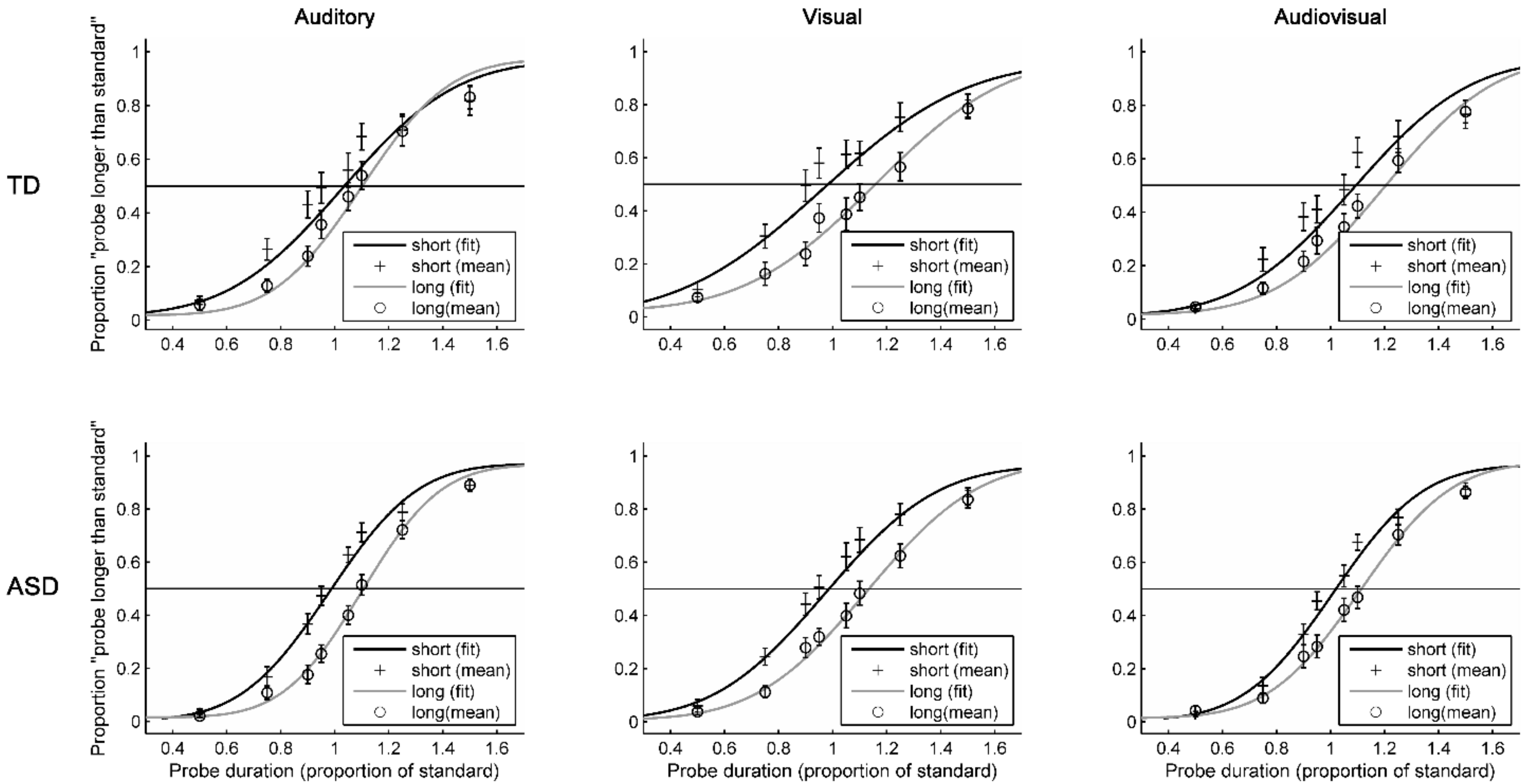

Figure 2.2 Response profiles for the TD and ASD groups, in the auditory, visual and audiovisual modalities.
Discrete points show average responses across included participants for short durations (+) and long durations (o). Lines show the average fit obtained for all included participants for short durations (black) and long durations (grey). Error bars show standard error of the mean.

2.3.1 Point of Subjective Equality

The point of subjective equality represents the proportional duration of the probe (relative to the standard) when the two were judged equal. Values above 1 indicate a tendency to perceive the probe as shorter than the standard when they were objectively equal, and vice versa.

In all three modalities, the ANOVA conducted on PSE measures revealed a main effect of range (Auditory: $F(38,1)=14.049$, $p<.005$, $\eta_p^2=.281$; Visual: $F(32,1)=28.996$, $p<.001$, $\eta_p^2=.491$; Audiovisual: $F(40,1)=18.270$, $p<.001$, $\eta_p^2=.325$) indicating that PSE was smaller for short durations than for long durations, i.e., the duration of the probe was overestimated for short durations and underestimated for long durations across groups. No main effect or interaction with the factor group was found in any of the modalities, providing no support for the idea that individuals with and without ASD show differences in accuracy in any of the sensory modalities.

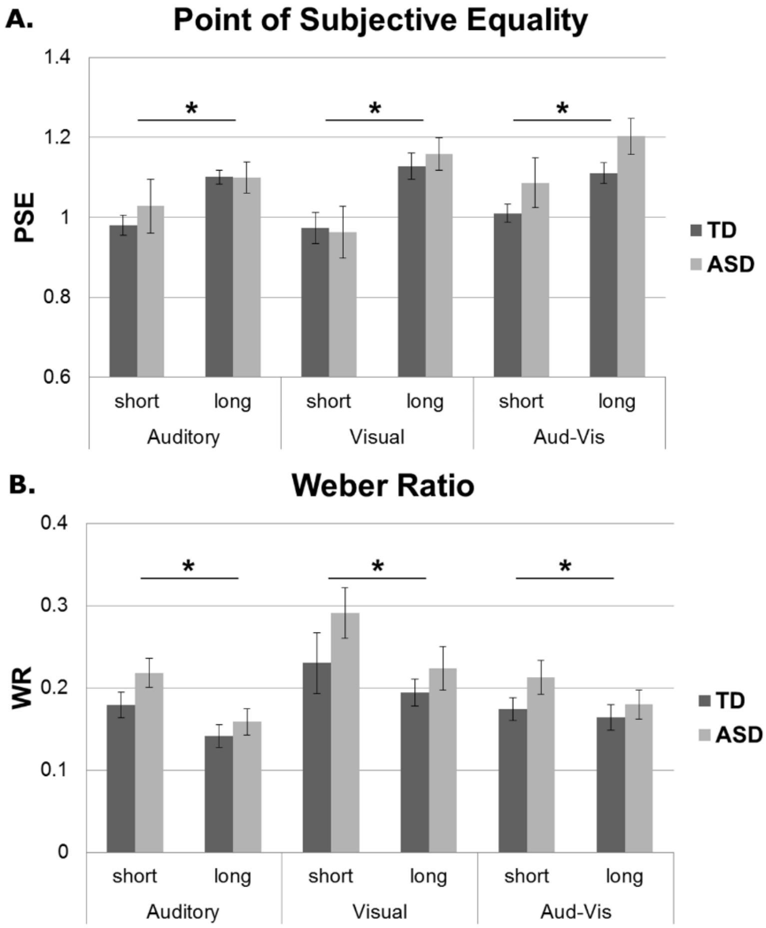

Figure 2.3 Mean Point of Subjective Equality (A) and Weber Ratio (B) in the auditory, visual and audiovisual modality for the TD and ASD groups. Both analyses revealed a main effect of duration range but no effect or interaction involving the factor group. Error bars show standard error of the mean.

2.3.2 Weber Ratio

The Weber ratio is a threshold measure that captures the proportional change in the probe that is required to reliably discriminate it from the standard. The scalar property (i.e., Weber's law for time) predicts that it will be constant for different standard durations. However, in the *auditory* and *visual modality*, the ANOVAs conducted on WR measures revealed a main effect of range (Auditory: $F(38,1)=20.896$, $p<.001$, $\eta_p^2=.367$; Visual: $F(32,1)=4.848$, $p<.05$, $\eta_p^2=.139$) indicating that WR was higher for short durations than for long durations, i.e., participants showed reduced *normalised* precision to discriminate between short durations as compared to long durations. Similarly in the *audiovisual modality* the ANOVA showed a marginal effect of range ($F(40,1)=3.829$, $p=.058$, $\eta_p^2=.092$) indicating a trend towards higher WR for short durations than for long durations. No main effect or interaction with the factor group was found in any of the modalities, showing no support for the hypothesis that individuals with and without ASD performed with a different degree of precision in any sensory modality.

2.3.3 Reaction Times

A 2 (short durations, long durations) x 8 (probe durations: ±0.5, 0.75, 0.9, 0.95 x standard duration) x 2 (TD, ASD group) mixed-design ANOVA on Reaction Times (RT) was conducted for each sensory modality separately. Results were Greenhouse-Geisser corrected where appropriate. Figure 2.4 presents RT in each modality. In all three modalities the ANOVA revealed a main effect of range indicating that participants took longer to respond in short-duration than in long-duration trials. Importantly, a main effect of probe duration was also found, showing that probe durations further away from the standard were responded to faster than probe durations closer to the standard, also known as the distance effect. This pattern presented a leftward skew (the shortest probe durations were responded to faster than the longest probe durations). In addition, in the auditory and audiovisual modalities, a range x probe duration interaction was found, showing that the distance effect was stronger in short-duration than in long-duration trials. Notably, no main effect or interaction with the factor group was significant in any of the modalities, suggesting that both groups responded with comparable speed and demonstrated a similar distance effect in all modalities.

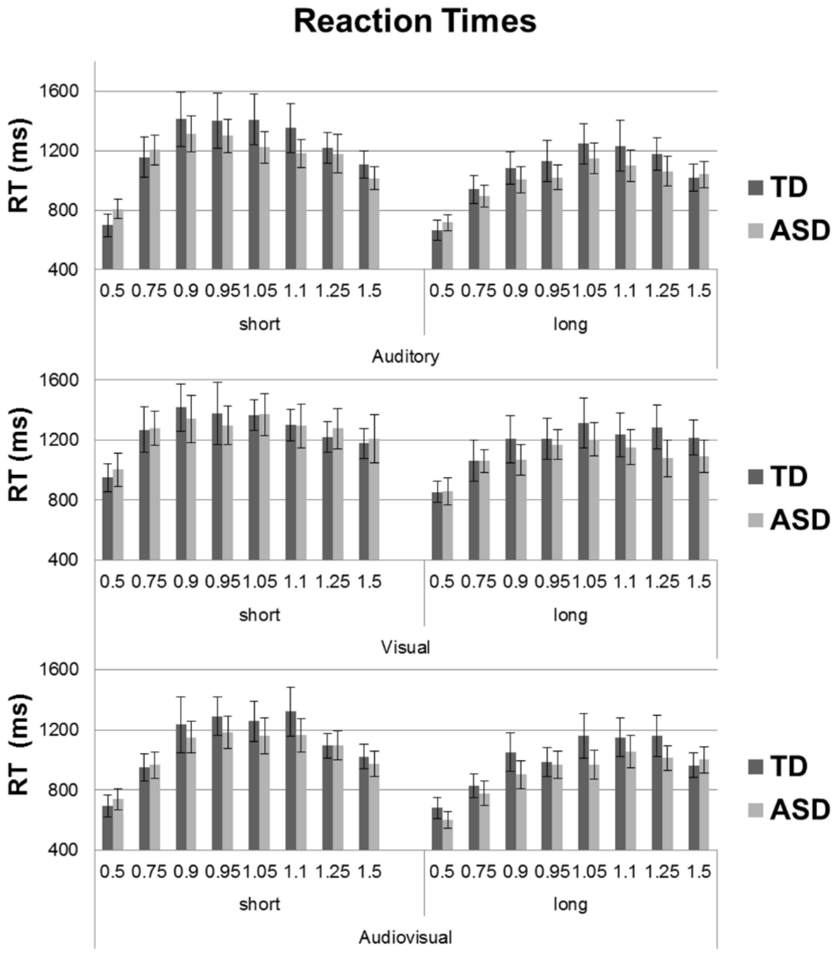

Figure 2.4 Mean Reaction Times in the auditory, visual and audiovisual modality for the TD and ASD groups in response to the range of short durations and long durations.
The analysis revealed a main effect of probe duration (distance effect) which appears clearly on the graph as well as an interaction between probe duration and duration range (flatter profiles in the range of long than in the range of short durations). No effect or interaction involving the factor group was found. Error bars show standard error of the mean.

2.3.4 Examining the excluded individuals

Before we discuss the results of the comparison task, it is necessary to take note of the unusually high number of excluded participants in the analysis, in particular in the ASD group in the visual condition. Finding and excluding participants who fail to perform the task above chance is fairly typical (and not always reported) in time perception studies, in which it is common to find that part of the population shows difficulties with explicit temporal judgements. However the number of participants excluded from at least one condition are particularly high in the present study. This is not unprecedented, although not often commented on: for instance Allman et al. (2011) excluded data from 6/19 (32%) ASD children but only 1/12 (8%) TD children in their temporal bisection task, Karaminis et al. (2016) excluded 6/29 (21%) ASD children and 8/31 (26%) TD children in reproduction and comparison tasks and Gil et al. (2012) excluded 5/17 (29%) ASD children in a temporal bisection task. Our inclusion criterion was that the data should be better fitted to a two-parameter model than a one-parameter model, in other words that the pattern of response should show some distinction between the different values of the probe. This is the condition to meet for the model, and therefore the PSE and WR measures, to be valid. Other studies using a comparison task have avoided this issue by using an adaptive paradigm, in which the probe duration is incremented until the performance reaches a certain threshold (Jones, Happé, Baird, Simonoff, Marsden, et al., 2009; Kargas et al., 2014). This procedure makes it possible to identify each individual's performance level, but does not constrain the range of durations to a definite window. In the current study we were specifically interested in a range of durations relevant for the timing of everyday conversation, and therefore constrained the task to use fixed stimuli. This leads us to two questions of interest: (1) what factors underlie difficulties with temporal processing and (2) do the same factors underlie difficulties with temporal processing in ASD and TD individuals?

Our participant database characterised participants on a range of measures, including the ADOS sub-scores and main score and the Autism Quotient (Simon Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001a). In order to better understand what differed between included and excluded individuals we examined age, IQ, ADOS and AQ scores. In addition, we formulated the post-hoc hypothesis that individuals who failed to perform the task might present higher levels of impulsivity and produce rushed responses which impaired their performance (Berlin & Rolls, 2004; Rubia, Halari, Christakou, & Taylor, 2009; Wittmann & Paulus, 2008). After completion of the experiment we were able to obtain impulsivity scores from most participants (using the Barratt Impulsiveness Scale: BIS; Patton, Stanford, & Barratt, 1995) and a score of time orientation measured by the Zimbardo Temporal Perspective Inventory (ZTPI; Zimbardo & Boyd, 1999). We considered "excluded" any participant who had performed at chance level or below in at least one condition.

	Factor	Mean incl.	Mean excl.	t	*p*
All participants	Age	42.3	45.9	-.893	*.377*
$n_{included}$ = 30	**¥FIQ◊**	**113.0**	**100.1**	**2.653**	***.015****
$n_{excluded}$ = 15	**¥VIQ**	**113.1**	**102.5**	**2.274**	***.034****
	PIQ	**110.4**	**97.5**	**3.026**	***.004****
	AQ	**21.9**	**30.4**	**-2.385**	***.022****
	BIS◊	57.0	63.5	-.985	*.331*
	ZTPI◊	157.8	161.7	-.205	*.839*
ASD participants only	Age	39.2	43.2	-.751	*.461*
	¥FIQ◊	**117.4**	**99.18**	**3.176**	***.006****
$n_{included}$ = 12	**VIQ**	**116.5**	**101.4**	**2.704**	***.013****
$n_{excluded}$ = 11	**PIQ**	**115.0**	**97.0**	**3.208**	***.004****
	ADOS◊	7.5	8.2	-.375	*.711*
	AQ	35.5	33.9	.599	*.556*
	BIS◊	45.9	60.4	-1.304	*.207*
	ZTPI◊	129.0	171.6	-1.391	*.180*
TD participants only	Age	44.3	53.3	-1.293	.211
	FIQ◊	110.1	102.8	1.013	.323
$n_{included}$ = 18	VIQ	110.8	105.8	.770	.450
$n_{excluded}$ = 4	PIQ	107.4	98.8	1.153	.263
	AQ	**13.6**	**20.8**	**-2.310**	***.032****
	BIS	64.2	72.3	-1.759	.095
	Attentional	**9.4**	**12.5**	**-2.538**	***.020****
	¥ZTPI	176.4	134.5	.922	.424

Table 2.4 Inspection of individuals excluded in at least one condition.
Outcome of independent t-tests comparing included vs excluded participants (uncorrected). ¥Failed Levene's test for Equality of Variance – Statistic reported for equality of variance not assumed. ◊ No subscores showed a significant difference on their own.

Table 2.4 reports descriptive and independent t-test statistics for all scores. We report uncorrected *p* statistics here for exploration purposes. Results indicated that overall, participants excluded in at least one condition presented lower IQ (FIQ, VIQ and PIQ, all *p*s < .05) and higher AQ score ($p < .05$). Further exploration of the results by diagnosis group shows a different pattern in the TD and ASD group between included and excluded participants. In the ASD group, included and excluded participants only differed in terms of IQ (FIQ, VIQ and PIQ, all *p*s < .05) whereas in the TD group they differed only on AQ ($p < 0.5$) and on the attentional subscore of the BIS ($p < 0.05$). This result has to be taken with caution since the number of excluded TD participants is very limited (n=4). Figure 2.5 illustrates this group difference by showing the distribution of included and excluded participants according to their FIQ and AQ scores.

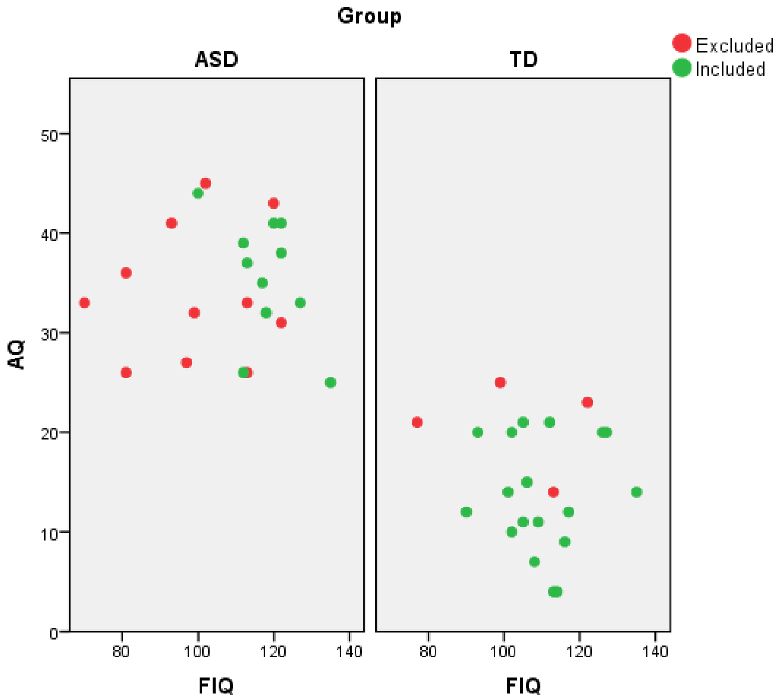

Figure 2.5 Scatterplot of included (green) and excluded (red) participants according to diagnosis group (ASD, TD), AQ and FIQ scores.

2.3.5 Relationship between temporal processing, IQ and AQ: post-hoc correlations

Inspection of excluded participants suggested that IQ (in the ASD group) and AQ (in the TD group) played an important role in temporal processing. We therefore re-examined the performance of included participants. For each group, and for each sensory modality, we computed the Spearman rank-order correlation coefficients between the PSE and WR produced by included participants and their FIQ and AQ scores. To control for family-wise error rate, a Benjamini-Hochberg procedure was applied. None of the correlations were statistically significant, providing no support to the idea that AQ or IQ contribute to temporal processing performance in a monotonic way.

2.3.6 Testing for proactive interference

Proactive interference occurs when material accumulated in working memory during one trial biases the processing of stimuli in the following trial (e.g., Jonides & Nee, 2006). Typically, proactive interference results in slower responses and lower accuracy when two consecutive trials require a different response (incongruent) compared to when they require the same response (congruent). In addition, proactive interference is stronger when the intertrial intervals (ITI) are shorter. In this task, the ITI were short and the design of the task resulted in a mixture of congruent and incongruent trials. However ITIs were pseudo-randomised (400-800ms) which is known to reduce proactive interference. The evidence available suggests that there is no overall differences in the susceptibility to proactive interference in ASD, although there may be a change in susceptibility over the course of the lifespan for autistic people (Lever, Ridderinkhof, Marsman, & Geurts, 2017).

To ascertain that differential proactive interference did not account for our results in this study, following Lever et al., we computed participants' accuracy scores in congruent and incongruent trials for each of the sensory modalities and duration ranges for the two groups separately. Shapiro-Wilk tests and Levene's tests of homogeneity indicated that all variables were distributed normally and that variance was not different between groups. Due to the uneven pattern of inclusion in the different sensory modalities, three separate 2 (ASD, TD groups) x 2 (congruent, incongruent) x 2 (short duration, long duration ranges) mixed-design ANOVAs were conducted on the accuracy scores. No main effect of congruency or congruency x group interaction were found to be statistically significant (all $ps > .1$), suggesting that proactive interference did not likely account for participants' level of performance.

2.4 Discussion

We tested time perception in individuals with and without ASD, in the auditory, visual and audiovisual modalities. For this purpose, pairs of durations were presented – a standard and a probe – and participants had to decide which of the two lasted longer. This procedure allowed us to derive for each individual (1) the mean duration at which the probe was judged equal to the standard (the PSE) and (2) the range-normalised precision with which the comparison was achieved (the WR). We tested two intermixed ranges of durations: 'short' durations (400 to 1200ms) and 'long' durations (600 to 1800ms). We found that ASD participants performed similarly to TD participants in terms both of accuracy and precision, as measured by the Point of Subjective Equality and Weber Ratio respectively. Moreover, task performance in the ASD group was susceptible to identical modulating effects of duration range. Notably, however, we found that a significantly larger proportion of ASD participants had to be excluded from the analysis for at least one sensory modality, particularly the visual modality, because they failed to discriminate between durations at above-chance levels. We first discuss the findings from the main analysis, with the caveat that in the visual modality in particular results have to be taken with caution as over 40% of ASD participants were excluded, before we discuss the factors that drove different inclusion patterns in each group.

We first found that TD and ASD participants performed with similar accuracy in each sensory modality. Moreover, in both groups, short durations were generally overestimated while long durations were underestimated (relative to one another). This finding suggests that participants' responses conformed to Vierordt's law (Vierordt, 1868) or central tendency: since durations from both ranges were intermixed within blocks, participants likely constructed an average representation of duration based on all durations experienced in the experiment (Grondin, 2005; Jazayeri & Shadlen, 2010). Under this account, on each trial, participants corrected their duration estimates using their internal average representation. For instance if estimations are corrected to be closer to the middle value (1000ms), then most probe durations will be overestimated in the short range and underestimated in the long range, leading to the observed difference in PSEs. Since the probe (16 possible durations) was more variable than the standard stimulus (2 possible durations), it seems plausible to assume that the standard duration estimate was less strongly corrected than the probe duration estimate. Furthermore, there is evidence that this kind of averaging is applied primarily to the first stimulus (in our case the probe) in comparison designs (Dyjas, Bausenhart, & Ulrich, 2012; Narkiewicz, Lambrechts, Eichelbaum & Yarrow 2015).

Observation of central tendency in both participant groups has important theoretical implications as it fails to support Pellicano & Burr (2012)'s hypothesis according to which ASD individuals have 'hypo-priors' and base their decision to a larger extent on the encoded stimulus itself and less on the predicted (i.e., average) stimulus, compared to TD individuals. If this were the case, we would expect ASD individuals to show a diminished central tendency, whereby duration would be estimated on a single-trial basis and thus be influenced less by an internal average representation. Evaluating the magnitude of central tendency was however not the focus of this study and would need to be further explored (for some data showing reduced central tendency in ASD see Karaminis et al., 2016). Typical central tendency effects in ASD would also sit uneasily with the WCC theory (Frith & Happé, 1994) according to which ASD is characterised by impairments in global/holistic processing but preserved local/detail-focused processing. Again, this framework would lead to the prediction that individuals with ASD would perform the task on a single-trial basis, comparing each pair of durations independently from previously presented pairs. The EPF theory (Mottron et al., 2006) can probably accommodate these results best as it acknowledges that enhanced low-level processing in ASD does not necessarily lead to impairments at higher levels. The extent to which each group relies on an average duration (global strategy) could be examined by manipulating the probes' distribution (e.g., the procedure proposed by Filippopoulos, Hallworth, Lee, & Wearden, 2012). Differentiating the average duration from the middle duration in the range would give more insight into how the internal average representation is constructed in each group.

In line with the PSE data, the Weber Ratio (WR) analysis revealed that TD and ASD participants exhibited similar precision in our temporal comparison task in each sensory modality. More surprisingly, we found that participants were more precise (in standard-normalised terms) when discriminating long than short durations. Hence our data strictly violate the scalar property that is commonly found for interval timing tasks (Wearden & Lejeune, 2008). Note, however, that minor violations of the scalar property are reported with some regularity (e.g., Lewis, & Miall, 2009), and in our particular case the deviant result could perhaps reflect the averaging processes discussed in relation to the PSE, above. Although ASD participants showed slightly higher WRs overall, the group difference was non-significant with a medium effect size ($F(32,1)=1.920$, $p=.177$, $\eta_p^2=.064$). Although the null effect provides support to other studies which found no atypicality of temporal processing in ASD (Gil et al., 2012; Jones, Happé, Baird, Simonoff, Marsden, et al., 2009; Mostofsky et al., 2000), this trend is in line with previous evidence that temporal processing precision is reduced in ASD (Falter et al., 2012; Kargas et al., 2014). Our results therefore bring some reconciliation to inconsistent results reported in the literature and suggest a small to medium difference in the precision of temporal judgements.

Last, RT analysis showed that ASD and TD individuals responded with comparable speed. Both groups also presented the classic distance effect (Moyer & Landauer, 1967) whereby probe durations further away from the standard are responded to faster than probe durations closer to the standard, reflecting the difficulty of the task. They also showed a similar leftward skew in their reaction times: participants responded faster to the shortest durations than the longest durations. The skewness can be accounted for by the time order error (Hellström, 1985): when two stimuli of equal durations are presented sequentially, a bias is often found. In the present study, the standard (always the second stimulus) was slightly overestimated, increasing the difference to a shorter probe and reducing the difference to a longer probe. Crucially, no difference was found between groups, indicating that ASD and TD individuals included in the analysis responded with similar speed, and were susceptible to the same distance and time order effects.

Overall, then, those ASD and TD individuals who could perform the task at above-chance levels were found to behave extremely similarly on our temporal comparison task in the auditory, visual and audiovisual modalities. They were susceptible to the same classic effects in time perception, suggesting the engagement of similar processes to perform the task. This finding appears, at first sight, consistent with work from those previous authors who have assessed a similar range of durations and also used a comparison task, albeit with younger ASD groups (Jones et al., 2009; Mostofsky et al., 2000) and who also found no differences from TD controls. It is also in line with Gil et al. (2012) who found no difference either in accuracy or precision in autistic children using a bisection task. These findings contrast with a sensitivity difference reported by Kargas et al. (2014) in an adaptive comparison task and Falter et al. (2012) using a generalisation task. Our results might thus be taken to suggest that the particular choice of perceptual timing task (rather than the age range of participants) yielded the deviant result reported by Kargas et al. (2014) and Falter et al. (2012).

However, in the present study, a high number of participants were not able to reliably discriminate between durations in at least one of the sensory modalities, most often in the visual modality. Lesser sensitivity in the visual modality is a classic finding in the time perception literature (e.g., Wearden, Edwards, Fakhri, & Percival, 1998) and can explain in part the imbalance in performance between modalities in both groups. The disproportionate rate of exclusion of ASD participants from the visual task analysis could also have been driven by enhanced low-level perceptual processing in ASD interfering with the higher level time perception task (Mottron et al., 2006). Importantly, a significantly larger number of ASD than TD individuals (almost half of the experimental group) had great difficulties when discriminating durations in at least one modality. This suggests that although many individuals with ASD show intact performance, another large subgroup experiences very substantial difficulties processing durations under 2 seconds. A closer look at excluded individuals indicated that difficulties with the time perception task seemed to be driven by different cognitive profiles in the two groups. In the ASD group, exclusion from the analysis was more likely for those with lower IQ scores (including VIQ, PIQ and FIQ). Examination of the relationship between IQ and performance (PSE and WR scores) in participants who managed to perform the task suggest that this association is not monotonic. There may be a threshold effect in the ASD group whereby participants need a sufficient level of cognitive resources to perform the task, which undermines the performance of individuals with a lower level of general cognitive functioning. This is in line with Brodeur et al. (2014) who found that lower-functioning children with ASD showed reduced sensitivity to duration changes using both a bisection and a generalisation task compared to mental age-matched typically developing children. Some support and further understanding for this idea comes from Lambrechts et al.'s (2017) neuroimaging data where it was found that ASD adults seemed to engage less neural resources towards temporal processing regardless of the context compared to their TD counterparts, which means that rather than allocating more resources to temporal processing in a time perception task, ASD participants may be sharing their neural resources between task relevant (e.g., duration) and task-irrelevant (e.g., pitch) features. We speculate that for this reason only ASD individuals with a higher overall level of cognitive resources are able to maintain sufficient neural resources to perform the duration comparison task to above-chance level. This interpretation is compatible with the idea that ASD individuals tend not to process information preferentially in a top-down, task-specific manner but rather in a more costly bottom-up manner (Mottron et al., 2006).

In contrast, in the TD group, failure to discriminate durations at above-chance level was not correlated with IQ scores but was associated with higher AQ scores, i.e., individuals with more autistic-like traits performed less well in a time perception task. This finding is in line with Stewart et al. (2015) who found that higher AQ scores were associated with lower performance in duration comparison tasks in the general population but at odds with Jones, Lambrechts, and Gaigg (2017) who found no correlation between performance on a time bisection task and AQ scores in a large sample of young adults. Excluded TD participants also presented a higher score on the attentional component of the impulsivity scale. Because only 4 TD participants were excluded from at least one condition altogether, however, these findings need to be replicated before we can speculate further.

Although the results of the present study are informative, it is necessary to acknowledge the small sample size, especially after exclusion of a significant number of ASD participants. As a result, subtle differences between groups might have been overlooked, reflective of atypical temporal processing per se or resulting from other differences in information processing generally, such as participants' susceptibility to proactive interference. For the same reason, accurately profiling the included and excluded group was limited by the number of participants in each group.

Overall, however, the increased prevalence of difficulties with time perception in particular in the visual modality in the ASD group could help reconcile discrepant evidence reported in the literature: depending on sampling, and on criteria for data exclusion, performance can appear either intact or atypical in ASD. In this context it is interesting to note that Allman et al. (2011) excluded data from 6/19 ASD children, 2 because they did not acquire temporal discrimination and 4 who produced "disorderly functions" and 1/12 TD children who also produced "disorderly function" in their temporal bisection task. The results of this study showed a difference in accuracy but not in precision for durations ranging from 1 to 4s. Karaminis et al. (2016) excluded 6/29 ASD children and 8/31 TD children and (about 20%) for poor temporal discrimination performance (WR out of range [0, 1]) in reproduction and comparison tasks. After excluding 5/17 ASD participants for mixed criteria (mental retardation, absence of language, attrition during the tasks, or inability to perform the training phase of the proposed procedure correctly), Gil et al. (2012) found no differences in a temporal bisection task with durations ranging from 0.5 to 2.5s. A recommendation for future research in this domain would be to harmonise not only performance indices but also criteria for inclusion or exclusion in the analysis.

Reduced proficiency in time perception can potentially have wide repercussions at higher cognitive levels. An inability to process short durations up to 2s might prove a disadvantage in the fine coordination of motor and sensory cues such as those used in communicative behaviour. For instance, the coordination of interpersonal speech or the integration of speech and gesture require that we combine auditory and visual information based on their duration and the delay between them (Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Leonard & Cummins, 2011; Treffner, Peter, & Kleidon, 2008), and the timing of turn-taking can affect the speaker's self-presentation both in the context of informal conversation or a high-impact situation such as a job interview (e.g., Brosy, Bangerter, & Mayor, 2016; Roberts et al., 2011). Although our results do not support a time perception deficit per se as a universal feature in ASD, they suggest that time processing of short durations engages a greater amount or level of cognitive resources than in the TD population and that as a result, a significant proportion of individuals on the spectrum (mostly individuals with less cognitive resources) could present more severe difficulties. Either profile could contribute to core aspects of the disorder such as social interactions and communication (e.g., de Marchena & Eigsti, 2010; Warlaumont, Oller, Buder, Dale, & Kozma, 2010), particularly if subsequent research were to reveal that even those individuals with ASD who ultimately succeed on interval timing tasks are subject to some developmental delay. Although examination of the included and excluded participants within the ASD group revealed no difference in the AQ scores or any of the ADOS subscores, it is important to remember that these measures have not been designed specifically to evaluate the timing of reciprocal social behaviour. Indeed, very little is known about how performance on psychophysical time perception task relate to timing and time processing in behaviour, or indeed whether temporal aspects of behaviour are atypical in autism. The second part of this work will therefore focus more closely on temporal aspects of social behaviour.

Chapter 3
Temporal dynamics of speech and co-speech gesture in ASD

3.1 Aims

The approach taken in chapter 2 to investigate temporal processing in autism was in continuation with the existing literature. Using a classic psychophysical task to assess duration comparison in the visual, auditory and audiovisual modality, the data provided evidence that time perception draws on cognitive resources more in ASD than in TD individuals and supports the notion that at least some autistic individuals with ASD show difficulties when making short duration judgements. This work contributes and adds to the existing body of evidence that points to temporal processing atypicalities in autism, albeit with some inconsistencies, and provides a starting point for the second part of the work developed in this thesis.

Based on the evidence reviewed so far, it is quite clear that time processing is not dramatically impaired in autism overall, nor is decreased temporal processing performance a universal feature of ASD. This seems to preclude the idea of a theory of autism based on temporal processing deficits, or the notion that a deficit in temporal processing could causally underlie the core features of autism as suggested by Allman (2011). However the appeal of studying time processing as a phenotypical feature of autism is its ubiquity: because neural and cognitive functions are all in some way concerned with time subjectively or objectively, dysfunctional temporal processing may be relevant for a wide range of activities and behaviours. Authors emphasize how potentially disruptive atypical timing could be to domains as varied as motor behaviour, sensory integration, memory and learning, and language and communication (Allman, 2011; Falter & Noreika, 2011). As mentioned in the introduction, they also draw potential relationships between time processing and the core features of autism: repetitive behaviour could be a putative strategy to parse time, and communication impairment could result partly from, or be aggravated by, difficulties integrating and coordinating incoming and outgoing pieces of information in time for the benefit of all interlocutors.

Surprisingly, the integrity of temporal aspects of behaviour and their relation to the autistic phenotype has received relatively little attention in research. One recent study (Kunchulia, Tatishvili, Lomidze, Parkosadze, & Thomaschke, 2017) looked at time-based event expectancies in children with ASD. The authors used a binary choice response task in which participants had to follow a target which could be moving to either the left or the right. For half the participants, a short pre-target interval (200ms) predicted a movement to the left with 80% validity and a long pre-target interval (800ms) predicted a movement to the right with 80% validity, and vice-versa for the other half of the participants. They found that children with ASD responded faster to frequent combinations between the long pre-target interval and the direction of the target, showing that they were able to form time-based event expectancies. This is an important piece of work because it shows one instance where autistic children were able to use temporal information effectively to optimise behaviour. It is worth noting that this was a small sample study, and that 7 out of 16 autistic participants had to be excluded because they could not do the task, whilst none of the TD children were excluded. However, in light of the findings from chapter 2 which indicated that autistic individuals with a lower IQ were more likely to show difficulties with short duration judgements, autistic participants included in Kunchulia et al. still presented a wide range of general intellectual abilities (IQ range 71-112), wider in fact compared to the TD children who presented a higher IQ overall (IQ range 97-119). Time-based expectancies are crucial in our interactions with the environment: by making use of temporal cues, we can predict not only what can happen but also within what time frame it is likely to happen, allowing us to deploy relevant resources for a limited, appropriate time window only. For instance, we know to hang up the phone after a number of tones, because we know that if we have had no answer after a certain time window, it is unlikely that we are going to have an answer at all. In the domain of communication in particular, we depend extensively on our expectations of timely responding. In conversational settings, the duration of inter-turn gap predicts listeners' perception of the speaker's willingness and agreement in the following utterance (Roberts et al., 2011). Taking into account the diminished sensitivity to short durations in autism, it is possible that atypical duration of pauses and inter-turn gaps in autistic individuals' conversation affect their listeners' judgement of knowledge and cooperativeness. Conversely, lesser sensitivity to the short durations of pauses and gap in conversation might reduce autistic listeners' ability to make judgements on their conversational partner's cognitive processes. In the case of factual questions (Brennan & Williams, 1995; Smith & Clark, 1993) as well as job interview questions (Brosy et al., 2016), response latencies are used by listeners to make inferences about respondents' knowledge and cognitive processes. When responses are delayed, listeners are more likely to form negative inferences about the respondents, such as judging them ignorant or uncooperative (de Ruiter, Mitterer, & Enfield, 2006; DeGroot & Motowidlo, 1999; Smith & Clark, 1993). Atypical temporal dynamics could therefore directly impact the outcome in situations such as job interviews or eye-witness testimony (Maras & Bowler, 2012) for individuals with autism. Our first aim in the second part of this thesis is therefore to evaluate whether the temporal dynamics of communication are atypical in autistic productions, and whether they affect listeners' judgements about the quality of what is being communicated.

The second main claim in the literature on time perception in autism is that atypical temporal processing might be related to clinical features of autism, which has also received relatively little empirical attention. In their time bisection study, Allman, DeLeon, and Wearden (2011) looked at the correlations between autistic children's individual psychophysics scores (bisection point and Weber ratio) and their ADOS, ADI-R and IQ scores and subscores. On a subset of 8 children, they found that children producing a smaller bisection point showed poorer language and communication and working memory scores. More recently, in a larger sample of 27 autistic children and adolescents, Brenner et al. (2015) recently found that diagnosis of ASD, younger age and poorer working memory performance predicted poorer performance in a time reproduction task (namely lower accuracy and lower consistency). Temporal aspects of behaviour, particularly communication, are all the more important to consider given that a diagnosis of ASD is based on subjective, interview-based observations. Because there are no biomarkers for autism, tools such as the gold-standard ADOS (Lord et al., 2000), which is widely used to inform diagnosis in research as well as in clinical settings, therefore requires an evaluation of fairly complex and abstract features such as the quality of communication displayed during the interview, or the ability to coordinate eye contact, language and gesture. These features each encompass a number of behaviours and skills, any of which might be typical or atypical. Moreover the evaluations are based on the interviewer's judgement rather than on the quantification of behaviours. The second aim for the second part of this thesis is to evaluate whether quantifiable temporal aspects of communicative behaviour can predict clinical scores, with a view to inform a diagnosis process.

3.2 Literature review

Communicating with one or several interlocutors is a complex, multidimensional process. Looking only at live conversational exchanges, we observe that each partner produces multiple auditory and visual streams of communication. Speech and vocalisations make up most of the auditory part, whilst mouth movements, gestures, facial expressions and other movements constitute most of the visual part. In addition to the first-level meaning conveyed by each word or gesture, a wealth of information is layered in the choice of words, the pace and intonation of the primary message. A piece of conversation gives us information about the speaker's physiological and emotional state (are they tired, angry, sad?), about their relation to the receiver (are they friends, colleagues or strangers to each other?), whether they mean their message literally or ironically and so on. These extra layers are expressed through (the list is not extensive) variations in pitch, loudness and speech pauses for the auditory part, amplitude, velocity, position in space and gestural holds for the gestural part. Together, these channels are timed exquisitely so that information in one modality can facilitate, complement or modulate information in another modality. For instance, gestures often precede the occurrence of their lexical affiliate by a few hundreds of milliseconds (Morett, O'Hearn, Luna, & Ghuman, 2016; Morrel-Samuels & Krauss, 1992), so that it prepares the receiver and helps them either disambiguate or capture more information about the target word (e.g., Goldin-Meadow, 1999).

Coming back to autism, the question can be asked whether the occasional diminished precision in temporal processing observed in psychophysics tasks extends to behaviour outside of the lab and in particular to the domain of communication. In the second part of this thesis, we are therefore choosing to turn the microscope around and rather than looking at how autistic individuals perform in time processing tasks, we shift our investigation to the timing of spontaneous behaviour in autistic individuals. Specifically, this second part will focus on the temporal dynamics of speech and gesture in spontaneous communicative productions in autism and their relation to the autistic phenotype. The following section will review some of the literature on temporal aspects of communication and social interaction, in typical development and in ASD. The review will distinguish between two aspects of social interaction: verbal (speech) and non-verbal (gesture) communication, and outline the temporal relationships between the two in the typical and autistic populations respectively.

3.2.1 Temporal dynamics of verbal communication: Speech and pauses

Speech is in essence dependent on time: it can be described as a series of changes in time which generate vowels, stressed and unstressed syllables and phrases (Kotz & Schwartze, 2010). **Temporal patterns of speech** include serial order (the succession of events in time), but also recurrence, or temporal regularity. Recurrence in particular allows the generation of predictions in time (Schwartze, Rothermich, Schmidt-Kassow, & Kotz, 2011), and facilitates speech processing. Often, speech, like music, is produced periodically: perceptual "beats" occurring near the onset of vowels are produced at a regular pace, albeit with significant variations between languages (e.g., Tilsen & Arvaniti, 2013). Port (2003) suggests that this periodicity aligns with neural oscillations that regulate attention. This allows the speaker to bias the motor production of speech in such a way that salient events in speech align with attentional attractors. Mastering the timing of speech, therefore, means controlling that the auditory information is going to be received at optimal engagement time.

Whilst the primary meaning of the spoken signal is carried by words and grammar, temporal organisation of speech is part of **prosody**, which encompasses properties of the speech signal that modulate and enhance its meaning and are not easily transcribed in the written form, including features such as intonation, speech rate and pausing. More technically speaking, prosody describes the characteristics of speech deriving from variations in the duration, amplitude and fundamental frequency of speech sounds which affect the communicative function (Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007). Prosody serves various goals: grammatical or syntactic prosody helps the listeners to segment and interpret speech in its intended grammatical acceptation. For instance, the lengthening of the final syllable or a pause can indicate the end of a phrase. Ambiguous phrases can also be resolved by using prosody. For example, the phrase "Ellen the dentist is here" can be understood as "Ellen, the dentist is here" or as "Ellen, the dentist, is here" (Peppé et al., 2007). In English, the intonation as well as the presence and duration of pauses in the utterance will indicate to the listener which of the two possible meanings is intended. Pragmatic prosody conveys social information beyond what is contained in the syntax of the sentence, for instance by employing emphasis or contrast in a sentence. The interpretation of the statement "I didn't say that" can change significantly by adding emphasis on the "I": "*I* didn't say that", implying that the speaker is denying the assertion that they had said something, but also suggesting that someone else did. By moving the emphasis to "that" for instance ("I didn't say *that*"), one understands a slightly different situation where the speaker challenges the content of what they are quoted to have said. Finally, affective prosody can convey information about the emotional state or mood of the speaker. For instance, anger and happiness/joy are generally characterised by higher mean pitch, wider pitch range, high speech rate, increases in high frequency energy, and usually increases in rate of articulation whilst sadness, as well as boredom, is characterised by a decrease in mean pitch, slightly narrow pitch range, and slower speaking rate (Juslin, Laukka, & Bänziger, 2018; Murray & Arnott, 1993). Although temporal aspects are only a part of prosody, it is possible that atypical temporal processing in the context of speech could affect autistic individuals' ability to disambiguate some phrases or convey non syntactic information.

Although atypical prosody in ASD is widely reported in clinical or anecdotal reports, often described as "monotonous", "robotic" or "exaggerated" (e.g., Asperger, 1944; from Wing, 1981), evidence is relatively scarce and conflicting as to what particular aspects of prosody and how they differ in autism (McCann & Peppé, 2003; Paul, Augustyn, Klin, & Volkmar, 2005). In recent studies (McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007; Peppé et al., 2011, 2007), evidence indicates that both expressive and receptive prosody development is delayed in many individuals with autism, although the studies don't look at temporal aspects of prosody in particular.

Temporal aspects of speech are also crucial when considering speech as an interaction between two or more partners. One of the simplest and strongest organisation pattern of verbal communication is **turn-taking,** which displays remarkably similar rules across languages despite language-specific variations in the gap durations involved (Stivers et al., 2009). In order to share communication time, and allow an exchange of information, partners in conversations, but also other types of speech exchanges such as an interview, a debate or a ceremony, adopt a turn-taking pattern: only one partner speaks at a time, and partners alternate turns with minimal overlap or gap between turns (Sacks, Schegloff, & Jefferson, 1974; Schlegglof, 2000). When dysfluency occurs, for instance when partners overlap, one of them usually interrupts their turn to repair the violation. Transitions between partners are incredibly fast and efficient. On average, 85% of the transitions in naturalistic conversations have less than 1 second gap or overlap, and 45% of the transitions have under 500ms gap or overlap (de Ruiter et al., 2006). In fact, listeners are not simply waiting for their partner's turn to end before taking their turn, but actively anticipate the end of the turn and predict its timing to minimise the transition time. Transition times can however be slightly longer in different contexts: for instance, when answering a factual question, the pause between the question and the answer was found to last between 2.65 and 8.83 seconds (Smith & Clark, 1993) and in a job interview the pause between the interviewer's question and the applicant's answer was found to last between 1.88 and 9.50 seconds (Brosy et al., 2016). In those two contexts, the respondent or interviewee is expected to take longer to respond, because they trade-off extra time in order to come up with the best possible answer.

In addition to its primary goal of exchanging information, verbal communication serves another social goal of **self-presentation** (Smith & Clark, 1993). It is not only about what we say, but how we say it. In that context, several pieces of evidence indicate that the timing of turn-taking can affect the speaker's self-presentation and the listener's evaluation of not only the verbal content of the message, but also the cognitive processes behind the message. Roberts et al. (2011) manipulated recordings of naturalistic conversations in three languages by introducing a 0ms, 600ms or 1200ms inter-turn gap between a speaker's request of assertion, and the affirmative response that followed them. They asked independent native speakers to rate the willingness of the addressee to comply with request or agree with the assessment. They found that irrespective of language background, all raters judged that the addressee was less ready to comply to the request or agree with an assertion when the inter-turn gap duration was longer. Brennan and Williams (1995) investigated the impact of the speaker's response time on the listener's *Feeling-Of-Another-Knowing*, the feeling of whether the speaker is confident and knowledgeable. They asked participants to rate how likely speakers knew the correct answer to a question based on their responses to trivia questions only (the actual question was replaced by a generic question so the rater's own knowledge wouldn't affect their judgement), manipulating the duration of the inter-turn gap between the question and answer to be either short (1s) or long (5s). They found that answers following 1s-gaps were rated to be more likely to be correct than answers following 5s-gaps, showing that the delay in responding to a question affects the listeners' trust in the answer, or in other words the speaker's credibility. In a job interview setting, Brosy et al. (2016) showed that the longer the gap between the interviewer's question and the applicant's reply, the less likely the recruiters were to produce a positive hiring recommendation. Although it is unclear in this study whether interviewers directly use temporal information (gap duration) to inform their recommendation, or whether gap duration is simply highly correlated with the quality of the response on which the interviewers base their assessment, this result suggests that temporal information could contain data relevant to employability skills.

Studies on reciprocal interaction in autism have consistently reported that autistic individuals struggle with turn-taking in conversation. Autistic children are more likely not to respond to a question than typically developing children (Capps, Kehres, & Sigman, 1998) and infants and children with ASD produce less turn-taking vocalisations than both their developmentally delayed and typically developing peers (Goldberg et al., 2005; Loveland, Landry, Hughes, Hall, & McEvoy, 1988; D. C. Wimpory et al., 2000). Once again, however, little research has been conducted to date quantifying the temporal dynamics of turn-taking in autism when it does take place, although some data shows that children with ASD leave a longer gap before taking a turn than matched TD children (over 200ms longer), and particularly so when their turn comes after a question (Heeman, Lunsford, Selfridge, Black, & Santen, 2010). Moreover, in the context of a mock job interview, Mitchell (2015) found that young adults with autism introduced more pauses between and within utterances, and that they were deemed less employable than their typically developing counterparts. In subsequent unpublished work, Mitchell and Volden (IMFAR 2015) asked fifty-nine raters to judge the quality of communication of the same interviews and found that ratings were poorer for recordings from the autistic compared to the typically developing interviewees. Taken together, therefore, the evidence suggests that increased pausing time could put not only the efficiency of exchanging information, but also self-presentation skills at risk in autism: failure to take turns or answer a question in the expected timeframe could affect the listener's assessment of an autistic person's willingness to communicate, their knowledgeability, level of confidence and credibility, and even their level of skills and employability.

3.2.2 Temporal dynamics of non verbal communication: Gestures

Although speech provides an incredibly rich and precise source of information in support of communication, speech constitutes only one half of spoken language. In the visual domain, gestures constitute a complementary channel of communication which doesn't directly interfere with auditory information but can be integrated simultaneously. Co-speech gestures are defined as spontaneous hand movements which accompany speech (McNeill, 1992, 2005). They differ from other movements such as self-touching movements (e.g., scratching), postural movements or conventional gestures (e.g., thumbs-up for "okay") which are cultural and socially learnt (Butterworth & Beattie, 1978; Kendon, 1972, 1980).

Gestures serve multiple purposes in communication (Goldin-Meadow, 1999). At the expressive level, they facilitate word searches and guide elocution (the skill to produce clear and distinctly articulated speech). At the receptive level, they can facilitate, complement and disambiguate speech (Cassell, 1998; McNeill, 1992). For instance, gestures can draw attention to a particular point in time when the speaker is adding emphasis and wants the listener's full attention (by adding a gestural beat to the vocal emphasis for instance), or to a point in space that is relevant to the message; they can speed up and improve the understanding process by simulating a situation or illustrating a sentence which would take a lot of words to describe precisely, but can be evoked easily by a movement (for example to describe the size and shape of an object or pattern); they can complete the meaning of a phrase by adding relevant information (for instance the phrase "look at this" only makes sense in relation to an accompanying gaze or point). Finally it has been shown that encoding the speaker's gestures improves episodic memory for the listener (So, Sim Chen-Hui, & Low Wei-Shan, 2012) so the impact of gesture extends beyond the immediate context of social interaction.

One of the most widely accepted classification of co-speech gestures is the one proposed by McNeill (1992). **Iconic gestures** depict, in form and/or manner of execution, aspects of the action or event being described. They have a close formal relationship to semantic content of speech. For instance, a forward movement of the index finger accompanying the utterance "he rang the bell for a long time" would illustrate the action of pressing the bell. **Metaphoric gestures** also illustrate an element evoked in the speech, but the concept they refer to is abstract. For instance, a brushing motion of the hand outwards during the same utterance "he rang the bell for a long time" could evoke the time spent ringing the bell. Both iconic and metaphoric gestures are *representational*: they directly illustrate a feature of the scene, action or concept mentioned in the co-occurring speech. **Deictic gestures** locate an action in space and/or time and consist of a pointing motion (with a finger, hand or other parts of the body). Continuing with the example above, pointing at an imaginary person in the gestural space would for instance refer to "he". Finally **beat gestures** are small, baton-like movements which keep the same form regardless of the content of the accompanying speech. They serve pragmatic functions such as emphasis or speech repair. For instance, by producing a beat gesture at different time points in the example utterance, the speaker could bring the listeners' attention more particularly to the agent ("he"), the action ("rang the bell") or on the commentary about the action ("for a long time").

3.2.3 Temporal dynamics of communication: Coordination of speech and gesture

Importantly, speech and gesture have a tight temporal correspondence which has been observed consistently for several decades (e.g., Kendon, 1980; McNeill, 1992; Wachsmuth, 1999; Wiltshire, 2007). Some theories argue that gestures are the residual traces of a proto-speech in which gestures were necessary to communication before language incorporated more and more vocalisations (Arbib, Liebal, & Pika, 2008; Corballis, 2003). In contrast McNeill (McNeill, 1992; McNeill, Bertenthal, Cole, & Gallagher, 2005), based on work by Kendon (1972), argues that speech and gestures are the expression of the same thought processes, conveyed through two different media, and that as a result they are produced in a coordinated way.

> *"Gestures and speech are partners in shaping communication and giving kinematic and temporal (visual and auditory) dimensions to our thoughts."* (Esposito & Esposito, 2011, p.256)

Speech and co-speech gestures together are co-expressive (McNeill et al., 2005), which means that they are believed to be the combined expression of the same underlying thought and intention to communicate. Speech in the auditory modality constitutes the combinatoric track where a series of symbolic elements are organised and produced sequentially, whilst gestures in the visual modality offer a synthetic track in which one movement sequence can embody several concepts or features and *exhibit meaning in its own right* (McNeill, 1992).

More specifically, a gesture and its related speech often overlap in time, and the 'stroke' of the gesture (the phase in which "the meaning of the gesture is expressed", McNeill, 1992), in other words the moment of the gesture with most emphasis, often the maximum extension in space or the point of fastest acceleration) usually occurs during the target word. Moreover in most cases the gesture is initiated before and not after the corresponding speech (Butterworth & Beattie, 1978; Morrel-Samuels & Krauss, 1992). Morrel-Samuels and Krauss (1992) asked young adults to describe a picture to a confederate and studied the correspondence between speech and co-speech gestures. They found that 60% of gestures were initiated less than 1s before the onset of the related speech, and over 80% of gestures were initiated less than 2s before the onset of the related speech. In addition, the asynchrony between gesture and speech increased when the speaker was communicating an unfamiliar concept. In contrast, Chui (2005) found that in a Chinese corpus of spontaneous conversation, the stroke phase of iconic gestures happened mostly in synchronisation with their speech affiliate. When this temporal correspondence is disturbed, the quality of communication decreases. In an ERP study by Habets, Kita, Shao, Ozyurek, and Hagoort (2011), videos of a person gesturing were paired with either a congruent or incongruent audio recording of a word. In addition, the asynchrony between video and audio was manipulated to be 0, 160 or 360ms (with the onset of gesture preceding or synchronised with the onset of speech). ERP recordings showed a differential response to congruent compared to incongruent pairings, but only for pairings with a 0 or 160ms. This shows that gesture and speech are processed as a whole only within a certain time frame. When they are produced too far apart, the association is broken.

The precise timing of speech and gesture can also affect the interpretation of the message. Treffner, Peter and Kleidon (2008) used an avatar display of a character pronouncing the phrase “put the book there now” whilst producing a simple beat gesture (simple two-phase gesture where a part of the body, often the finger or hand, is displaced in one direction then reverses back to its initial position). They manipulated the time window during which the gesture was performed by the avatar so that it was centred on various frames falling somewhere during the “book-there” segment of speech (the temporal distance between the earliest and latest position of the gesture being 760ms). They asked participants to rate where the focus of the sentence was. They found participants’ perception of the focus in the same spoken sentence changed from “book” when the beat stroke (maximal extension) was centred on “book”, to “there” when the beat was centred on “there”, demonstrating that the timing of gesture can directly affect the way speech is perceived and interpreted. In another study manipulating the timing of beat gesture, Leonard and Cummins (2011) used naturalistic audio-video recordings of a man speaking a simple sentence which each included one simple beat gesture associated to one emphasized word. In each trial, they asked participants to compare two recordings – one untouched recording , and one recording where the audio track was shifted by 200, 400, 600 or 800ms (either ahead of the video or delayed) – and decide which of the two recordings was “out of synch”. They found that participants detected asynchrony easily when the audio track was ahead of the video track (the “gesture-lag” condition), even for lags as short as 200ms, but not when the audio track was delayed compared to the video track (the “gesture-lead” condition). In the gesture-lead condition participants needed a greater asynchrony (600 or 800ms) to reliably detect which recording was “out of synch”. This shows that beat gestures have an asymmetrical temporal relation to speech event: there is a certain temporal window before the speech event during which the gesture can be produced (either ahead or synchronously to the speech event), but if the gesture happens after the speech event, even with a very short delay it is perceived as out of synch. It could be that beat gestures need to be timed correctly ahead or closely synched to the speech event to announce the imminent emphasis in the speech, and draw the listeners attention to an important part of the message. An alternative hypothesis is that gestures facilitate the production of speech for the speaker and therefore appear first in the communication process. In the case of hesitation or word search, it is also the gesture that appears first. In any case, this shows again that the timing of gestures in relation to speech needs to be quite precisely coordinated in time to serve its purpose.

A less well-known aspect of speech and gesture coordination resides in the synchronisation of speech pauses and gestural pauses or "holds". The dynamics of holds and their relationship to speech and speech pauses have generated very little research, but some evidence suggests that holds appear to be distributed similarly to speech pauses and to overlap with them in adults (Esposito, McCullough, & Quek, 2001) and in children (Esposito & Esposito, 2011). The authors suggest that the synchronisation of speech pauses and holds reflects the dual essence of language and indicates that a common mechanism regulates the production of speech and gestures.

3.2.4 Speech and gesture in autism

In early reports already, clinicians have noted atypicalities not only in verbal but also non verbal communication in ASD (Kanner, 1943). The first description of Asperger's syndrome (Asperger, 1944; from Wing, 1981) states that both production and perception of gestures are affected:

> *"Gestures are limited, or else large and clumsy and inappropriate for the accompanying speech. Comprehension of other people's expressions and gestures is poor and the person with Asperger's syndrome may misinterpret or ignore such non-verbal signs."* (Wing, 1981, p.116)

The clinical literature is assertive about a deficit of communicative gesture in ASD, frequently reporting a lower frequency of gestures and poorer integration of co-speech gestures with the accompanying speech. In fact gold-standard diagnostic tools such as the ADOS (Lord, Rutter, Dilavore, & Risi, 2002) and the Autism Diagnostic Interview (Lord, Rutter, & Le Couteur, 1994) rate the absence or scarcity of gestures as a symptom of ASD. However exactly how gestures differ in autism is generally under-researched and little understood.

Much of the research on gestures in ASD focuses on quantifying various types of gestures, and in particular proto-gestures in the context of development (e.g., Iverson et al., 2017). Wetherby et al. (2004) reported that inventory of gestures at 2 years of age was the strongest predictor of autism and Colgan et al. (2006) found that infants who would later be diagnosed with ASD produced a lesser variety of gestures, but with similar frequency and initiation than their Typically Developing (TD) counterparts. A good amount of evidence suggests that protodeclarative pointing gestures (pointing to share attention) are less frequent in autism (Camaioni, Perucchini, Muratori, & Milone, 1997; Mundy, Sigman, & Kasari, 1990; Mundy, Sigman, Ungerer, & Sherman, 1986; Watson, Crais, Baranek, Dykstra, & Wilson, 2013), but there is surprisingly little research and consistency to show that other types of gestures are less frequent in autism than in typical development (de Marchena & Eigsti, 2010; but see Iverson et al., 2017; Tantam, Holmes, & Cordess, 1993). In fact, three studies have found that autistic children and adolescents use iconic gestures just as much as their typical counterparts (Capps et al., 1998; de Marchena & Eigsti, 2010; Morett et al., 2016), although the overall variety of gestures appears to be reduced (Colgan et al., 2006).

Whilst quantitative differences in gesture production in ASD are not strongly replicated in empirical studies (particularly in adults), such differences are consistently implicated in the diagnostic criteria of the disorder along with wider qualitative atypicalities in the use of gestures and integration of gesture and speech. The *quality* of gestures, much like the *quality* of speech, has been repeatedly described as “odd” in autism (Asperger, 1944a; Lorna Wing, 1981), so much so that gold-standard diagnostic instruments, such as the ADOS, consider e.g., “exaggerated” gestures as diagnostically important in individuals with fluent language. Little empirical research, however, has addressed the question of qualitative differences in non-verbal communication systematically. García-Pérez, Lee and Hobson (2007) collected subjective ratings of the quality of communication between an experimenter and children and adolescents. They reported lower subjective ratings of affective engagement between conversational partners in autism and lower “flow” in the exchange but the study did not clarify what specifically was different between autistic and non-autistic patterns of communication. More recently, de Marchena and Eigsti (2010) and Morett et al. (2016) collected ratings of quality of communication displayed by adolescents in a naturalistic narrative task. Both studies report that narratives from autistic participants were rated as less coherent and less engaging than the narratives of their typical counterparts.

Importantly, the latter two studies also unveiled another way in which gesture production might differ in autism: their temporal coordination to speech. De Marchena and Eigsti (2010) found that autistic adolescents' iconic gestures were less closely coordinated in time with the co-occurring speech, with an average asynchrony of 240ms between a gesture and its corresponding speech in the TD group, but an average of 490ms in the ASD group. In light of Habets and colleagues' findings, it is therefore possible that gestures produced by an autistic individual are not as well integrated to co-occurring speech by a listener as they are on average further apart in time. In line with this idea, ratings of quality of communication collected from independent raters by de Marchena and Eigsti revealed that ASD productions were judged to present poorer ratings than TD productions, and that gesture-speech asynchrony accounted for a significant 20% of the variance in quality of communication ratings (after correcting for IQ). Using a similar experimental set-up, Morett, O'Hearn, Luna, and Ghuman (2016) reported that TD and ASD adolescents produced iconic and deictic gestures equally frequently, but that the ASD group produced fewer metaphoric and beat gestures than the TD group. The asynchrony between speech and gesture was not different between TD and ASD productions, however ASD individuals produced significantly more gestures that were more than 200ms away from the corresponding speech. Interestingly, Morett et al. added a condition in which the listener was not visible to the speaker, and reported that both TD and ASD adolescents produced fewer gestures in the non visible condition, suggesting that gestures are at least partly aimed at the listener in both groups, and that the production of gesture is appropriately modulated by the social context in ASD.

Other aspects of co-speech-accompanying movements in ASD could contribute to modify the efficiency of gesture in autism. Repetitive motor behaviour as well as self-stimulatory gestures (Tantam, Holmes, and Cordess, 1993) might compete with the use of co-speech gestures. Moreover, Cook, Blakemore, and Press (2013) found that the quality of movement was atypical in ASD: they tested the kinematics of simple movement in adults and found that ASD individuals did not minimize jerk to the same extent as their TD counterpart, and moved with greater acceleration and velocity. If this profile extends to co-speech gestures, this could affect autistic individuals' quality of communication.

To our knowledge, no research has investigated gestural holds in the context of ASD. The next chapter will provide the first data on holds and their synchronisation with speech pauses.

If gesture production in autism is understudied, gesture comprehension claims almost no research. Addressing this gap in the literature, Dimitrova, Özçalışkan, and Adamson (2017) asked young children with and without ASD matched on receptive language to choose the target picture out of two choices based on a word alone (e.g., "sofa"), a gesture alone (e.g., pointing at sofa), a word and a reinforcing gesture (e.g., "sofa" + point at sofa) or a word and a supplementing gesture (e.g., "sitting" + point at sofa), and found no difference in the pattern of gesture comprehension between groups for iconic, deictic and conventional gestures. Silverman, Bennetto, Campana, and Tanenhaus (2010) addressed the underlying mechanics of gesture comprehension in autism in an interesting eye-tracking study in which the authors asked ASD and TD adolescents to identify a target shape amongst four candidates based on a speech-only or on a speech-and-gesture description. They found that although both groups succeeded to identify the correct target, the presence of gestures facilitated comprehension for TD participants compared to the speech-only condition, but actually hindered ASD participants' performance: when the speaker produced a disambiguating gesture alongside speech, TD participants' eye-gaze fixated to the target earlier than in the speech-only condition for the TD group, but later in the ASD group. Both groups performed close to ceiling on a gesture-only control task so it seems that the diminished performance in the presence of gesture and speech in the ASD group cannot be explained by a deficit in gesture-only processing, but rather by a difficulty to integrate multimodal communicative information – speech and gesture simultaneously. In line with this interpretation, Hubbard et al. (2012) found that whilst the presence of beat gestures in communication modulates the auditory neural response in typical listeners, autistic listeners do not show this modulation of the auditory response but rather an enhanced visual response. This suggests that whilst in typical listeners speech and gestures are integrated as one multimodal signal, in autism the different modalities are processed separately, which could explain why the presence of gesture actually hindered performance in Silverman et al.'s ASD group. More research is needed to have a better understanding of gesture comprehension pathways in autism.

3.3 Research objectives

Although atypical temporal processing has been put forward as a potential factor underlying core features of ASD, the integrity of temporal aspects of behaviour and their relation to phenotypical features of autism has received relatively little attention in research. The literature review in this chapter has underlined how the functions of both speech and gestures are served by accurate timing within each modality (auditory and visual), and by the temporal coordination of various streams of information, and that this fine temporal tuning might differ in autism. Our first objective in the second part of this thesis is therefore to evaluate whether the temporal dynamics of speech and gesture are atypical in autistic productions, and whether this relates to the subjective perception of autistic communication.

A main reason for investigating the temporal dynamics of communication in autism is that a diagnosis of ASD is based on subjective observations made during a staged social interaction, for instance using tools like the ADOS. These observations are usually qualitative (relying on "atypical", "odd" or "exaggerated" behaviours) or limited to the count of occurrences of a particular behaviour during a limited time (e.g., the number of iconic gestures over a one hour interaction). Such qualitative observations are prone to personal interpretation and biases and what is considered 'atypical' or 'odd' is most likely heavily influenced by cultural contexts and it may also change over time. In addition, relatively short clinical observations may not provide sufficient opportunities to observe differences in relatively crude measures such as the frequency with which an individual produces iconic gestures. In the absence of reliable biomarkers for autism, it would be valuable to identify objective quantifiable aspects of reciprocal communication behaviours with a potential to support and/or validate a diagnosis of ASD. The objective for this second part of this thesis is therefore to evaluate whether quantifiable temporal aspects of communicative behaviour could inform a diagnosis process.

Chapter 4
A systematic analysis of temporal aspects of communication in Autism Spectrum Disorders

Part A – Methods

4.1 Introductory note

The opportunity for the study reported in this chapter stemmed from a collaboration with then post-doctoral fellow Dr Katie Maras on a study primarily investigating eyewitness testimony in autism, published in the Journal of Autism and Developmental Disorders (Maras, Memon, Lambrechts, & Bowler, 2013). The original study specifically aimed to extend the literature on eyewitness research in ASD by testing memory for a live and personally experienced event, in which AL acted as an experimenter. The live event consisted of a first-aid scenario centred around a manikin who was the purported victim of a car crash. AL and the participant took turns in performing a series of first aid actions on the manikin. Later on, the participant was interviewed by KM following the procedure recommended by the Home Office for professionals who interview eyewitnesses, including a free recall part where the participant reported everything they could remember about the event without guidance or feed-back, followed by a Q&A section where the interviewer asked more details about the elements mentioned by the participant.

Beyond its original purpose, this study provided an ideal opportunity to obtain ecologically valid communication samples in a context that would elicit monological speech with accompanying gestures. The first section in this chapter provides a detailed description of the methods used to analyse temporal aspects of the communication in audiovisual recordings of these eyewitness interviews.

4.2 Objectives and choice of variables

Chapter 4 will present a breakdown of some temporal aspects of speech and gesture behaviours in a corpus of spontaneous communicative samples produced by adults with and without ASD. Because we were interested in the temporal dynamics of communication, we chose variables that reflected prosodic aspects rather than semantic content. Following Peppé, McCann, Gibbon, O'Hare and Rutherford (2007) who defined prosody as the characteristics of speech deriving from variations in the duration, amplitude and fundamental frequency of speech sound, we selected the fundamental frequency (which is perceived as **pitch**) and intensity (which is perceived as **loudness**) as acoustic variables, and quantified the **speech/pause behaviour**. Considering the data available, we mirrored these choices in the gesture domain by choosing **motion energy** (a continuous measure of the amount of movement produced over time) as kinematic variable, and quantified the **motion/hold behaviour**. Because the literature on temporal aspects of gestures in autism generally focuses on particular type of gestures, and because data on gesture production in autistic adults is scarce overall, gestures were annotated to provide a better description of the dataset, and a basis for comparison with previous studies. In particular, the **type of gesture** and their **characteristic times (onset, stroke, offset)** were coded manually. In order to measure the impact of temporal aspects of communication on the autistic phenotype, we collected ratings of **quality of communication** that we used throughout the chapter as a point of reference, or an "outcome" measure of the hypothetical differences between TD and ASD groups. Standard measures were also used to measure cognitive functioning (**IQ**) and clinical features of autism (**ADOS, AQ**). Again, they were used throughout the chapter as an "outcome" measure.

The objectives of this study were fourfold. **The first objective** was to establish whether quantifiable temporal aspects of communication were atypical in ASD. For each variable, we were therefore primarily interested in direct group comparisons and interactions involving the factor group. **The second objective** was to investigate the relation between temporal aspects of communication and phenotypical features of ASD. To that end, we explored the correlations between each variable and quality of communication ratings, ADOS and AQ scores, and IQ scores. **The third objective** was to provide quantifiable variables with the potential to inform or support a diagnosis of autism. With this in mind, we endeavoured to provide incremental levels of description of temporal dynamics of speech and gestures, which depended less and less on manual and subjective measures and analyses and more and more on automated and objective measures and analyses. Finally, **the fourth objective** was to evaluate the predictive power of temporal aspects of communication to determine both the diagnosis group and scores on clinical features of autism (ADOS, AQ) for each participant. For this purpose, in the last section of this chapter all variables were therefore pooled together and machine learning algorithms were used to evaluate how accurately we could predict a) the diagnosis of a participant and b) their scores on the ADOS and AQ. Specific predictions will be reported at the beginning of each section, and discussed at the end of each section. A general discussion will then follow at the end of the chapter.

4.3 Participants

The Maras et al. (2013) study provided audio-video recordings of 18 ASD and 18 matched Typically Developing (TD) adults. We contacted the participants again asking for additional consent to use the recordings for the purpose of the current study, which 17 ASD and 17 TD participants agreed to. One TD participant was subsequently excluded from the analysis because he was diagnosed with schizophrenia during the time that separated the two studies, leaving 17 ASD (15 males and 2 females) and 16 matched TD (13 males and 3 females) participants in the groups. All participants were recruited from the Autism Research Group's research participants database. ASD participants held a clinical diagnosis delivered by local health authorities according to DSM-IV (American Psychiatric Association, 2000) criteria for Autistic Disorder or Asperger Disorder, and diagnoses were confirmed for all participants by assessment with the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, Dilavore, & Risi, 2002). All but one participant met the criteria for ASD on the ADOS (the participant who did not meet criteria scored 2 in the communication domain and 3 in the social interaction domain, total score of 5 and was nevertheless included on the basis of their clinical record). TD participants were matched with ASD participants and age, VIQ, PIQ and FIQ. Independent sample t-tests confirmed that groups did not differ significantly on any of these measures (all *p*s > .40). Autism Quotient scores (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) was collected to provide a continuous measure of autistic traits across groups. Table 4.1 presents participants characteristics and t-test statistics.

	ASD (n=17)		**TD (n= 16)**		**t-test**		**Cohen's**
	Mean *(SD)*	*Range*	Mean *(SD)*	*Range*	**t**	**p**	**d**
Age in years	41.1 *(13.58)*	25-62	44.8 *(12.07)*	25-60	-0.824	.416	.287
M/F ratio	15/2		13/3				
Handedness (R/L/A/NA)	13 / 2 / 1 / 1		13 / 1 / 0 / 2				
AQ	32.8 *(7.0)*	21-45	16.9 *(6.3)*	4-28	2.988	.010*	2.384*
ADOS total	9.71 *(3.29)*	5-17					
Comm	2.88 *(1.69)*	0-6					
RSI	6.82 *(2.58)*	3-12					
Im / Crea	1.25 *(0.58)*	0-2					
Sens Behav	1.23 *(1.09)*	0-3					
VIQ	110.35 *(11.21)*	81-123	111.88 *(12.81)*	82-128	-0.364	.718	.127
PIQ	106.71 *(13.51)*	84-128	109.69 *(14.14)*	75-136	-0.619	.540	.215
FIQ	109.69 *(12.35)*	81-122	111.88 *(13.97)*	77-135	-0.537	.595	.166

Table 4.1 Participant characteristics.
Handedness: R = Right-handed; L = Left-handed; A = Ambidextrous; NA = information Not Available; AQ = Autism spectrum Quotient (Baron-Cohen, 1995); ADOS = Autism Diagnosis Observation Schedule (Lord et al., 1999) ; VIQ, PIQ, FIQ = Verbal, Performance and Full-scale Intellectual Quotient (The Psychological Corporation, 2000).

4.4 Procedure

Full details of the original study are available in Maras et al. (2013), but the general procedure is summarised here for ease of reference. In the original study, participants took part in a live event first aid scenario. They were informed that their task was to help the experimenter (AL) carry out some first aid on a manikin, and that the experimenter would instruct them on what they needed to do. The participant and experimenter went into a room to find a life-size manikin lying on the floor, with various staged injuries. Pictures on the wall indicated that the context of the event was a fictitious car crash, and a number of items were present in the room (hazard triangle, first aid kit, blanket, foil blanket). The experimenter proceeded to guide the participant through a scripted series of first aid actions (for instance: putting out the hazard triangle, rolling a bandage up, placing a blanket under the manikin's leg). Following the event, participants engaged in unrelated tasks for an hour, after which they were interviewed by a different experimenter (KM) in a different, quiet room. The interview followed the standard structure recommended by the government for professionals who interview witnesses. The participant first was instructed to take their time and to recall as much as they could from the scenario, prompting a monological free report of the events by the interviewee without any further prompts or interruptions (free recall, part 1). The interviewer then gave one prompt ("can you remember anything else?") to encourage participants to provide any extra information (free recall, part 2). This was followed by a dialogical Q&A session in which the interviewer asked more detailed questions about elements that the interviewee had evoked in the free recall part.

4.5 Recordings

All interviews were audio and video recorded using a camera placed on a tripod close to the interviewer and pointing to the participant, so that it captured the participant's production from a point of view as close as possible to the interviewer's. The distance between the camera and the participant was approximately 1 meter but was not measured exactly, because the original purpose for recording was only for speech coding. The nature of the first aid task and the fact that participants had taken active part in the scenario provided a strong occasion for eliciting rich reports, both in terms of speech and gestures. Independent t-tests indicated that the groups did not differ in term of the length of free recall produced, either for the full free recall or for the first part of free recall until they got a prompt (see table 4.2).

Recording length (sec)	ASD (n=17)			TD (n= 16)			t-test		Cohen's
	Mean	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	**t**	**p**	**d**
Full free recall	382.9	177.2	186-733	361.4	124.5	165-577	.401	.691	.140
Free recall part 1	321.5	148.8	117-642	305.7	105.4	149-522	.351	.728	.123

Table 4.2 Recordings length in seconds.
Mean, standard deviation and range for each group. Full free recall: entire free recall production including answer to the initial question and one prompt. Free recall part 1: free recall production following the initial question up until the prompt.

For all subsequent analyses / features extraction, the first 10s of each time-series and the second part of the free recall (including the experimenter's prompt and the participant's reply to prompt) were discarded to eliminate the majority of productions from the experimenter (some recording presented occasional back-channelling or prompt, but this represented a minority of the audio signal).

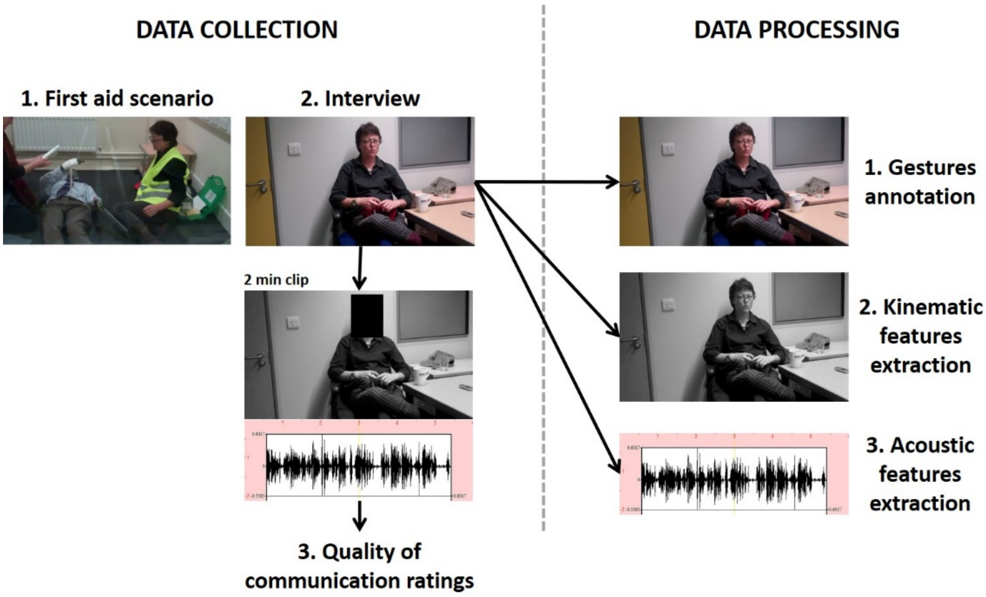

Figure 4.1 Experimental procedure for data collection and data processing[1].
Data collection: 1. Participant takes part in a first aid live-event scenario with AL. 2. Participant is interviewed in a separate room by KM about what happened in the scenario. 3. After faces are masked, audio-only and audiovisual recordings are rated for quality of communication. Data processing: 1. Video recording is compressed and each gesture is annotated on a frame-by-frame basis. 2. Video recording is converted to black and white and kinematic features are extracted. 3. Acoustic features are extracted from the audio recording.

[1] All images are provided with the participant's permission.

4.6 Gesture annotation

Two coders (AL and an independent coder JK) annotated the video recordings of the first part of the free recall using ANVIL (Kipp, 2001). AL annotated 21 recordings and JK annotated 13 recordings. For each recording, the coder identified each gesture's **onset time**, **offset time** and **stroke time** (defined as the moment of the gesture with most emphasis, often the maximum extension in space or the point of fastest acceleration) with precision using frame by frame display; the **type of gesture** (iconic, metaphoric, deictic or beat); the **referent** (for iconic and metaphoric gestures); a **rating of confidence** in identifying the referent (on a 1-5 scale); the **bodypart** used to execute the gesture (finger(s), hand, wrist, forearm, arm, shoulder or other); the **handedness of the gesture** (left, right, left and right or none); the **target** (for deictic gestures: self, Anna or the manikin, position, none or other); and the **number of repetitions** of the same gesture (e.g., to illustrate rolling a bandage participants typically rolled their finger in the air a number of times).

4.7 Quality of communication ratings

Quality of communication ratings were collected for 16 out of 17 ASD participants, and all 16 TD participants because 1 ASD participant only gave permission to use their recordings in the context of this study after this section was completed.

Recordings processing: The video recordings were converted to grayscale and edited down to a 2-minute clip using VirtualDub (http://www.virtualdub.org/). The 2-minute section was selected by browsing through the video to find a section containing a good number of gestures (relative to the participant filmed) and was created by cutting out the beginning and the end of the recording (not by combining different sections of the recording). A dynamic, black rectangular mask was placed over the participants' faces in the video recordings using Wax 2.0., both to ensure the participants' anonymity and also so that the visual information accessible to the viewer would relate mainly to gestures and not facial movements. Finally, 3-second-long fade in and fade out transitions (fade from/to black for the video track, fade from/to silence for the audio track) were added using AviSynth (http://avisynth.nl/) at the beginning and the end of the clip respectively, to avoid splitting the recording abruptly in unnatural places. The corresponding 2 min auditory track was exported from the new 2 minute clip.

Ratings collection: 30 naïve judges were recruited from City, University of London Sona participants sign-up system and divided in two groups. One judge in group 1 was replaced because they were missing data for six of the recordings. Another judge in group 1 was excluded from the analysis because they had difficulty hearing most participants and showed a tendency to rate all questions similarly and used very low scores compared to the rest of the judges. Therefore in the final analysis group 1 consisted of 14 judges (6 female, mean age 25.4 ± 5.5) and group 2 consisted of 15 judges (7 female, mean age 27.5 ± 7.4). Groups did not differ in age ($t(27) = -.86$, $p > .3$, Cohen's $d = .32$).

Ratings were collected in group sessions where the recordings were projected on a large screen and played through the university sound system, at the same volume and in the same amphitheatre for all groups. Judges were widely spread across seats in the room so that they could not see each other's ratings. They were first informed that they would be watching audio-only and audio-video recordings of participants recalling the same live first aid scenario, and that their faces would be masked for confidentiality. It was explained that their task was to assess the quality of communication displayed in each recording on 6 items. Each item was defined and clarified where needed. Judges were unaware of the fact that participants in the recordings belonged to particular diagnostic groups until they were debriefed at the end of the session.

Each group was presented with half the recordings in the audiovisual modality and the other half in the auditory modality only, with the modality counter-balanced between group so that each recording was presented in both modalities to an equal number of judges. For each recording, judges rated 6 items on a 7-point Likert scale assessing comprehension (2) and quality of communication (4):

<u>Comprehension:</u>

1. How well were you able to follow what the person was saying?

2. Was the person's report organised in a clear sequence of events?

Quality:

3. How well did the person express himself/herself?

4. Was the person speaking fluently and clearly?

5. How engaged were you while listening to the recording?

6. How well could you picture the scene based on the person's description?

Two questions (1 and 5), which focussed on the listener, were taken from de Marchena and Eigsti's study (2010). Three questions (2, 3, 4) were added to assess the perceived quality of communication demonstrated by the speaker in the recording. One question (6) specifically addressed how well the listener could "picture" the scene, an aspect that we expected to be directly related to the production of gestures.

4.8 Acoustic features extraction

Regularly sampled time-series of fundamental frequency f0 (in Hz) and intensity (in dB) were extracted from the audio recordings every 50ms using Praat (http://www.praat.org/). Pitch was bandpass filtered at 50-700 Hz and intensityF was bandpass filtered at 0-75dB. Voice/pause behaviour was extracted as a binary variable every 50ms (1 for speech, 0 for no speech) using Praat, with a sampling rate of 50ms.

4.9 Kinematic features extraction

In order to obtain a continuous variable quantifying the amount of movement produced, we measured the motion energy of the recordings, defined as the difference in grayscale pixels between consecutive video-frames (Grammer, Honda, Juette, & Schmitt, 1999). The original video recordings were resampled to 20fps and compressed to 800x720 px to increase processing speed and converted to grayscale for movement analysis using VirtualDub. The time series of motion energy was then computed as the number of pixels that changed in luminance between frame n and frame n+1 (see figure 4.2), with a minimum threshold of 30 pixels (to discard changes due to light flicker). In the following chapter motion energy will be abbreviated as ME.

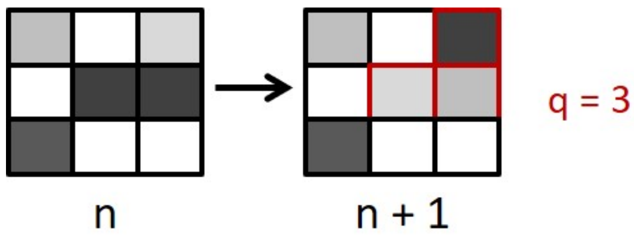

Figure 4.2 Motion energy extraction procedure.
The video recording is converted to grayscale, then read frame-by-frame and the motion energy is computed as the number of pixels that change in luminance between frame n and frame n+1.

Motion/hold behaviour was extracted as a binary variable every 50ms. After inspection of individual motion energy profiles, we used a threshold of 1.25% of the range of ME to define motion (> 1.25%) and holds (≤ 1.25%). For instance, if a participant produced motion energy in the range of 0-30,000, any datapoint with ME ≤ 375 was counted as a hold.

Part B – Quality of Communication

First, we set out to assess the quality of communication displayed in the recordings, in audio-only and audiovisual conditions. Based on previous research, which found that autistic productions were less clear to follow and associated with reduced level of the listener's engagement compared to non autistic productions (de Marchena & Eigsti, 2010; García-Pérez et al., 2007; Morett et al., 2016), we predicted that quality of communication ratings would be lower in the ASD compared to the TD group, across conditions. Second, we predicted that gestures would improve the reported quality of communication in both groups, with higher scores in the audiovisual than in the audio-only condition. Finally, we predicted that we would replicate de Marchena and Eigsti's (2010) finding that gesture improve quality of communication ratings less in the ASD compared to the TD groups, with a smaller gain between audio-only and audiovisual conditions in the ASD group.

Results in this section are reported for 16 ASD and 16 TD participants because 1 ASD participant gave permission to use their recordings for the purpose of this study only after the quality of communication ratings were collected.

4.10 Reliability

We measured the internal consistency of the 6-item quality of communication scale by computing Cronbach's alpha on individual judges' responses, which yielded a high reliability score of .987.

Next we measured inter-rater reliability by computing the intraclass correlation coefficient (ICC) for judges in group 1 (14 judges) and group 2 (15 judges) separately. The average measure ICC for group 1 was .785 with a 95% confidence interval ranging from .723 to .835 ($F(167,2171) = 5.895$, $p < .001$). The average measure ICC for group 2 was .875 with a 95% confidence interval ranging from .836 to .906 ($F(160,2240) = 10.481$, $p < .001$). Both groups of judges therefore showed acceptable to good inter-rater reliability.

4.11 Average quality of communication

For each recording, we computed the average ratings for each of the quality of communication scale item, for audiovisual and audio-only modalities separately. We also averaged single-item scores into a composite total score (see Figure 4.3).

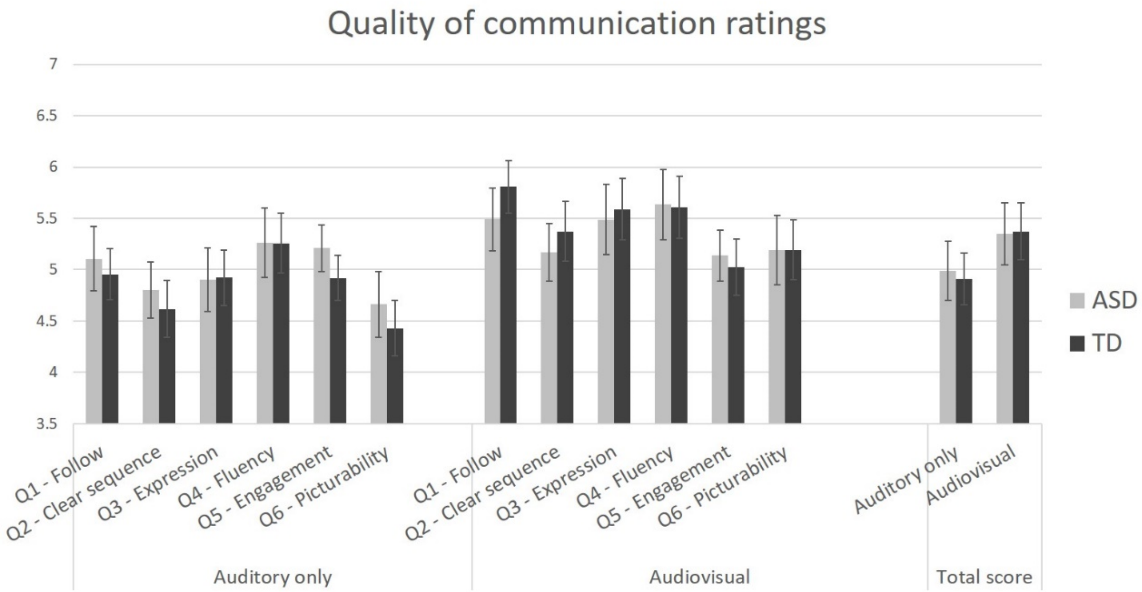

Figure 4.3 Mean quality of communication ratings in the audio-only modality and audiovisual modality on a scale of 1-7.
The left and middle panel show ratings per individual scale item, whilst the right panel show the total average score. Q1: How well were you able to follow what the person was saying? Q2: Was the person's report organised in a clear sequence of events? Q3: How well did the person express himself/herself? Q4: Was the person speaking fluently and clearly? Q5: How engaged were you while listening to the recording? Q6: How well could you picture the scene based on the person's description?

Shapiro-Wilk tests revealed multiple violations of the normality assumption for the ratings data in both groups. In the ASD group, ratings on Q1 ($D(16)=.837$, $p<.01$), Q2 ($D(16)=.782$, $p<.005$), Q3 ($D(16)=.829$, $p<.01$), Q4 ($D(16)=.765$, $p<.005$), Q6 ($D(16)=.844$, $p<.05$), and the total score ($D(16)=.797$, $p<.005$) in the auditory modality, and ratings on Q1 ($D(16)=.850$, $p<.05$), Q2 ($D(16)=.844$, $p<.05$), Q3 ($D(16)=.846$, $p<.05$), Q4 ($D(16)=.822$, $p<.01$), Q6 ($D(16)=.868$, $p<.05$) and total score ($D(16)=.832$, $p<.005$) in the audiovisual modality showed a negative skew. In the TD group, ratings on Q1 ($D(16)=.844$, $p<.05$), Q2 ($D(16)=.872$, $p<.05$), Q3 ($D(16)=.853$, $p<.05$), Q4 ($D(16)=.873$, $p<.05$), and total score ($D(16)=.802$, $p<.005$) in the audiovisual modality showed either a negative skew or a platykurtic distribution. Levene's test of homogeneity indicated that variance was equal between groups or all scores.

Mann-Whitney U tests were conducted to compare the ratings between groups. The results indicated no significance difference between any of the ratings between ASD and TD speakers (all ps >.3). The inferential statistics are reported in Table 4.3. A Benjamini-Hochberg procedure was applied to correct for family-wise error rate. Ratings were significantly different between the audio-only and audiovisual modalities for all items except Q5 (engagement rating), revealing that overall the quality of communication was rated higher for audiovisual than audio-only recordings (ASD total scores: mdn_{Aud} = 5.25, mdn_{AV} = 5.69; TD total scores: mdn_{Aud} = 4.47, mdn_{AV} = 4.99).

Group	Item	z score	p value	Effect size r
ASD	Q1 – Follow	2.172	0.030*	0.384
	Q2 – Clear sequence	2.741	0.006*	0.485
	Q3 – Expression	2.999	0.003*	0.530
	Q4 – Fluency	2.327	0.020*	0.411
	Q5 – Engagement	-0.414	0.679	-0.073
	Q6 – Picturability	2.689	0.007*	0.475
	Total score	2.689	0.007*	0.475
TD	Q1 – Follow	3.464	0.001*	0.612
	Q2 – Clear sequence	3.516	0.000*	0.622
	Q3 – Expression	3.413	0.001*	0.603
	Q4 – Fluency	2.301	0.021*	0.407
	Q5 – Engagement	0.454	0.650	0.080
	Q6 – Picturability	3.067	0.002*	0.542
	Total score	1.965	0.049*	0.347

Table 4.3 Statistics from related sample Wilcoxon signed-rank test conducted on quality of communication ratings between auditory-only and audiovisual conditions in each group, for each item.

* Statistically significant after applying the Benjamini-Hochberg procedure to correct for family-wise error rates.

Q1: How well were you able to follow what the person was saying? Q2: Was the person's report organised in a clear sequence of events? Q3: How well did the person express himself/herself? Q4: Was the person speaking fluently and clearly? Q5: How engaged were you while listening to the recording? Q6: How well could you picture the scene based on the person's description?

4.12 Visual gain

Second, we wanted to test whether watching gestures improved the perceived quality of communication similarly in both groups. For each question and for the total composite score, we computed a measure of visual gain, which was simply the difference between the score in the audiovisual modality and the score in the audio-only modality. These values are presented in Figure 4.4. A positive gain indicated that the quality rating increased when the visual track was provided alongside the audio track.

Shapiro-Wilk tests indicated that gain scores for all items were normally distributed in both groups. Levene's test of homogeneity indicated that variance between groups was equal for all scores.

One sample t-tests showed that the gain in quality of communication across groups was significantly different from zero for all questions ($p \leq .001$) except Q5 ($p > .5$), confirming that overall quality of communication improved with visual information.

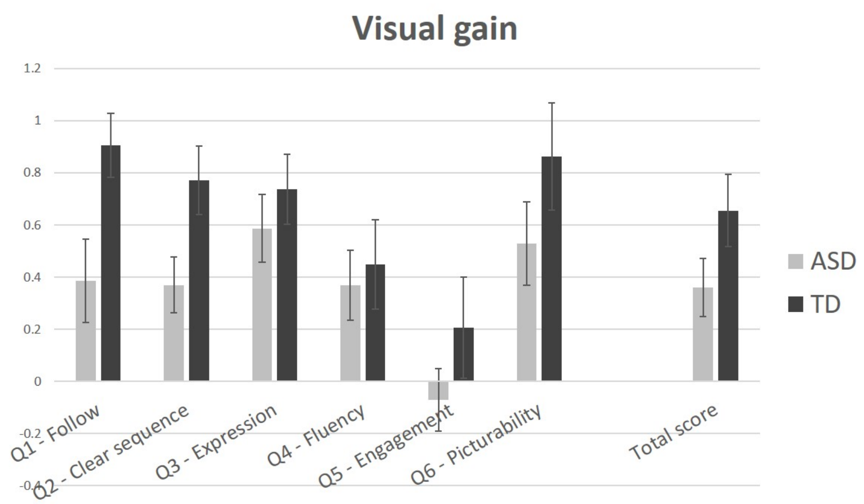

Figure 4.4 Mean visual gain in quality of communication ratings per question.
Gain = difference in quality of communication ratings between audiovisual and audio-only modalities. Q1: How well were you able to follow what the person was saying? Q2: Was the person's report organised in a clear sequence of events? Q3: How well did the person express himself/herself? Q4: Was the person speaking fluently and clearly? Q5: How engaged were you while listening to the recording? Q6: How well could you picture the scene based on the person's description?

Planned paired t-tests were conducted to compare the gain between groups. The Benjamini-Hochberg procedure was applied to correct for multiple comparison. Results are reported in table 4.4 below. Statistics showed that the gain was significantly larger with large effect sizes, in the TD compared to the ASD group for questions Q1 and Q2, which were related to aspects of the comprehension (both ps <

.05). Notably, although comparisons did not reach significance level, the group differences for the total score as well as for Q5 and Q6 showed moderate effect sizes, again indicating a trend that the visual gain was larger for TD compared to ASD participants.

		ASD (n=16)	TD (n= 16)	t-test		Cohen's
		Mean *(SD)*	Mean *(SD)*	t	p	d
Visual gain (AV-Aud)	Total score	0.36 (.45)	.66 (.55)	-1.65	.109	.59
	Q1 – Follow	0.39 (.64)	0.91 (.45)	-2.56	.016*	.94
	Q2 – Clear sequence	0.37 (.43)	0.77 (.52)	-2.38	.024*	.84
	Q3 – Expression	0.59 (.52)	0.74 (.54)	-.80	.429	.28
	Q4 – Fluency	0.37 (.53)	0.45 (.68)	-.37	.717	.13
	Q5 – Engagement	-0.07 (.48)	0.21 (.78)	-1.21	.235	.43
	Q6 – Picturability	0.53 (.64)	0.86 (.83)	-1.27	.212	.45

Table 4.4 Independent t-test statistics for mean visual gain analysis.
Gain = difference in quality of communication ratings between audiovisual and audio-only modalities. Q1: How well were you able to follow what the person was saying? Q2: Was the person's report organised in a clear sequence of events? Q3: How well did the person express himself/herself? Q4: Was the person speaking fluently and clearly? Q5: How engaged were you while listening to the recording? Q6: How well could you picture the scene based on the person's description?

4.13 Relationship between quality of communication and clinical profile

We assessed the relationship between quality of communication and clinical features by computing Spearman's correlation coefficients between our total auditory and audiovisual quality of communication score and total visual gain score on the one hand, and total ADOS score, the ADOS subscores in the communication and reciprocal social interaction domains (for ASD group only), and the AQ (to have a continuous measure between groups). The Benjamini-Hochberg procedure was applied to correct for family-wise error rate. There were no significant correlations between any of the variables (≥.05), suggesting that perceived quality of communication was not directly associated to clinical severity in the ASD group, or to autistic traits in any of the two groups.

4.14 Summary and provisional discussion

Overall, there was no main difference in the quality of communication ratings between the ASD and TD groups. This is not consistent with García-Pérez et al.

(2007), de Marchena & Eigsti (2010) and Morett and colleagues (2016) who found that stories narrated by ASD participants were judged to be less easy to follow and less engaging, and generally does not provide support for the notion that quality of communication as a whole is poorer in autism. The incongruency with previous studies could be accounted for by the type of task the participants engaged in. In both de Marchena and Eigsti's and Morett et al.'s studies, participants had to tell a fun story based on animal characters either from a cartoon or a short video (ADOS materials and an episode of Tweety and Sylvester, respectively), whilst in the current study participants were recalling and reporting a real event that they personally took part in. Reduced quality of communication in previous studies could, therefore, be accounted for by reduced imaginative skills (Crespi, Leach, Dinsdale, Mokkonen, & Hurd, 2016) and difficulties with theory of mind (Frith & Happé, 1994) in autism rather than difficulties with aspects of communication per se. Specifically, autistic adults might provide narratives that are less clear and less engaging than their TD counterparts when the task relies on their imaginative skills and theory of mind to "fill in the gaps" and imagine what the characters in the story are thinking or intending to do. In the current study by contrast, the emphasis was on factual aspects of the event, rather than intentionality and interpretation of behaviour, as the interviewer was mainly interested in what actions had happened and who had performed them. This hypothesis could be tested by directly contrasting communicative behaviours based on the generation of narratives versus the retrieval of personal experiences. In the case of García-Pérez et al.'s (2007) study, they used recordings taken from 12 recordings of adolescents with and without ASD interviewed about themselves. The interviews touched on emotional content, which again might have put ASD adolescents at a disadvantage (e.g., Gaigg, 2012). An alternative hypothesis is that the difference between ours and previous results could be due to the demographics of participants: our sample consisted of adults whereas all three previous studies tested children and adolescents. It is therefore possible that the development of communicative skills are simply delayed in ASD, explaining why differences observed in adolescence are no longer detected in adulthood. Finally, a limitation of this study is that the 2-minute clips extracted from each recording for the purpose of rating quality of communication were subjectively selected to include a maximum amount of gestures. This might have resulted in an overestimation of the overall quality of communication of individuals who typically produce less gestures. We will see however that the overall number of gestures did not significantly differ between groups, therefore this would not necessarily have affected the current results.

As expected, however, the quality of communication displayed in audiovisual recordings was rated as higher than the one displayed for audio-only recordings, showing that the visual modality contributed positively to the perceived quality of communication. In particular, access to visual information increased the ratings of how well the listener could picture the scene based on the participant's report, how easy the narrative was to follow overall and how well the participant expressed themselves. Because participants' faces were masked, the visual information consisted of contextual information and body movements, including gestures. The relation between gestures and picturability is intuitive and it was expected that "seeing" the participants' descriptions and having access to iconic (illustrative) and deictic gestures (showing the positions of persons and objects) in particular would improve how well the listener could imagine the scene. Interestingly, visual information also benefitted the listener's comprehension and their perception of the speaker's fluency and expression, showing that gestures broadly enhanced the communicative quality of speech including structure and perceived fluency. Only the listener's reported level of engagement did not differ between modalities, which may have been due to the experimental conditions: judges were requested to pay close attention to all recordings and this was supported by their task to come up with quality of communication ratings. This was also the case in Morett and colleagues' study who found no difference in level of engagement for audio-only vs audiovisual conditions. However, contrary to the current study, Morett et al. found no modality difference for ratings of narrative coherency (how easy it was to follow). A number of factors can explain this discrepancy: the ratings in Morett and colleagues's are produced by two judges only, so their results might have been underpowered to show a modality effect on ratings of coherency. In addition, the material they used to generate narratives (video stories of Tweety and Sylvester) were short and probably familiar to the judges, so comprehension might not have relied as heavily on visual information. Finally, and importantly, in Morett and colleagues's study participants were instructed to try and remember the story whilst they watched the videos as they would have to tell them to the experimenter later on. It is therefore possible that participants were able to better structure their narratives, whereas in the current study participants were unaware of the fact that they would have to recall and report the live event. In addition the first aid scenario was longer and unknown to the listener, which could have exacerbated differences in the quality of reporting the event as a narrative.

By examining the actual difference in ratings between the audio-only and audiovisual conditions (or visual gain), we were able to show that visual information improved the listener's comprehension ratings, but less so in the ASD compared to the TD group. Based on the assumption that gestures are the main source of communicative information in the visual domain, this suggests that gestural information in autistic participants is not as efficient at improving the listener's comprehension and how clear they perceived the report to be. This could be either because ASD participants produced less gestures overall, or because the quality of the gestures or their integration in the communication system are atypical. Surprisingly, we found no significant correlation between any of the quality of communication ratings or visual gain and the ADOS scores, offering no support to the idea that qualitative differences might contribute to the clinical picture of ASD. This differs from de Marchena and Eigsti's study, which found a negative correlation between the ratings of how engaged the judges reported to be during the story and the ADOS reciprocal social interaction and communication scores. There was a trend showing that greater visual gain was associated with lower AQ score, but the result was small in size and at risk to be a false positive, so it would need to be replicated before we could accept it as a result. More research is needed to establish whether or not quality of communication ratings relate to clinical criteria in ASD.

The next three sections will systematically examine gestures and temporal aspects of the coordination between speech and gestures to provide a better understanding of how speech-accompanying gestures contribute to communication in autism and how it might differ from the typical population.

Part C – Descriptive Analysis

In the following three sections, temporal aspects of speech and gesture were explored at different levels, in a gradually less subjective/manual way. First, we offer a more traditional analysis of the data and report temporal characteristics of both speech and gesture using descriptive statistics. Using this information, and in continuity with previous studies, we next move onto exploring the temporal coordination between characteristic timepoints in speech and characteristic timepoints in gesture. We then move away from subjective annotation of the data and investigate the temporal coordination of acoustic and kinematic features as continuous time series using cross-correlation. Finally, we remove ourselves from a linear perspective on the dynamics of speech and gesture and propose a quantification of the degree of recurrence within the various time series.

In order to describe the corpus and compare it to that of previous studies, we first report descriptive statistics on both acoustic and kinematic variables, as well as variables obtained from the gestures annotation. Based on previous research (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017) and in line with historical reports describing autistic speech as "monotonous" and "robotic" (Wing, 1981), we predicted that mean pitch and pitch range would differ between groups, and that mean pitch would correlate with diagnostic features of ASD. To our knowledge, no other study has quantified motion energy in autism, so the group comparison here was exploratory.

Turning to speech/pause behaviour, evidence from the turn-taking literature suggest that autistic children leave a longer gap before taking a conversational turn (Feldstein, Konstantareas, Oxman, & Webster, 1982; Heeman et al., 2010) and that autistic adolescents and adults produce more pauses between and within utterances (Mitchell, 2015; Morett et al., 2016). Previous studies also report a greater amount of empty pauses with durations over 2s (Lake, Humphreys, & Cardy, 2011; Morett et al., 2016). Finally, unpublished data suggest that increased amount of pauses was associated with poorer quality of communication ratings (Mitchell & Volden, 2015). We therefore predicted that the overall proportion and mean duration of pauses would be higher in recordings of ASD participants compared to TD participants, and that both proportion and mean duration of pauses would correlate negatively with quality of communication ratings. Again, there is, to our knowledge, no evidence regarding the patterns of holds in autism. Because speech pauses and motion holds have been shown to overlap and to be distributed similarly in typical productions (Esposito & Esposito, 2011; Esposito & Marinaro, 2007), we tentatively predicted that holds would mirror pause behaviour, with a greater proportion, and longer mean duration of holds in the ASD group compared to the TD group, and that both measures would correlate negatively with quality of communication ratings.

Regarding the overall amount of different types of gesture, recent evidence is inconclusive: one study reported no difference between groups on any of the gesture categories (de Marchena & Eigsti, 2010) whilst another one reported fewer metaphorical and beat gestures in ASD compared to TD adolescents narratives (Morett et al., 2016). As a result we did not have strong predictions about the rates of gestures of any types in ASD compared to TD group, but we speculated that because of its nature, our task would elicit a high proportion of iconic gestures, which would give us more power to detect potential group differences in that category. Based on de Marchena and Eigsti (2010), we did not expect group differences in the duration of gestures. Because a core feature of autism is the production of repetitive behaviours, we investigated the number of times a gesture is repeated. For instance, when demonstrating rolling a bandage, the same rolling movement is often repeated a few times. We anticipated that the average number of gesture repetitions would be greater in the ASD group, and that it would correlate positively with the total ADOS score. Although it was not the main focus of this project, we also provided an analysis of gesture handedness to control that any difference in the characterisation of gestures or their coordination to speech could not be attributed to a difference in motor execution.

4.15 Continuous acoustic and kinematic features: Pitch, Loudness and Motion energy

For each participant, **pitch, loudness and motion energy** values were averaged over the first part of the free recall. Shapiro-Wilk tests showed that f0 was not normally distributed in either the ASD ($D(17)=.796$, $p<.005$) or the TD ($D(16)=.839$, $p<.01$) groups, and that loudness was not normality distributed in the TD group

(D(16)=.825, p<.01). Levene's test of homogeneity indicated that variance between groups was equal for all scores.

Mann-Whitney U tests were performed on each of the descriptive measures (statistics for the mean comparisons are reported in table 4.5). Results showed no group difference in the mean pitch or motion energy, but revealed a significant group difference in average speech loudness with a large effect size which hold when female participants were excluded. Specifically, ASD productions were on average louder than TD productions.

	ASD	TD	Mann-Whitney		
	(n=17)	(n= 16)	U	p	r
Pitch (Hz)					
mean *(SD)*	**152.8** *(63.3)*	**132.6** *(40.5)*	116	.488	.125
median	115.9	117.4			
range (min:max)	95.5:286.7	93.3:231.9			
Male only:	(n = 15)	(n = 13)	77	.363	
mean *(SD)*	**145.2** (62.5)	**121.6**			.164
median	115.7	(36.3)110.3			
range (min:max)	95.5:286.7	93.3:231.9			
Loudness (dB)					
mean *(SD)*	**54.4** *(2.22)*	**51.9** *(2.27)*	55	.003**	.508
median	54.2	51.3			
range (min:max)	49.1:57.6	49.5:58.1			
Male only:	(n = 15)	(n = 13)		.002**	
mean *(SD)*	**54.3** (2.26)	**51.5** (1.69)	33		.517
median	54.2	51.3			
range (min:max)	49.1-57.6	49.5:54.9			
ME (px)					
mean *(SD)*	**1709** *(1003.7)*	**1869** *(1228.3)*			
median	1700	1651.9	148	.683	.075
range (min:max)	0:37518	0:32948			

Table 4.5 Continuous acoustic and kinematic features.
Mean, median, standard deviation and range of pitch (in Hz), loudness (in dB) and motion energy (in pixels) for each group, computed over the first part of the free recall. SD = Standard Deviation of the Mean. Range = min:max.

Spearman's correlations revealed no significant association between the average pitch and loudness of speech, and the ADOS and AQ scores in either group (all ps >.05).

4.16 Speech pause behaviour

For each participant, we computed the **percentage of pauses** in the analysed section of recording, the **rate per second** and the **duration** of pauses. Following Campione and Véronis (2002), we computed the proportion of pauses by duration category following a **trimodal distribution**: brief (less than to 200 ms), medium (200 to 1000 ms included) and long (over 1000 ms), illustrated in figure 4.5. One ASD participant was excluded from the following analysis because their percentage of pause behaviour was over 2.5 standard deviation away from the mean of the group (and represented 95% of the recording). Examination of their profile indicated that they showed the lowest IQ across groups (FIQ = 81). Shapiro-Wilk tests and Levene's test of homogeneity indicated that variables were normally distributed, and that the variance between groups was equal for all scores.

Planned independent-sample t-tests conducted on the remaining 16 ASD and 16 TD participants showed a non significant trend for ASD participants to produce a smaller percentage of pauses overall (with a medium to large effect size, $mean_{ASD}$= 66.0%, $mean_{TD}$= 71.7%, t(30)= -1.990, p = .056, Cohen's d = .70). The **pauses produced were on average shorter** in the ASD than in the TD group by about 100ms with a fairly robust effect size ($mean_{ASD}$= 449ms, $mean_{TD}$= 554ms, t(30)= -2.282, $p < .05$, Cohen's d= .81). There was no difference in the rate of pauses between groups, however ($mean_{ASD}$= 1.50/s, $mean_{TD}$= 1.36/s, t(30)= 1.668, $p > .1$).

A mixed model ANOVA with group (ASD, TD) as between subjects factor and pause duration (brief, medium, long) as within subjects factor, revealed a **significant main effect of pause duration** [F(2,32) = 112.06, $p < .001$, $\eta_p^2 = .79$] on the rate of pauses and a **significant group x pause duration interaction** [F(2,32) = 6.83, $p < .05$, $\eta_p^2 = .019$]. Post-hoc paired t-tests confirmed that overall participants produced long pauses significantly less frequently than both brief ($mean_{brief}$= .47/s, $mean_{long}$= .14/s, t(31)= 13.860, $p < .001$) and medium length pauses ($mean_{medium}$= .40/s, $mean_{long}$= .14/s, t(31)= 14.637, $p < .001$), and that they produced more brief than medium duration pauses (t(31)= 2.348, $p < .05$). However ASD participants produced significantly more brief pauses ($mean_{ASD}$= .50, $mean_{TD}$= .43, t(30)= -1.990, $p < .05$. Cohen's d = .81) and less long pauses ($mean_{ASD}$= .12, $mean_{TD}$= .16, t(30)= -2.176, $p < .05$. Cohen's d = .77) than TD participants.

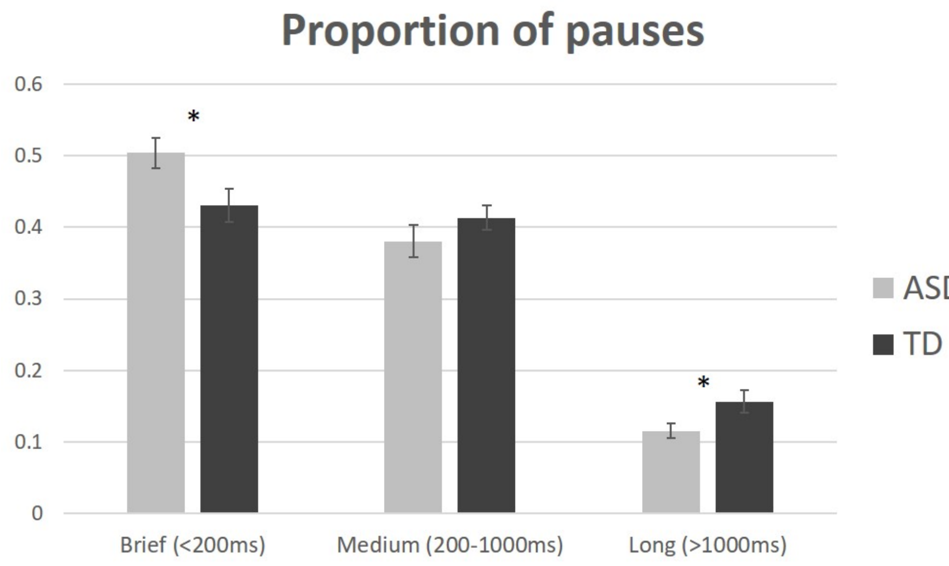

Figure 4.5 Proportion of speech pauses by duration category.
Following Campione and Véronis (2002): brief (< 200ms), medium (200 to 1000ms) and long (> 1000ms) and per group. Error bars represent the standard error of the mean.

Overall, then, the analyses of speech pause behaviour indicated that the proportion of medium length pauses related to cognitive skills both in ASD and TD individuals, whilst the proportion of brief and long pauses differed between diagnostic groups with autistic participants producing more brief and less long pauses compared to TD participants.

4.17 Motion hold behaviour

For each participant, we computed the **percentage of holds** in the analysed section of recording, the **rate per minute** and the **duration** of holds. After examination of the distribution of holds and to mirror the speech pause analysis, we computed the frequency of holds by duration category following a **trimodal distribution**: brief (less than 200 ms), medium (200 to 500 ms included) and long (over 500 ms), illustrated in figure 4.6.

Two ASD participants and one TD participant were excluded from the following analysis because their number, frequency and/or duration of holds were over 2.5 standard deviation away from the mean of their group. Examination of their profile indicated that one ASD participant was spinning on his chair during the recording, which may have produced an unrepresentatively low amount of holds (no movement), whilst the second ASD participant and the TD participant excluded produced very few gestures (lowest or second lowest rate in their respective groups), again an unrepresentatively high amount of hold behaviour.

Planned independent-sample t-tests conducted on the remaining 15 ASD and 15 TD participants showed that **ASD participants produced holds more frequently than TD overall** (with a medium to large effect size, $mean_{ASD}$= 1.16/s, $mean_{TD}$= 1.00/s, t(28)= 1.855, p = .074, Cohen's d = .68), but this was not significant. The percentage of hold behaviour ($mean_{ASD}$= 44.8%, $mean_{TD}$= 48.6%, t(28)= -.467, p > .6) and the mean duration of holds ($mean_{ASD}$= 391ms, $mean_{TD}$= 490ms, t(28)= -1.329, p > .1) did not differ significantly between groups.

A mixed model ANOVA with group (ASD, TD) as between subjects factor and hold duration (brief, medium, long) as within subjects factor, revealed a **significant main effect of hold duration** [F(2,30) = 263.312, p < .001, η_p^2 = .90] on the frequency of holds. The group x hold duration interaction was not significant [F(2,30) = .390, p > .6, η_p^2 = .014]. Post-hoc paired t-tests confirmed that overall participants produced brief holds significantly more frequently than both medium ($mean_{brief}$= .67/s, $mean_{medium}$= .16/s, t(29)= 22.830, p < .001) and long duration holds ($mean_{brief}$= .67/s, $mean_{long}$= .18/s, t(29)= 14.384, p < .001). The frequency of medium and long holds did not differ significantly (t(29)= -1.605, p > .1).

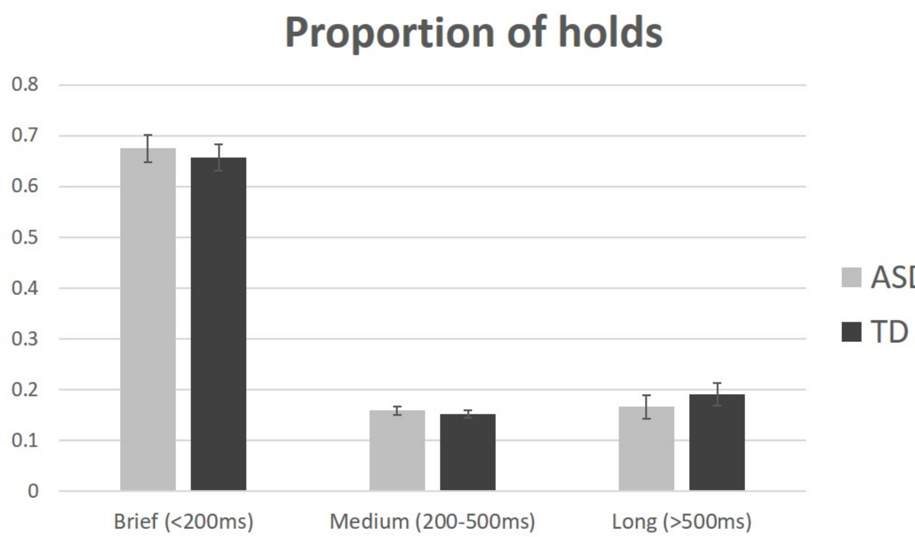

Figure 4.6 Proportion of motion holds by duration category.
Brief (< 200ms),medium (200 to 500ms) and long (> 500ms) and per group. Error bars represent the standard error of the mean.

4.18 Gesture coding

For each of the four categories of gestures (iconic, metaphoric, deictic and beats), we computed the mean and standard deviation of the following variables: total number, rate (number of gestures per minute), average duration, average number of repetitions and percentage of right-handed, left-handed and two-handed gestures (see Table 4.6).

	ASD (n=17)				TD (n= 16)			
	Iconic	*Meta*	*Deictic*	*Beat*	*Iconic*	*Meta*	*Deictic*	*Beat*
Number	**27.59**	**21.06**	**5.76**	**8.82**	**36.25**	**29.50**	**9.38**	**7.38**
(SD)	*(31.34)*	*(29.50)*	*(6.85)*	*(12.81)*	*(34.78)*	*(19.35)*	*(12.09)*	*(7.99)*
range	0-120	0-115	0-25	0-54	2-134	0-53	0-40	0-32
Rate (/min)	**4.88**	**3.42**	**1.19**	**1.79**	**6.61**	**4.08**	**1.58**	**1.39**
(SD)	*(4.72)*	*(3.63)*	*(1.26)*	*(2.70)*	*(5.37)*	*(3.92)*	*(1.77)*	*(1.26)*
Range	0-16.07	0-13.56	0-3.80	0-11.45	0.77-18.87	0-14.92	0-6.00	0-4.34
Duration (s)	**2.17**	**1.42**	**1.53**	**1.02**	**1.72**	**1.47**	**1.74**	**1.06**
(SD)	*(0.62)*	*(0.48)*	*(0.75)*	*(0.99)*	*(0.35)*	*(0.55)*	*(0.70)*	*(0.99)*
range	1.35-3.51	0.83-2.50	0.72-3.26	0.24-4.01	1.31-2.28	0.74-2.75	0.82-3.13	0.34-3.46
Repetitions	**2.07**	**1.86**	**1.78**	**1.44**	**1.79**	**1.56**	**1.78**	**1.45**
(SD)	*(0.57)*	*(0.77)*	*(1.30)*	*(0.97)*	*(0.37)*	*(0.33)*	*(0.73)*	*(0.90)*
range	1-3	1-3.67	1-6	1-4.79	1-2.4	1-2.2	1-3.5	1-4.33
Handedness								
% Right	**36.08**	**40.17**	**64.78**	**44.62**	**33.89**	**29.49**	**49.84**	**30.13**
(SD)	*(22.45)*	*(30.27)*	*(33.41)*	*(41.36)*	*(25.92)*	*(27.36)*	*(32.22)*	*(32.39)*
range	7.14-100	0-100	0-100	0-100	0-80	0-82.35	0-85.71	0-100
% Left	**17.72**	**17.08**	**14.27**	**25.25**	**15.28**	**11.94**	**30.06**	**33.37**
(SD)	*(17.81)*	*(21.22)*	*(20.19)*	*(29.42)*	*(18.98)*	*(20.20)*	*(23.17)*	*(32.44)*
range	0-66.67	0-60	0-62.5	0-100	0-66.67	0-78.72	0-71.43	0-100
% R+L	**45.67**	**41.04**	**9.34**	**29.80**	**50.18**	**57.61**	**18.18**	**36.50**
(SD)	*(22.78)*	*(30.79)*	*(11.00)*	*(32.64)*	*(30.21)*	*(34.47)*	*(26.95)*	*(32.03)*
range	0-80	0-100	0-33.33	0-100	0-100	4.26-100	0-100	0-86.67

Table 4.6 Descriptive statistics (mean and standard deviation) of gesture characteristics by diagnostic group (ASD, TD) and type of gesture (iconic, metaphoric, deictic, beat).
Total number of gestures, rate (number of gestures per minute), duration (in seconds), average number of repetitions for each gesture, percentage of right-, left-, and two-handed gestures.

4.18.1 Gestures rate

Because the duration of the recording varied between participants, analyses were performed on the rate of gestures per minute rather than the overall number of gestures counted over the duration of the recording. Shapiro-Wilk tests revealed that most of the variables were not normally distributed. Mann-Whitney independent-sample U tests were performed to compare gesture rates between groups. There was no significant group difference for any of the gesture types (all ps > .3).

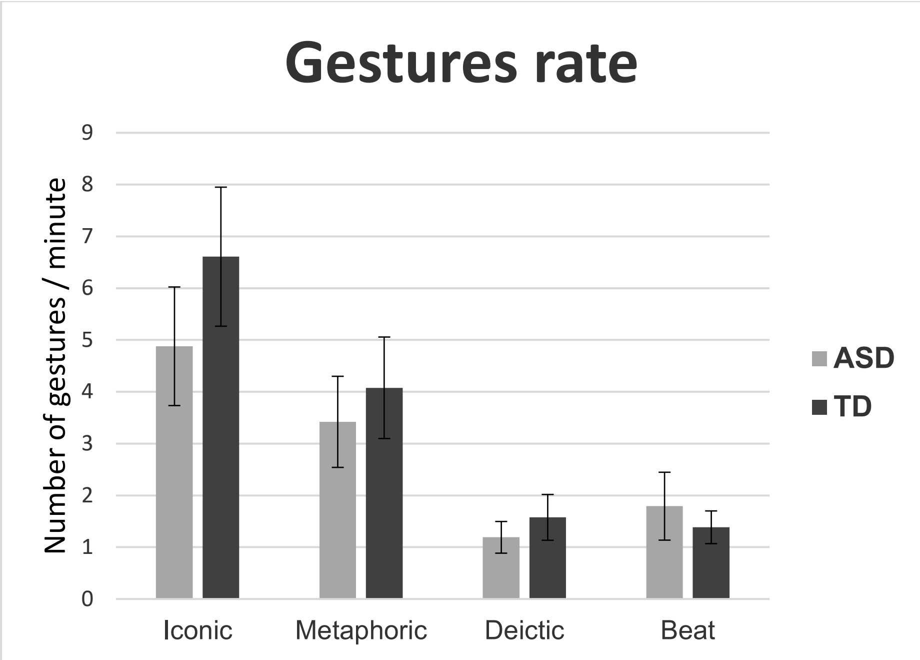

Figure 4.7 Mean gesture rate (in gesture/minute) by diagnostic group (ASD, TD) and type of gesture (iconic, metaphoric, deictic, beat).
Error bars represent the standard error of the mean.

4.18.2 Gestures duration

Shapiro-Wilk test revealed that the mean duration of beat gestures were not normally distributed in either group (ASD: D(12) = .677, p<.005, TD: D(12) = .735, p<.005) but showed a positive skew. To correct for this, a log transformation was applied to the data. Analyses were first conducted with the untransformed data and then a second time with the transformed data. The pattern of findings remained the same in all analyses. Because transformed data make it difficult to interpret results, and because ANOVAs are considered to be fairly robust to deviations from normality (e.g., Schminder, Ziegler, Danay, Beyer, & Bühner, 2010) especially when group sizes are equal, findings from the original (untransformed) data are presented below. A mixed model ANOVA with group (ASD, TD) as between subjects factor and type of gesture (iconic, metaphoric, deictic, beat) as within subjects factor, revealed a **significant main effect of gesture type** [$F(3,24) = 7.37$, $p < .001$, $\eta_p^2 = .25$] on the mean gesture duration. However, the main effect of group [$F(1,24) = .005$, $p > .9$, η_p^2

= .00] and the group x gesture type interaction [$F(3,24) = .72$, $p > .5$, $\eta_p^2 = .03$] were not significant.

Post-hoc paired-sample t-tests indicated that overall iconic gestures lasted significantly longer than metaphoric gestures ($t(28) = 4.65$, $p < .001$), deictic gestures ($t(26) = 2.31$, $p < .05$) and beat gestures ($t(27) = 4.91$, $p < .001$), and beat gestures were significantly shorter than metaphoric gestures ($t(25) = 2.21$, $p < .05$) and marginally shorter than deictic gestures ($t(24) = 1.98$, $p = .06$) but the duration of metaphoric and deictic gestures did not differ significantly ($t(25) = -1.65$, $p > .1$). Mean gesture durations are illustrated in Figure 4.8.

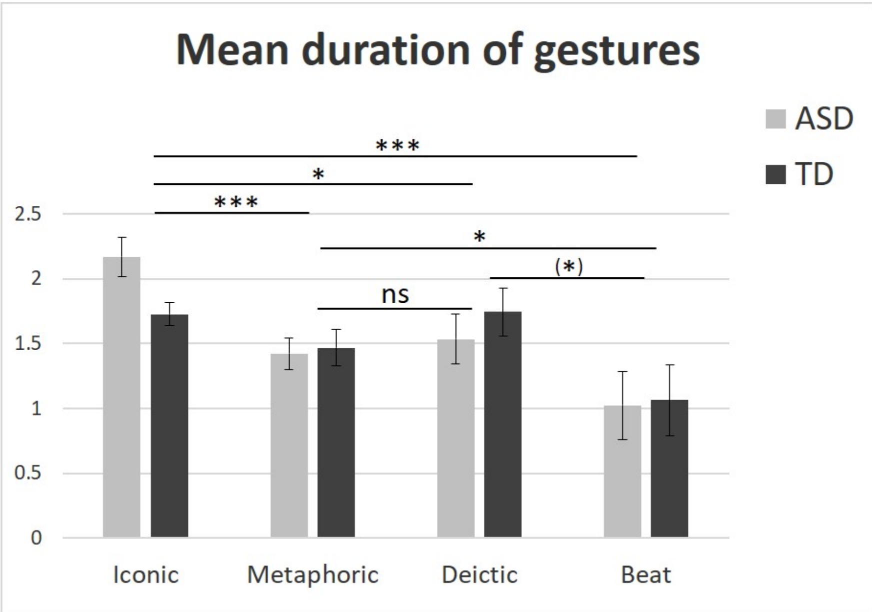

Figure 4.8 Mean duration of gestures (in seconds) by diagnostic group (ASD, TD) and type of gesture (iconic, metaphoric, deictic, beat).
Asterisks indicate the level of significance in paired sample comparison (* < .05; ** < .01, *** < .005). Error bars represent the standard error of the mean.

4.18.3 Gestures repetition

Shapiro-Wilk tests revealed several violations of distribution normality for the number of repeated gestures: in the ASD group this was the case for metaphoric gestures ($D(12)=.835$, $p<.05$), deictic gestures ($D(12)=.638$, $p<.001$) and beat gestures ($D(12)=.652$, $p<.001$). In the TD group this was the case for beat gestures only ($D(12)=.609$, $p<.001$). Mann-Whitney U tests found no significant group difference for any of the gesture types (IG: $U(27)=212.5$, $p>.05$; MG: $U(27)=86.5$, $p>.4$; DG: $U(27)=109$, $p>.4$; BG: $U(27)=108$, $p>.8$).

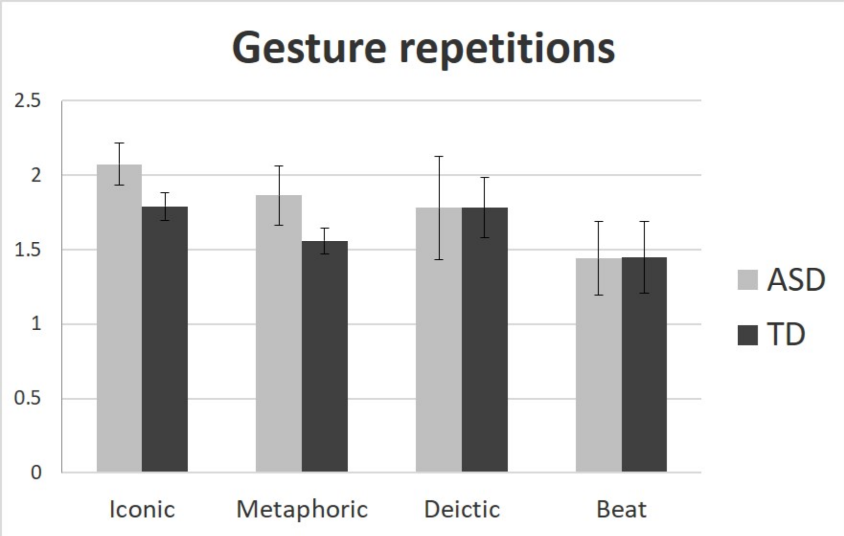

Figure 4.9 Average number of repetitions per gesture by diagnostic group (ASD, TD) and type of gesture (iconic, metaphoric, deictic, beat).
Error bars represent the standard error of the mean.

4.18.4 Gesture handedness

Gesture handedness was not the main focus of this study, and we did not have any predictions concerning the laterality of the different types of gestures or whether it should differ between groups. We did however want to check that any differences in gesture dynamics was not due to a difference in hand dominance between group. In addition, there is evidence of differences in laterality in language processes in ASD, which also translate into laterality differences in gestures (see Lindell & Hudry, 2013, for a review). The analysis revealed some interesting patterns of laterality between the different types of gestures which may be of interest for a follow-up project but will not be explored further in this thesis.

A mixed model ANOVA with group (ASD, TD) as between subjects factor and type of gesture (iconic, metaphoric, deictic, beat) and handedness (right, left, right and left) as within subjects factors, revealed a **significant main effect of handedness** [$F(2,24) = 4.67$, $p < .05$, $\eta_p^2 = .18$] and a **significant gesture type x handedness interaction** [$F(6,24) = 11.42$, $p < .001$, $\eta_p^2 = .34$]. There was no significant main effect of group [$F(1,24) = 1.57$, $p > .2$, $\eta_p^2 = .07$], no significant main effect of gesture type [$F(3,24) = 1.09$, $p > .3$, $\eta_p^2 = .05$] and the group x gesture type interaction [$F(3,24) = 1.42$, $p > .2$, $\eta_p^2 = .06$], group x handedness interaction [$F(2,24) = .52$, $p > .4$, $\eta_p^2 = .02$] and group x gesture type x handedness interaction

[$F(6,24) = .35$, $p > .9$, $\eta_p^2 = .02$] were not significant. Gesture handedness is illustrated in figure 4.10.

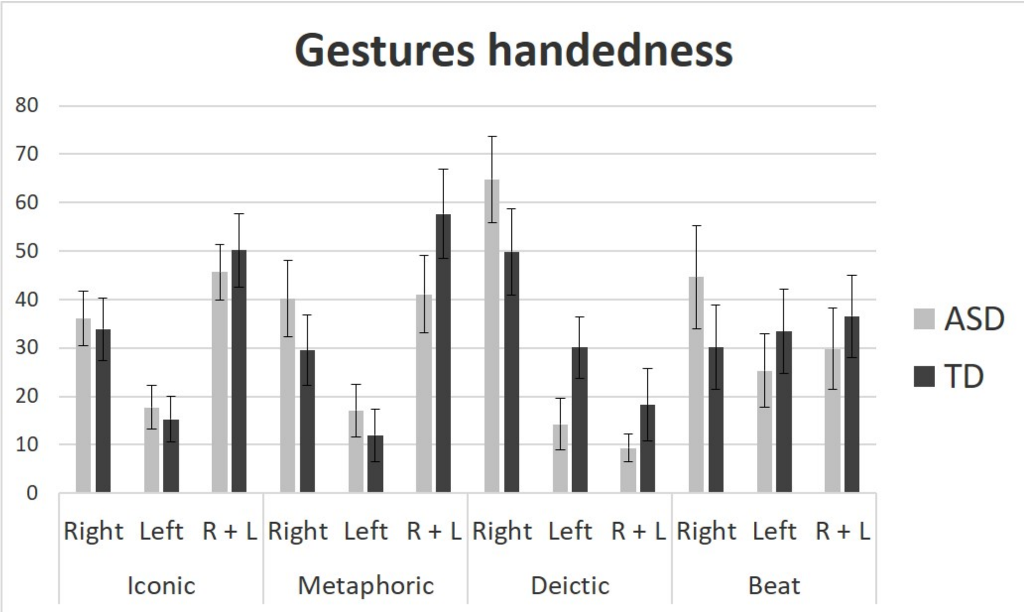

Figure 4.10 Percentage of gestures performed according to handedness.
Right: with the right side only; Left: with the left side only; R + L: with both sides, for each type of gestures (iconic, metaphoric, deictic, beat) and diagnostic group (ASD, TD). Error bars represent the standard error of the mean.

Post-hoc paired-sample t-tests were conducted to explore the gesture type x handedness interaction, for each gesture category. We found that a greater percentage of iconic gestures were performed with the right hand than the left hand ($t(31) = 3.127$, $p < .005$), and a lower percentage of iconic gesture were performed by the left hand than by both hands ($t(31) = -4.608$, $p < .001$). Similarly, a greater percentage of metaphoric gestures were performed with the right hand than the left hand ($t(28) = 2.843$, $p < .01$), and a lower percentage of metaphoric gesture were performed by the left hand than by both hands ($t(28) = -3.977$, $p < .001$). Deictic gestures were performed with the right hand more than by the left hand ($t(26) = 3.689$, $p < .005$) and by both hands ($t(26) = 4.817$, $p < .001$). There was no difference in the different percentage handedness for right, left and both hands for beat gestures (all $ps > .4$). none of the other comparisons were significantly different (all $ps > .1$).

4.19 Summary and provisional discussion: descriptive statistics

Descriptive analysis of acoustic and kinematic features of ASD and TD participants productions revealed a mixed pattern of differences and similarities between groups.

In the acoustic domain, and contrary to prediction, pitch did not differ between groups, and there was no relation between pitch and clinical features of ASD. Although the systematic review by Fusaroli and colleagues (2017) report significant differences in pitch between ASD and TD groups, they found a small to medium effect size (Cohen's d of 0.4-0.5), which is not too far from our effect size of 0.38. It may be, therefore, that our analysis is simply underpowered to detect small differences in pitch. In contrast, we did find that speech was on average louder in autistic participants, which is in line with reports that autistic individuals produce a significantly larger proportion of "loud" utterances compared to matched TD individuals (Shriberg et al., 2001). This result should be taken with some caution because the original experiment was not designed to record controlled auditory data, therefore the exact distance between the participant and the camera (although approximately the same) was not measured and controlled for. As a result the possibility that louder speech in the ASD group is an artefact resulting from the position of the camera cannot be entirely excluded, although examination of the data and the large effect size (Cohen's d = 1.11) suggest that this is a true positive result. Alternatively, it is possible that autistic participants behaved differently and oriented to the camera more or more often than TD participants, resulting in an artificially inflated measure of loudness. Close examination of the recordings did not however support this idea.

The pattern of speech pauses was surprisingly opposite to the one predicted: whilst Mitchell (2015) reported that young autistic adults produced more pauses between and within utterances, and Morett et al. (2016) and Lake et al. (2011) reported that autistic adolescents and adults produced more pauses over 2s than TD participants, autistic participants in our task produced less pauses, and those pauses were shorter by about 100ms compared to the pauses produced by TD participants. Specifically, ASD participants produced more brief pauses (under 200ms) and less long pauses (over 1s) than their TD counterparts. There were differences in the way the pauses were counted in the different tasks. In particular in previous studies pauses were coded manually whereas in our task they were computed based on the absence of pitch in the acoustic signal. However this would not be expected to reverse the pattern of results. The discrepancy between our and previous studies might be better accounted for by the nature of our task which was essentially a memory task:

participants had to retrieve information about a long, complex event retrospectively, and emphasis was placed on the accuracy of information (e.g., who performed which action), which we would expect to introduce more and longer pauses in order to reflect or to express the degree of certainty of the information. In Mitchell's task, which was a mock job interview, participants were arguably under pressure to provide an answer quickly and convincingly in order to create a more favourable impression on their listener (Brosy et al., 2016). In Lake et al.'s task, communication samples were created in the context of a naturalistic, spontaneous conversation about the participants' hobbies, with no particular memory load or pressure to answer positively (other than compliance to the experimental setting). Finally in Morett et al., the participants' task was to tell a short, fun story which would again not be expected to put extra load on memory or the pressure to make good impression. We speculate that in the TD group, the amount and duration of pauses might vary more depending on the context than in the ASD group. Specifically, we hypothesize that in a situation where self-presentation is important, pauses are fewer and shorter, whereas in a situation where content accuracy is important, there are more, longer pauses. Autistic participants might not adjust pause production to the context as much as their TD counterparts, therefore producing seemingly more and longer pauses in contexts where self-presentation is most important and fewer, shorter pauses in situations where accuracy is more important. Unfortunately previous studies do not provide the average duration of pauses in each group, but this hypothesis could be tested by contrasting the pattern and duration of pauses in different contexts with a gradient of emphasis between self-presentation and accuracy.

In the kinematic domain, the analysis yielded no difference in the average motion energy between groups. The pattern of motion holds however differed between groups, with the ASD group producing marginally more holds than the TD group. In addition, in the ASD group, a greater proportion holds (in particular holds of duration 200-500ms) was associated with greater difficulties in the communication domain as measure by the ADOS. This suggests that although the overall amount of motion energy did not differ between group, the patterns of motion / hold was more interrupted in ASD, possibly affecting the flow of communication. The next two sections (Chapter 4, part D and E) will look closer at the periodicity and recurrence of holds to provide further insight.

In addition, ASD and TD adults were found to produce comparable amounts of gestures of similar duration in each category. Notably, there was a huge variability in the amount of gestures produced within groups. This corroborates previous findings by de Marchena and Eigsti (2010) who also found no difference between ASD and TD adolescents in the amount and duration of gestures in a story telling task (after looking at cards depicting a story) but a large range in the amount of gesture produced in both groups. In contrast, Morett and colleagues (2016) found that autistic adolescents produced fewer metaphoric and beat gestures than their TD counterparts in a story-telling task (after watching a video), which indicated that various supports for the task (live event, video, pictures) might elicit different types of gestures. Overall in our data participants produced more representative gestures (iconic and metaphoric) than non representative gestures (deictic and beats), and produced more iconic than metaphoric gestures, which is probably due to the fact that they were describing a real-life event, rich in contextual details and involving a series of manual acts (first aid scenario). Iconic gestures also lasted longer than any other type of gesture.

Finally, in order to preclude that disparities in the temporal aspects of gestures were not simply due to differences in dexterity, we looked at the handedness of gestures, which we found varied according to the type of gesture: most iconic and metaphoric gestures were performed either with the right hand (approx. 35%, the dominant hand for most participants) or both hands (approx. 50%), whilst deictic gestures were performed mostly with the right hand only (over 50%) and beats were produced equally with the right-only, left-only or both hands. Although handedness analysis is not the main focus in this thesis, we can speculate that the differential use of the two hands for gesture execution likely reflects the complexity of the gestures and the precision of the motor skills required: gestures with semantic content and multiple functions such as iconic and metaphoric gestures required the use of both hands (for instance to illustrate how they rolled a bandage around the manikin's arm, participant represented the manikin's arm with one arm and illustrated the rolling motion with the other hand). Representative gestures as well as deictic gestures require spatial precision and good motor coordination to be meaningful, which could be the reason why they tended to engage participant's dominant hand. Finally, and interestingly, beat gestures showed no preferential use of any combination of hand, highlighting that they did not depend as much on spatial information and fine motor skills, and mobilised whichever side was available. Corballis (2003) argues that dominant right-handedness in humans is related to the fact that vocalisation or speech is lateralised to the left brain hemisphere in a majority of people. He hypothesises that speech evolved from manual gestures (performed predominantly with the right hand, therefore engaging the left hemisphere): gestures were accompanied at first by vocalisations which involved into speech, which is why communication, for most people, developed a left hemisphere dominance. In light of this theory, it is interesting to consider that dominantly right-handed iconic, metaphoric and deictic gestures have a semantic content, whereas beat gestures seem to stand apart and engage a separate motoric allocation mechanism. In this context, it is interesting to note that although children with ASD have been found to show atypical lateralisation of language with a right-hemisphere dominance (Flagg, Cardy, Roberts, & Roberts, 2005), in our sample ASD and TD participants did not differ overall in the laterality of co-speech gestures, suggesting typical language lateralisation.

The next section will go beyond descriptive statistics of temporal aspects of communication and look into the temporal coordination of speech and gestures.

Part D – Temporal coordination of speech and gesture

The analysis of separate aspects of speech and gestures in the recordings so far has revealed a few differences between ASD and TD groups. Autistic participants were found to speak louder than typical participants on average, and the two groups showed different patterns of pauses and holds that suggested atypical adaptation to context in ASD. In addition, some relations were uncovered between temporal aspects of speech and gesture behaviours (particularly the rate and duration of gestures) and outcome measures including clinical scores, cognitive skills and quality of communication.

However by considering only static measures, we may be overlooking a wealth of information with respect to the temporal dynamics of speech and gesture in autism. The literature on speech-accompanying gestures is unanimous about the fact that gestures and their lexical affiliates happen in close temporal proximity (Kendon, 1980; McNeill, 1992) and that this proximity is crucial to reap the benefit of complementary auditory and visual information (e.g., Leonard & Cummins, 2011; Treffner, Peter, & Kleidon, 2008). Specifically, a majority of studies reported that the onset of gestures usually precede the onset of the corresponding speech referent (e.g., Morrel-Samuels & Krauss, 1992) although one study reported that when considering the stroke phase of the gesture (the phase which contains the meaning of the gesture, as opposed to the preparation and retraction), most iconic gestures happened synchronously with their lexical affiliate (Chui, 2005). In the case of autism, two studies (de Marchena & Eigsti, 2010; Morett et al., 2016) provided some evidence that the onset of representational gestures stroke phase preceded the onset of their lexical affiliates and that this lag was on average 250ms longer in adolescents with ASD. In addition, the studies report that this gesture-speech asynchrony explains 20% of the variability in quality of communication ratings, suggesting that temporal coordination of speech and gesture heavily impacts on a direct outcome measure. In this section, we seek to reproduce these findings and extend them to non representative gestures (deictic and beat gestures). To that end, we used two different methods: the first one relies on manual gesture annotation and measures the temporal lock (or time-lock) between gesture characteristic times (hereafter "gesture timepoints") and salient moments in the acoustic signal. The second method is an attempt to reduce the need for manual coding further and investigates the cross-correlation between kinematic (motion energy) and acoustic (pitch, loudness) prosodic time-series. The temporal relationship between pauses and holds is also examined in order to better understand results from the previous section. A small literature suggests that similarly to speech and gestures, speech pauses and gestural holds are closely coordinated in time (Esposito & Esposito, 2011; Esposito & Marinaro, 2007). As a working hypothesis, we predict that the coordination of pauses and holds will be stricter in the TD compared to the ASD group, with a smaller correlation and a longer lag between the two in the ASD group.

4.20 Temporal coordination of speech and gestures: Coding data

First, we were interested in how closely gestures were time-locked to speech, in other words how strict the temporal relationship was between speech and gestures. Previous studies have typically measured the asynchrony between particular timepoints of a gesture (onset, offset or stroke) and the onset, offset or vowel onset of its lexical affiliate (e.g., Eigsti, De Marchena, Schuh, & Kelley, 2011; Morett, O'Hearn, Luna, & Ghuman, 2016). Hereafter we will refer to this asynchrony as "gesture-speech lag", or simply "lag". This approach presents several challenges: first, there is rarely a one-to-one correspondence between a gesture and a single word or even a phrase, and even then it might illustrate information which is not conveyed in the speech so it is difficult to identify the best word or unit of speech which is related to the gesture. For instance, the speech might convey the notion that an event happened "two months ago", whilst a co-occurring gesture could indicate that the temporal information is approximate. Some gestures are also produced in the absence of speech, during a pause or vocalisation (for instance when searching for the right word), whilst beat gestures by definition have no semantic content but rather an emphatic function which is dependent on when they happen in time. Finally, the choice of characteristic timepoint itself is often arbitrary. A solution to this is offered by Leonard and Cummins (2011) who compared several timepoints within beat gestures and their temporal relationship to speech. They found that the lag variability was lowest when it was computed from the stroke of the gesture (the point of maximal extension), and concluded that stroke was the point most closely time-locked to speech.

To address these challenges, we followed two steps: 1) we substituted the identification of a lexical affiliate by using the nearest salient event in speech acoustic features and 2) we computed the speech-gesture lag using three different gesture timepoints (onset, offset and stroke), and compared their variability to select the timepoint which was most closely time-locked with speech features. In order to identify the nearest salient event in speech, we used the findpeak function in Matlab to extract local maxima in pitch (f0) and loudness (loudness) series. For each gesture, and for each timepoint, we identified the nearest peak in each of the two time series, pitch and loudness (see Figure 4.11) and computed the mean (**lag**) and standard deviation (**lag variability**) of the interval between that peak and the gesture timepoint. A positive lag indicates that the peak in a given acoustic stream occurred before the gesture's timepoint, and a negative lag indicates that the peak occurred after the gesture's timepoint.

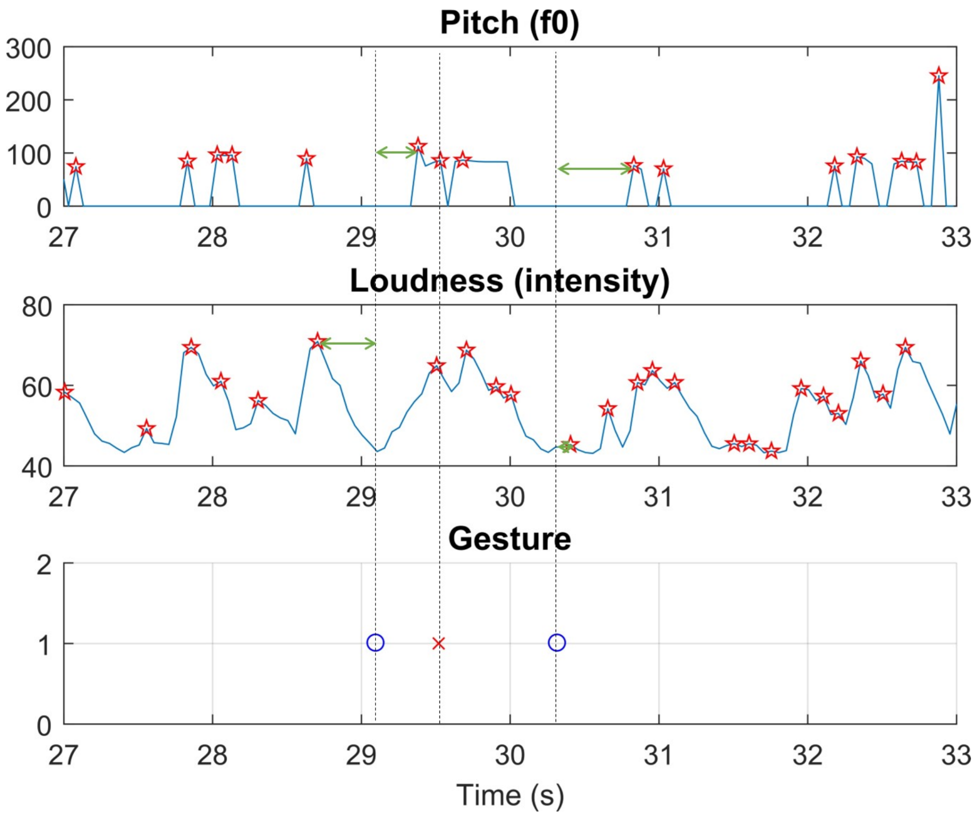

Figure 4.11 Determining the temporal lock between a gesture's timepoints and co-occurring speech events.
The top and middle panels show an interval of the speech pitch (top) and loudness (middle) timeseries respectively, with the local maxima ('peaks') marked as red stars. The bottom panel shows the onset and offset times (blue circles) and the stroke time (red cross) of one iconic gesture happening during that interval. The vertical dotted lines project when in time the characteristic times of the gesture happen in relation to events in speech. The horizontal green arrows represent the lag between the various gesture timepoints and the nearest peak in the pitch and loudness time series (there are no arrows for the stroke because it falls exactly on a peak both in the pitch and in the loudness time series).

For each category of gesture we excluded participants who produced only one or no gesture of that type (and therefore showed an artificially low variability). This led us to exclude 1 ASD participant for iconic gestures, 2 ASD and 2 TD participants for metaphoric gestures, 6 ASD and 4 TD participants for deictic gestures and 3 ASD and 2 TD participants for beat gestures. In addition we excluded 1 ASD participant whose lag values were more than 3 standard deviations above the mean. Inspection of individual data revealed that this last excluded participant had the lowest IQ scores in the sample (FIQ = 81). This was the same participant who was excluded from the speech pause analysis because their percentage of pause behaviour was over 2.5 standard deviation away from the mean of the group (and represented 95% of the recording). Atypical lag was likely due to the small amount of data in the recording, which in turn might have been linked to cognitive abilities and memory skills. To retain a maximum of statistical power, we ran the analysis separately for each gesture category.

In light of the literature, we predicted that gesture stroke would be the timepoint most closely time-locked to speech for beat gestures, and tentatively predicted that this would also extend to other types of gestures. Although previous studies consistently reported that gestures precede speech, authors usually considered the lag between the onset of gesture and the onset of the related speech. If our prediction is correct and gesture strokes are more closely time-locked to speech than both onset and offset of gestures, then we expect the speech-gesture lag to be close to zero (in line with Leonard & Cummins, 2011). Finally, based on de Marchena & Eigsti (2010), we expect the speech-gesture lag to correlate with quality of communication ratings.

4.20.1 Identifying the best characteristic timepoint for gesture

First, we wanted to identify which gesture timepoint(s) (onset, offset and stroke of the gesture) were the most relevant when considering speech and gesture coordination. To that end, following a similar reasoning to Leonard and Cummins (2011), we compared the individual standard deviations of the lag separately for the different timepoints and gesture type. We reasoned that if a particular timepoint during the gesture was especially important relative to the accompanying speech, its distance to the nearest peak in acoustic time series should show the lowest variability.

For each type of gesture, a mixed model ANOVA with group (ASD, TD) as between subjects factor and acoustic feature (pitch, loudness) and gesture timepoint (onset, offset, stroke) as within subjects factor was conducted on lag variability for each gesture category separately. In each gesture category, the ANOVA yielded a large, significant main effect of acoustic feature (Iconic: $F(1,32) = 88.74$, $p < .001$, $\eta_p^2 = .75$; Metaphoric: $F(1,29) = 46.42$, $p < .001$, $\eta_p^2 = .63$; Deictic: $F(1,27) = 24.66$, $p < .001$, $\eta_p^2 = .50$; Beat: $F(1,29) = 4.39$, $p < .05$, $\eta_p^2 = .14$), with a significantly greater variability for pitch than for loudness, which led us to analyse pitch and loudness data separately. Results are illustrated in figure 4.11.

4.20.1.1 Lag variability between gesture and pitch

Iconic gestures: a mixed model ANOVA with group (ASD, TD) as between subjects factor and gesture timepoint (onset, offset, stroke) revealed a significant main effect of gesture timepoint ($F(2,32) = 3.58$, $p < .05$, $\eta_p^2 = .11$) and a significant main effect of group ($F(1,32) = 4.83$, $p < .05$, $\eta_p^2 = .14$) on lag variability. Post-hoc t-test showed that the lag variability was smaller when computed from the gesture stroke than either gesture onset (marginal effect, $p = .07$) or gesture offset ($p < .05$). In addition, overall lag variability was greater in the TD compared to the ASD group.

Metaphoric gestures: a mixed model ANOVA with group (ASD, TD) as between subjects factor and gesture timepoint (onset, offset, stroke) revealed a significant main effect of gesture timepoint ($F(2,29) = 6.31$, $p < .005$, $\eta_p^2 = .19$), a significant main effect of group ($F(1,29) = 6.87$, $p < .05$, $\eta_p^2 = .20$), and a significant timepoint x group interaction ($F(2,29) = 4.83$, $p < .05$, $\eta_p^2 = .15$) on lag variability. Post-hoc paired-sample t-test showed that lag variability was smaller when computed from the gesture stroke than either gesture onset ($p < .005$) or gesture offset ($p < .001$). In addition, overall lag variability was greater in the TD compared to the ASD group. The timepoint x group interaction was driven by the fact that the lag variability computed from the gesture offset was significantly higher for the ASD compared to the TD group ($p < .005$), whereas lag variability computed from gesture onset and stroke did not differ between groups ($p = .89$ and $p = .65$ respectively). In addition, in the TD group, the lag variability computed from the gesture offset was significantly greater than from the gesture stroke ($p < .001$), whereas the two did not significantly differ in the ASD group ($p > .8$).

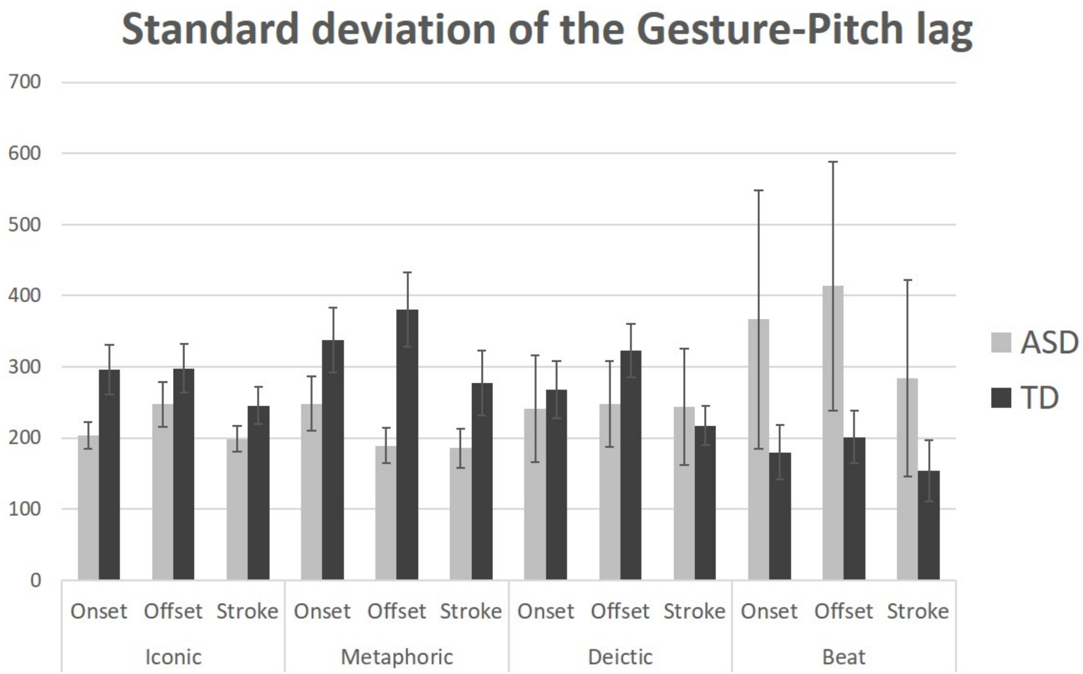

Figure 4.12 Average standard deviation of the lag between gesture and pitch.
The lag is computed between gesture timepoints (onset, offset and stroke) and the nearest peak in the pitch time series, per group and gesture type. Error bars represent the standard error of the mean.

Deictic gestures: a mixed model ANOVA with group (ASD, TD) as between subjects factor and gesture characteristic timepoint (onset, offset, stroke) revealed no significant main effect or interaction (all $ps > .2$) on lag variability.

Beat gestures: a mixed model ANOVA with group (ASD, TD) as between subjects factor and gesture timepoint (onset, offset, stroke) revealed a significant main effect of gesture timepoint ($F(2,29) = 5.81$, $p < .01$, $\eta_p^2 = .18$) on lag variability. Post-hoc paired sample t-test showed that the lag variability was smaller when computed from the gesture stroke than either gesture onset ($p < .05$) or gesture offset ($p < .01$).

Overall, the analysis just presented suggested that the standard deviation of the gesture-pitch lag was smallest when the lag was computed from the stroke of the gesture, supporting the idea that gesture stroke is the timepoint which is most strictly time-locked to pitch. Although the main effect of gesture timepoint was not significant for deictic gestures, numerically the lag variability was also smallest for the lag computed from gesture stroke. Interestingly, the speech-pitch lag variability for iconic and metaphoric gestures was greater in the TD than in the ASD group, suggesting that iconic and metaphoric gestures were more strictly time-locked to pitch peaks in the ASD group. By contrast, lag variability for beat gestures appeared greater in ASD than TD participants although this difference was not statistically significant due to large individual differences in the ASD group.

Interestingly, we found that the ADOS communication score correlated negatively with the lag variability between deictic gesture stroke and pitch ($r(14) = -.621$, $p < .05$) and marginally with the lag variability between iconic gesture and pitch ($r(14) = -.470$, $p = .066$), indicating that tighter temporal coordination between deictic gesture and speech pitch was associated with greater difficulties in the communication domain. Moreover, lag variability between deictic gesture stroke and pitch correlated negatively with visual gain ($r(27) = -.424$, $p < .05$).

4.20.1.2 Lag variability between gesture and loudness

Mixed model ANOVAs with group (ASD, TD) as between subjects factor and gesture timepoint (onset, offset, stroke) revealed no significant main effects or interactions on lag variability for any of the gesture categories (all ps > .1). This result does not allow us to make inferences on the gesture timepoint which is most closely time-locked to speech loudness.

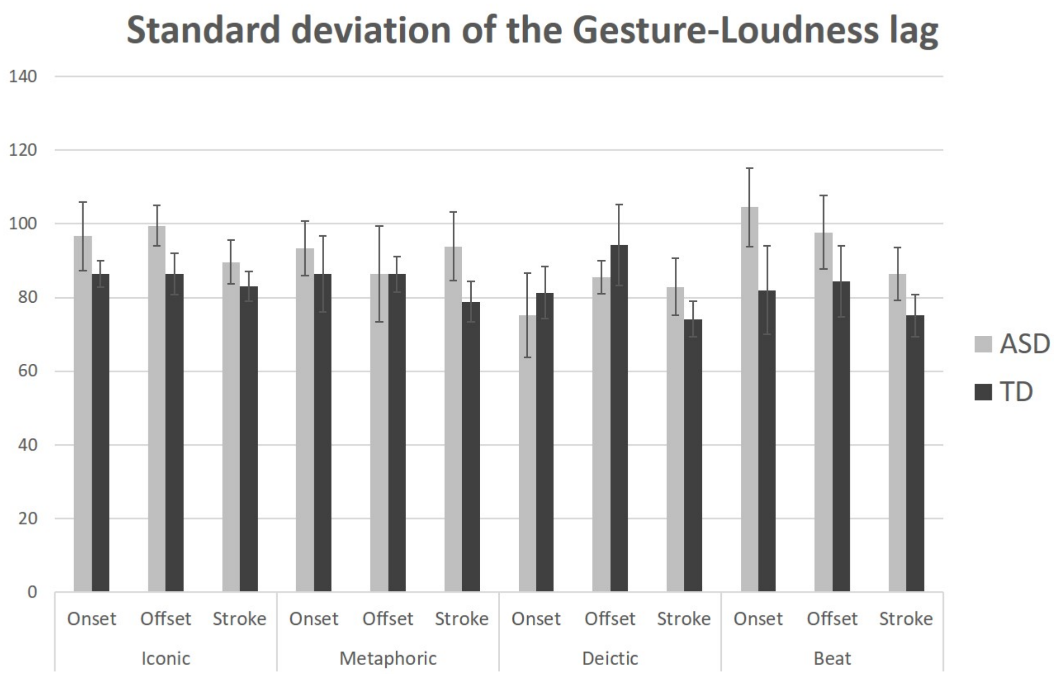

Figure 4.13 Average standard deviation of the lag between gesture and loudness.
The lag is computed between gesture timepoints (onset, offset and stroke) and the nearest peak in the loudness time series, per group and gesture type. Error bars represent the standard error of the mean.

However, because numerically the gesture-loudness lag variability was lowest when computed from gesture stroke (see figure 4.12), and in the interest of simplicity, **we will select gesture stroke as the marker of gesture time for both pitch and loudness features** in the following section.

Interestingly, we found that the lag variability between DG stroke and loudness highly correlated with the ADOS communication score ($r(14) = -.766$, $p < .005$) and with the ADOS restricted and repetitive behaviour score ($r(14) = -.655$, $p < .05$), indicating again that tighter temporal coordination between deictic gestures and speech loudness was associated with greater difficulties in the communication domain.

4.20.2 Temporal lock between gestures and speech features

Next, we examined the average gesture-speech lag, by computing the average duration of the interval between the stroke of a gesture and the time of the nearest local maximum in the pitch and loudness time series. In order to preserve a maximum number of participants in the analyses, the data was again analysed separately for the different gesture categories (as not all participants produced sufficient numbers of all types of gestures).

Planned independent samples t-tests were conducted on average lag values, and results are reported in Table 4.7. Results indicated that the average lag between metaphoric gestures and their nearest pitch peak was significantly different in the ASD compared to the TD group, with a large size effect (Cohen's $d = .8$), indicating that in the ASD group metaphoric gesture stroke tended to precede a local maximum in pitch, whereas in the TD group metaphoric gesture stroke tended to follow a local maximum in pitch. The absolute lag was also smaller in the ASD group, suggesting that on average metaphoric gestures were performed closer to the nearest pitch peak by ASD compared to TD participants. There was no other significant differences between groups (all $ps > .08$).

		ASD	TD	t-test		Cohen's d
				t	p	
Pitch	**Iconic**	n_{ASD} =16	n_{TD} =16			
	mean (ms) *(SD)*	**2** *(66)*	**-1** *(43)*	.141	.889	.054
	Range (min:max)	-150:190	-92:67			
	Metaphoric	n_{ASD} =15	n_{TD} =14			
	mean (ms) *(SD)*	**18** *(50)*	**-45** *(99)*	**2.147**	**.041***	**.803**
	Range (min:max)	-33:122	-227:133			
	Deictic	n_{ASD} =14	n_{TD} =13			
	mean (ms) *(SD)*	**-6** *(65)*	**6** *(75)*	-.408	.688	.171
	Range (min:max)	-97:128	-118:187			
	Beat	n_{ASD} =15	n_{TD} =14			
	mean (ms) *(SD)*	**10** *(81)*	**7** *(69)*	.099	.992	.040
	Range (min:max)	-118:174	-88:176			
Loudness	**Iconic**	n_{ASD} =16	n_{TD} =16			
	mean (ms) *(SD)*	**-5** *(17)*	**-3** *(16)*	-.281	.781	.121
	Range (min:max)	-60:10	-30:32			
	Metaphoric	n_{ASD} =15	n_{TD} =14			
	mean (ms) *(SD)*	**-18** *(32)*	**-5** *(35)*	-1.014	.320	.388
	Range (min:max)	-113:34	-54:92			
	Deictic	n_{ASD} =14	n_{TD} =13			
	mean (ms) *(SD)*	**14** *(53)*	**4** *(23)*	.634	.533	.245
	Range (min:max)	-44:125	-31:63			
	Beat	n_{ASD} =15	n_{TD} =14			
	mean (ms) *(SD)*	**11** *(43)*	**-4** *(30)*	.997	.328	.405
	Range (min:max)	-77:63	-67:32			

Table 4.7 Average lag between gesture stroke time and its nearest peak in the pitch (top section) and loudness (bottom section) time series in seconds, by group and type of gesture. SD = Standard Deviation of the Mean. Range = min:max. A positive lag indicates that the stroke of the gesture preceded the nearest peak in the acoustic feature time series, whereas a negative lag indicates that the stroke of the gesture follows the nearest peak in the acoustic feature time series.

Looking at correlations between the gesture-pitch lag and outcome variables, we found that the lag between metaphoric gestures and pitch was marginally, negatively correlated with the ADOS communication score in the ASD group ($r(14) = -.480$, $p = .083$) and marginally, positively correlated with the AQ score across groups ($r(28) = .370$, $p = .052$). Finally, we found that the lag between iconic gestures and pitch correlated positively with the visual gain in quality of communication ($r(31) = .471$, $p < .01$, see figure 4.14).

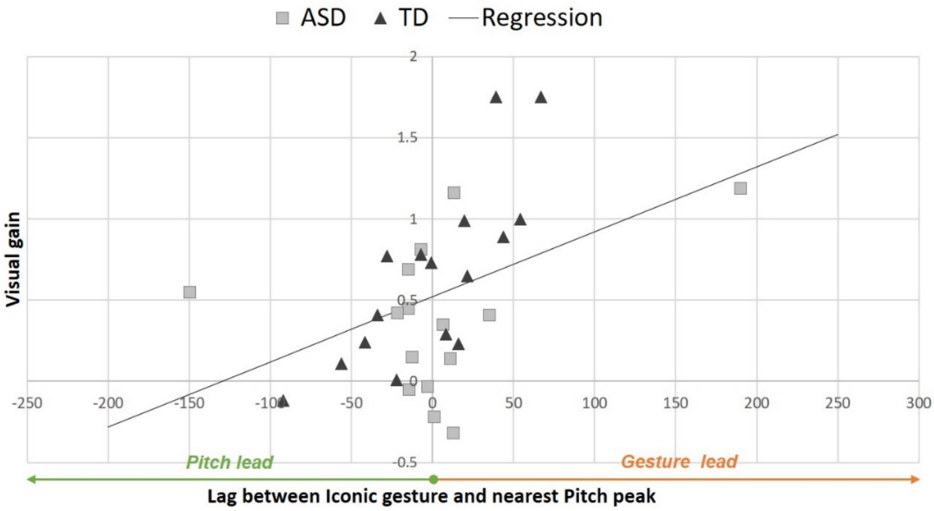

Figure 4.14 Relationship between iconic gesture-pitch lag and visual gain.
On the x-axis, a negative lag indicates that on average gestures followed the nearest pitch peak (pitch lead, green arrow) whilst a positive lag indicates that on average gestures preceded the nearest pitch peak (gesture lead, orange arrow). The correlation was computed across groups, but the graph shows ASD datapoints (light grey squares) and TD datapoints (dark grey triangles) for illustration purposes.

Turning to the correlations between the gesture-loudness lag and outcome measures, we found that in the ASD group, the lag was marginally, positively correlated with the ADOS reciprocal social interaction score ($r(14)=.491$, $p = .075$, see figure 4.15). The lag between beat gestures and loudness also showed a positive correlation with VIQ (marginal: $r(27) = .377$, $p = .052$), PIQ (marginal: $r(27) = .371$, $p = .057$) and FIQ ($r(27) = .425$, $p = .027$). There was no correlation between speech-gesture lag and quality of communication scores for loudness.

Figure 4.15 Relationship between metaphoric gesture-loudness lag and ADOS RSI score.
On the x-axis, a negative lag indicates that on average gestures followed the nearest loudness peak (loudness lead, green arrow) whilst a positive lag indicates that on average gestures preceded the nearest loudness peak (gesture lead, orange arrow). The correlation was computed for the ASD group only (light grey squares).

Whilst we postulated differences between groups at the prosodic level, the notion that prosody is constrained by syntax is widely accepted (e.g, Nespor & Vogel, 1986; Selkirk, 1986). Specifically, in English, the focus on different syntactic constituents can be realised by varying pitch peaks (e.g., Grabe, 1998). In order to check whether group differences in temporal lock between pitch and gestures were due to differences in the syntactic structure of participants' productions, we submitted the transcripts (generated by K. Maras in her original study) to the Stanford's CoreNLP parser (https://stanfordnlp.github.io/CoreNLP/). There was no significant difference in the number of lemmas between groups (M_{ASD} = 602 ±319, M_{TD} = 608±284, $t(31)=-0.052$, $p>0.9$). Close inspection of Part-Of-Speech (POS) tags as well as constituency parses did not reveal any noticeable differences between groups. Representative sample sentences from each group and their constituency parses are presented in Appendix 2.1. The proportions of POS tags for each group are illustrated in Appendix 2.2, and suggest high similarity between the productions of autistic and non autistic participants. As a very crude measure of syntactic structure, the proportion of coordinating conjunctions (CC) and the number of prepositions or subordinating conjunctions (IN) were analysed separately as an index of complex sentence constructions. Independent-sample Mann-Whitney tests revealed no significant difference in the proportion of CC ($U=125$, $p=.709$) or IN ($U=115$, $p=.465$) between groups. Overall, differences in syntactic structure are unlikely to account for differences in prosody between the ASD and TD groups in this dataset.

4.20.3 Summary and provisional discussion

Examination of the variability in speech-gesture lag revealed that the temporal lock between gesture and pitch events was greater when the lag was computed from the stroke timepoint in the gesture, for iconic, metaphoric and beat gestures. Although the differences did not reach significance, the variability was also lowest numerically for deictic gesture-pitch lag and all gesture-loudness lags when computing lag from the gesture stroke. As predicted, this result confirms and extends Leonard and Cummins (2011) finding in beat gestures to representative gestures (iconic and metaphoric), with a similar trend for deictic gestures. Conceptually, it is quite intuitive that the moment of the gesture with most emphasis (often the maximum extension in space or the point of fastest acceleration) or "culminating point" of a gesture should be most closely related to speech. However, lags based on gesture onset and offset also showed low variability which indicated that gesture onset and offset also closely related to speech temporally, in line with other studies (de Marchena & Eigsti, 2010; Morett et al., 2016). It is perhaps useful to reiterate that in our design, one gesture could be annotated with several strokes, for instance if the gesture was repeated (e.g., rolling a bandage) or if it had multiple phases (e.g., pointing "from there to there"). Therefore we could capture multiple time-locks between one gesture and one speech unit.

Overall, the average lag between speech feature peaks (both pitch and loudness) and gesture strokes in our dataset were very close to zero (between approximately -50 and 20ms) which was comparable to the values reported by Leonard and Cummins (between approximately -100 and 0ms). This result challenges the literature which widely reports that gestures precede speech by several hundreds of milliseconds. However in previous studies, authors often compute the gesture-speech lag from the onset of gesture (Morrel-Samuels & Krauss, 1992) or the onset of the stroke phase (de Marchena & Eigsti, 2010; Morett et al., 2016) to determine the gesture's timing. In our study, we used manually-coded single timepoints to capture the stroke(s) of a gesture, defined as the moment(s) which carried the most meaning for that gesture, usually the maximum extension (particularly for deictic and beat gestures) or maximum acceleration. Similarly, previous studies used various temporal timepoints to define the timing of speech, including voice or word onset (de Marchena & Eigsti, 2010; Morett et al., 2016; Morrel-Samuels & Krauss, 1992), P-centre (defined as timepoint halfway through a local rise in the loudness envelope of the speech waveform), vowel onset or pitch peak (Leonard & Cummins, 2011). Because we were looking for a measure that would allow automatic processing, we chose the nearest pitch or loudness peak to the gesture stroke as a marker of speech salience. Taken together, and in line with Chui (2005), results from previous studies and from our current study suggest that the onset of a gesture (whether the onset of the entire gesture or the onset of the stroke phase) precedes the onset of the related speech; however the stroke 'apex' is near-synchronous with the pitch peak in the related speech. Despite this, we found that the iconic gesture-pitch lag correlated positively with visual gain. In other words, when the stroke of iconic gestures preceded the nearest peak in pitch, visual information tended to improve quality of communication ratings more significantly, replicating findings by de Marchena and Eigsti (2010). This brings us back to the idea that in order to benefit communication, gestures should occur shortly ahead of their related speech.

Interestingly, the metaphoric gesture-pitch lag was closer to zero in the ASD compared to the TD group, indicating that the timing of metaphoric gestures was more tightly time-locked to the pitch time-series in autism: metaphoric gestures occur closer to the pitch peak and less variably so. This was reflected in the fact that greater / more positive lags between metaphoric gestures and pitch (more common in the ASD group) were associated with higher AQ scores. In addition, we found that the lag between metaphoric gestures and pitch was negatively associated with the ADOS communication score in the ASD group. More precisely, lags which were more positive were associated with lower ADOS communication scores, suggesting once again that the tendency to produce gestures ahead of speech was beneficial to individual communication skills, extending de Marchena and Eigsti's (2010) finding to metaphoric gestures. This result (based on a single correlation) should however be replicated before a firmer conclusion can be reached. The lag between gesture and loudness did not differ between groups in our sample, however we found that lags indicating that beat gestures preceded the nearest peak in loudness were associated with higher IQ compared to lags indicating that loudness preceded beat gestures. This suggests that the timing of beat gestures in particular relies on cognitive skills. Beats are short, baton-like gestures for which meaning is purely based on timing: they have no semantic content of their own and cannot be related to speech by any other mean than the time at which they occur. Therefore the control of beat gestures timing may rely particularly strongly on cognitive skills, compared to other gesture categories which contain other information than timing .

4.21 Temporal coordination of speech and gestures: cross-correlation between time series

"Traditional" frame-by-frame coding analysis is extremely rich in terms of the level of detail that can be examined in social interaction. However this type of analysis presents some limitations: it is extremely costly in time and resources. Depending on the amount of features to code, a few-minute interaction can take several hours to code accurately, which is hardly practicable in applied clinical contexts (e.g., to facilitate diagnosis). Whilst tools such as the ADOS require examiners to record the occurrence of iconic gestures and their coordination with speech during particular presses (e.g., Demonstration task) and throughout the assessment, the coding of these gestures as ‘typical’ or ‘atypical’ does not factor in the rich complexity of gestures, but rather relies on a relatively crude impression of deviance from the norm. Moreover, video annotation and live assessments (such as the ADOS) alike are prone to individual bias, for instance categorising a gesture is often difficult as in essence gestures often capture various aspects of the verbal information (object, movement, confidence, emphasis) and the ultimate decision is based on the annotator’s understanding of the utterance and the inferences they make. Determining the characteristic timepoints also requires subjective judgement, for instance in situations where consecutive gestures don’t have clear boundaries or when gestures have an ambiguous meaning or function. For this reason, a minimum of two coders is required for at least part of the data coding so that agreement scores can be computed and data considered valid. Again, this is not a set-up that can be expected to be enforced in either a clinical or an educational setting to contribute to a diagnosis or follow-up on an individual’s progress over time.

In the following sections, we explore objective, automatic methods of speech and co-speech gesture analysis which assess various aspects of the temporal dynamics of communication in ASD. With a view that such information could in the long-term be used in the framework of diagnosis or to monitor an individual’s communication skills, we used materials that (1) could be collected easily during an interaction with an autistic individual (in a clinical setting or otherwise); (2) do not require much pre-processing; (3) do not depend on expensive or rare equipment and software.

4.21.1 Motion energy as a proxy for gestures

Our aims for the following section is to reduce subjective input to generate the data on the one hand, and optimise the analysis and reduce its cost in time and resources on the other hand. To that end, we selected motion energy (ME) as a proxy for gesture analysis. As a reminder, ME is the positive, continuous variable extracted automatically from a video recording (see Chapter 4, Part A-9 for details), and consists of the amount of change in pixels between one frame and the next. Motion energy has been used in past research to investigate non-verbal behaviour in both intra- and interpersonal interaction (Grammer et al., 1999; Kupper, Ramseyer, Hoffmann, Kalbermatten, & Tschacher, 2010; Ramseyer & Tschacher, 2011). Because of its continuous nature, ME lends itself to exploring on-going temporal relationships during a recording rather than restricting the analysis to the relation between discrete events (a gesture and speech 'unit'), identified with a high risk of subjective bias.

Figure 4.15 shows two examples of alignment between gesture timepoints as we defined them in the previous two sections, and motion energy. We can see that gesture timepoints (onset, offset, stroke) are associated with a significant increase in motion energy, but it is also clear that there are considerable differences between the ME profiles of different gestures that are lost by the cruder identification of discrete gesture timepoints. Specifically, the example on the left in Figure 4.15 is characterised by much more discrete ME peaks during gesture onset and offset than the example on the right.

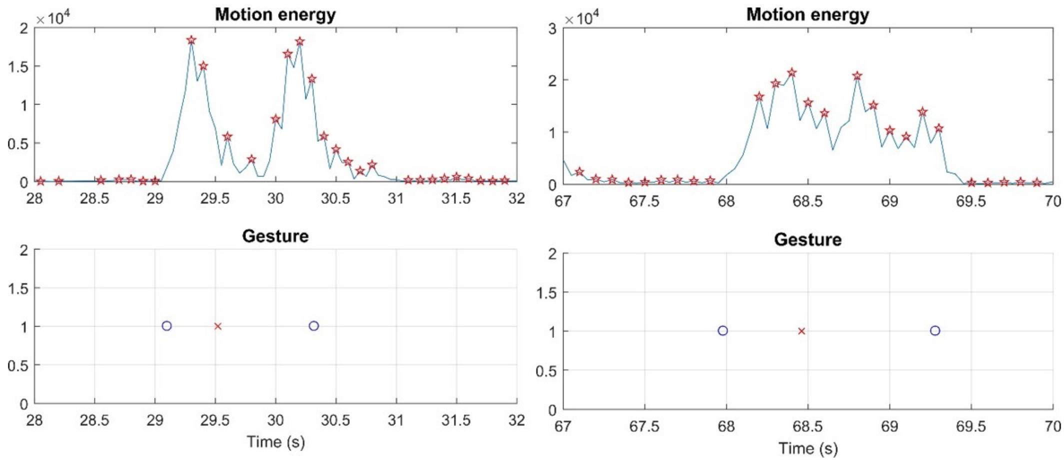

Figure 4.16 Two example of temporal alignment between motion energy (top) and gesture characteristic time points (bottom), taken here from a TD participant.

Importantly, just like pitch and loudness time series for the voice, ME provides a way to examine gestures as a dynamic system spanning in time, rather than reducing the signal to a few discrete datapoints. It therefore provides a more sensitive measure to detect on-going atypical timing of gestures in relation to speech in ASD.

4.21.2 Cross-correlation: automatised investigation of the temporal relationship between speech and gesture

The following analysis addresses both the issue of efficient, automatized processing and the issue of including gestures as a whole in the characterisation of temporal characteristics of communication in autism using **cross-correlation**. Cross-correlation is a measure of similarity between two time-series at various time lags. A cross-correlation profile illustrates the degree of similarity between the two time-series at each lag. By identifying peaks in the profile, it is possible to find out at which time lags two signals are most similar, and whether there is a leading-lagging relationship between them. Measures of cross-correlation between two signals consist of a **coefficient of correlation** at each defined time lag in a set temporal window. This method provides a fast, efficient and objective way of identifying not only if speech and movement features are correlated, but also what the characteristic temporal lags are, if any, between the two signals. For example, if two time series are synchronised with a 0s lag, their cross-correlation profile will peak at 0s. However if one time-series leads the other by 0.5s, their cross-correlation profile will peak at 0.5s (or -0.5s).

In the following sections, we investigate the cross-correlation between ME, pitch and loudness time series on the one hand, and pause and hold behaviour on the other hand, to compare the degree of coupling and the characteristic time delay between time-series.

First, because the time series were sampled using different softwares (pitch and loudness in Praat and ME using a custom script in Matlab), pitch, loudness and motion energy vectors were aligned into one matrix based on the timestamps of each time-series. Specifically, for each timestamp in the ME series, we sought the closest timestamp in the pitch and loudness series to match acoustic information with kinematic information. Empty cells (e.g., at the start of the recording, pitch / loudness are not registered if null) were filled with zeros. Pauses and holds did not require temporal alignment or further processing. Since the series were sampled at 50ms and each sampled value was matched to the closest value in time in the other time-series, cross-correlation lag results have a precision of ±25ms. Cross-correlation were run in Matlab using the xcorr function. Based on Morrel-Samuels and Krauss (1992) who provided evidence that most gestures happen within 4s of their lexical affiliate, and after exploring different windows of cross-correlation, we allowed a maximum lag of ±5s between signals to compute cross-correlation. Figure 4.16.A shows an example of cross-correlation profile. From these cross-correlation profiles the following information was extracted from the main peak (i.e., the local maxima of the cross-correlation profile that indicate at which time-delay the different signals are most closely time-locked, see figure 4.16.B): the **correlation value at peak** (i.e., the strength of the association between signals), the **peak lag** (i.e., the time-delay at which the correlation between signals is strongest), **peak prominence** (i.e., the height of the peak, indicating how the correlation at peak lag differs from baseline correlation throughout the recording) and **peak width** (i.e., the width of the peak measured at half the prominence, indicating the interval within which events in acoustic and kinematic series are likely to happen in proximity). In order to select meaningful peaks (as opposed to small local maxima in a noisy time series), only peaks with a prominence of more than 2 standard deviations (computed over the entire time-series) were retained for the analysis. This allowed us to detect not only the highest peak, but also any other peaks in the data which would indicate a closer time-lock between the two time series. In the case of multiple peaks, we selected the peak with the highest correlation value at peak for the analysis.

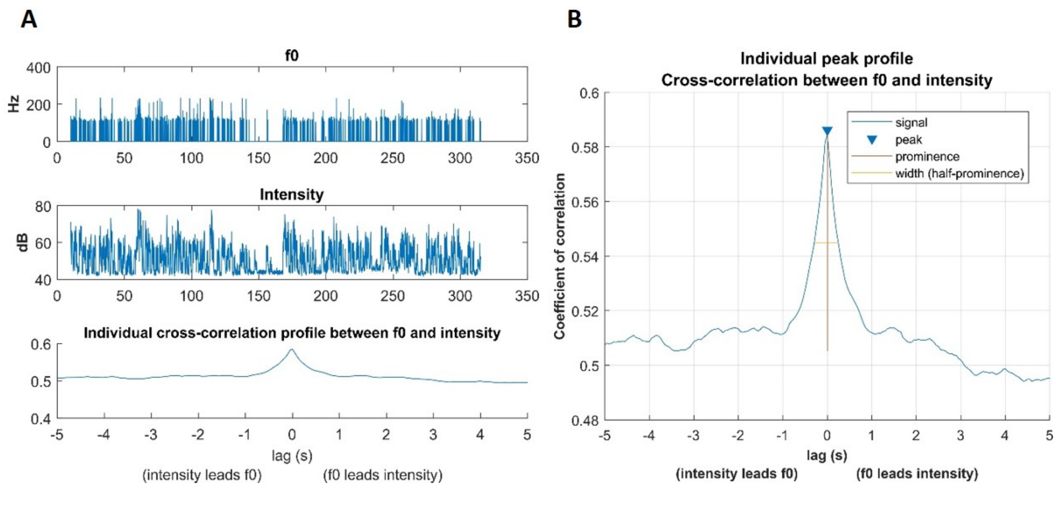

Figure 4.17 Example of an individual profile of cross-correlation between pitch (f0) and loudness of speech, taken here from a TD participant.
A. Relation between the f0 time-series (top section), the loudness time-series (middle section), and the cross-correlation profile (bottom section) between f0 and loudness. B. Peak characteristics: peak lag (time of the peak), prominence (height of the peak) and width (measured at half the prominence).

Finally, in order to detect differences in lead/lag relationship between time-series, we computed the **area under the curve** on the left and the right of lag 0ms respectively. The window of integration to compute this area was defined based on the average peak width for each condition (see below). We then subtracted the area on the left from the area on the right to obtain a measure of **skewness**. A positive difference meant that the cross-correlation curve was skewed to the right, which meant that overall time-series 1 was leading time-series 2. Conversely, a negative difference meant that time-series 1 was generally lagging behind time-series 2.

As described in section 4.20.2, no differences in the syntactic structure of the productions were observed between ASD and TD participants, suggesting that any group differences detected in the temporal dynamics of prosodic cues cannot be accounted for by differences in syntax between the groups.

4.21.2.1 Cross-correlation between pitch and loudness

We first explored the temporal relationship between pitch and loudness. Because they are two features of the same complex signal (speech) and produced through the same physical action, we expected pitch and loudness to be highly correlated and show a time lag close to 0ms. This prediction was confirmed in both groups by a robust single-peak profile centred on a lag close to 0s in all but one participant (one TD participant presented 2 peaks, one centred on 50ms and another one centred on 1600ms). Two ASD and one TD participants were excluded from the analysis because their peak characteristic values were outside of ±2.5 standard deviation of the mean of their group. Inspection of the participants revealed one of the two excluded ASD participants was the participant with the lowest IQ who was also excluded from descriptive data and temporal lock analyses. There were no obvious differences between the other two excluded participants and the rest of the sample. Because the main peak in the cross-correlation profiles had an average width (measured at half-prominence) of around 500ms, we chose a window of integration of [-500 to -50ms] on the left and [50 to 500ms] on the right of lag zero to measure skewness. Figure 4.17 shows the averaged profile of cross-correlation between groups.

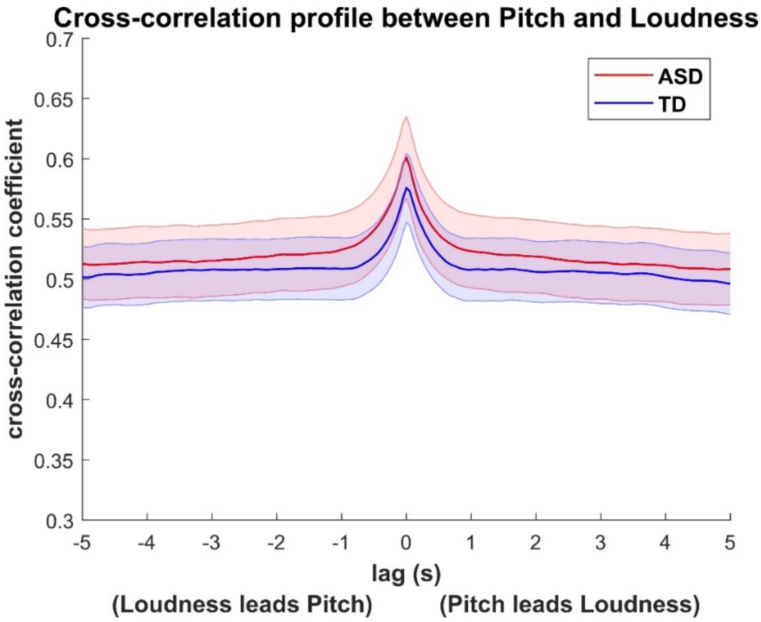

Figure 4.18 Average profile of cross-correlation between pitch and loudness between the ASD group (in red) and the TD group (in blue).
Shadowed areas represent the standard error of the mean.

Planned independent samples t-tests were conducted to compare peak characteristics and skewness of the curve between groups on the remaining 15 ASD and 15 TD participants. Results are reported in table 4.8.

	ASD (n=15)	TD (n=15)	t-test		Cohen's d
			t	p	
Number of peaks					
mean *(SD)*	**1.00** *(0)*	**1.06** *(0.25)*			
Range (min:max)	1:1	1:2			
Peak correlation coef					
mean *(SD)*	**.616** *(.081)*	**.565** *(.106)*	1.473	.152	.541
Range (min:max)	0.446:0.748	0.377:0.781			
Peak lag (ms)					
mean *(SD)*	**10** *(2)*	**20** *(25)*	-1.183	.247	.564
Range (min:max)	0:50	0:50			
Peak width (ms)					
mean *(SD)*	**574** *(212)*	**515** *(119)*	.938	.356	.343
Range (min:max)	341:1116	313:729			
Peak prominence					
mean *(SD)*	**.095** *(.029)*	**.075** *(.026)*	2.016	.054	.726
Range (min:max)	0.065:0.159	0.037:0.135			
Skew (Area R - Area L)	($x10^{-3}$)	($x10^{-3}$)			
mean *(SD)*	**-.928** *(2.966)*	**2.222** *(2.933)*	**-2.925**	**.007****	**1.068**
Range (min:max)	-6.083:4.501	-2.359:6.683			

Table 4.8 Statistics of independent samples t-tests conducted to compare the characteristic of the main peak in the cross-correlation profile between pitch and loudness.

Results showed a significant difference in skewness with a large effect size whereby ASD participants' profiles showed a leftward skew indicating that loudness generally led pitch, whereas TD participants' profiles showed a rightward skew indicating that pitch generally led loudness. In terms of prosodic profile, this indicates that for autistic participants, a change in pitch was typically preceded by a change in loudness, whereas in TD a change in pitch was followed by a change in loudness. The ASD group also presented a marginally greater peak prominence than the TD group (with a medium to large effect size), showing a greater degree of correlation between pitch and loudness.

Moreover, in the ASD group, results showed that peak prominence correlated marginally with ADOS communication score ($r(15) = .451$, $p = .092$) and total ADOS score (marginal: $r(17) = .471$, $p = .056$, see figure 4.18), with greater peak prominence associated with higher ADOS score (or greater difficulties) in the communication domain.

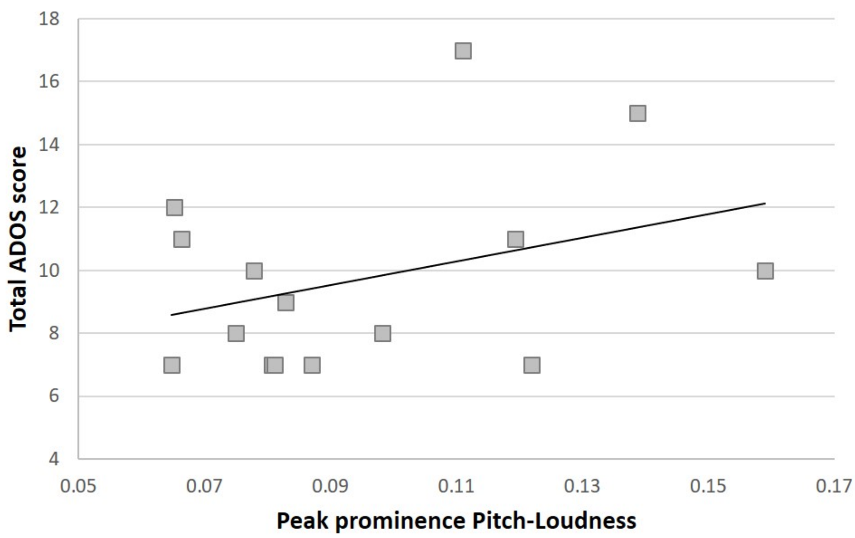

Figure 4.19 Relationship between peak prominence of the cross-correlation profile between pitch and loudness and total ADOS score.

In addition, peak width and prominence correlated negatively with VIQ (width: r(30) = -.464, p = .010, see figure 4.19, prominence: r(30) = -.462, p = .010), PIQ (width: r(30) = -.382, p = .037, prominence: r(30) = -.382, p = .037) and FIQ (width: r(30) = -.430, p = .018, prominence: r(30) = -.423, p = .020), showing that larger peaks (in width and height) were associated with lower IQ scores overall. There were no significant correlations between peak characteristics and quality of communication ratings.

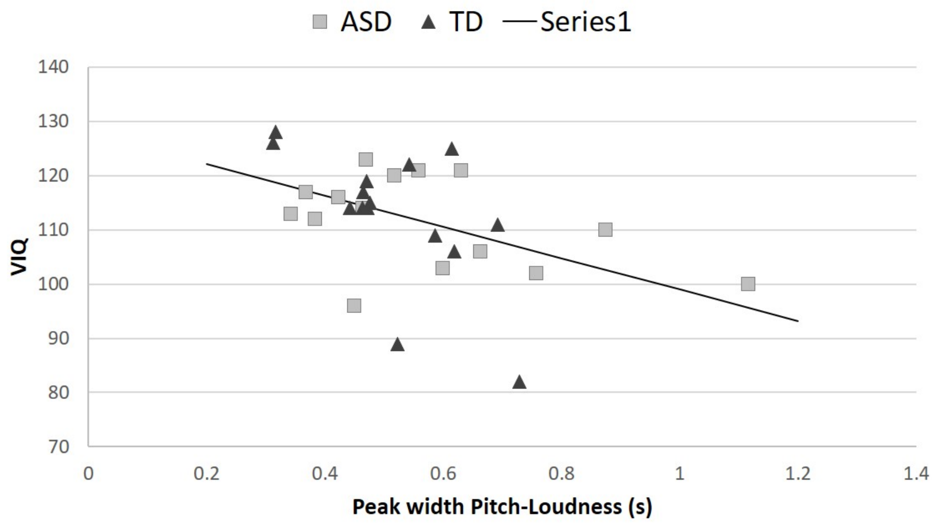

Figure 4.20 Relationship between peak width of the cross-correlation profile between pitch and loudness (in seconds) and VIQ.
The correlation was computed across groups, datapoints for ASD (light grey squares) and TD (dark grey triangles) participants are differentiated for illustration purposes.

4.21.2.2 Cross-correlation between pitch and motion energy

Peak analysis on the cross-correlation profiles of ME and pitch revealed some variability in the number of peaks identified. No peaks were detected in the profile of 3 ASD participants, suggesting that their movement and variation in pitch were not strongly coordinated in time. Examination of the three participants recordings revealed that one participant was spinning on his chair during the recording, one participant (unlike other participants) had his hands partly hidden by a table during the recording and that the third one, although producing gestures, was holding a pen in one of his hands, which may have added noise to their cross-correlation profile. 6 ASD and 9 TD participants presented a 1-peak profile; 4 ASD and 2 TD participants presented a 2-peak profile; 1 TD participant presented a 3-peak profile; and 1 ASD and 1 TD participant presented a 4-peak profile, suggesting that the temporal coordination between movement and pitch in speech is quite complex.

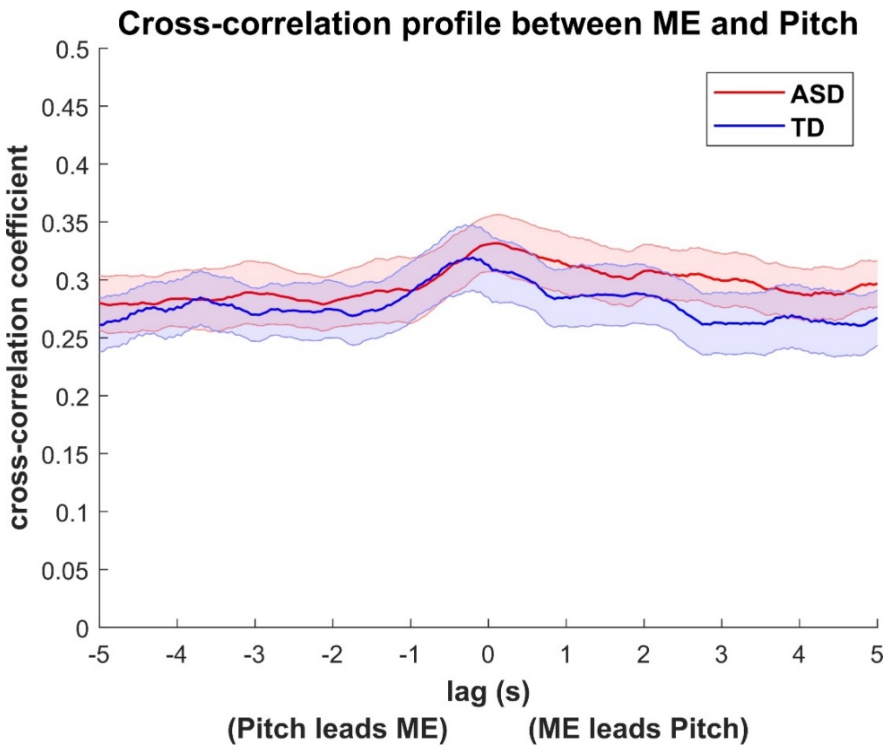

Figure 4.21 Average profile of cross-correlation between movement (ME) and pitch (f0) between the ASD group (in red) and the TD group (in blue).
Shadowed areas represent the standard error of the mean.

In addition to the participants who presented no peak, one ASD and two TD participants were excluded from the analysis because their peak characteristic values were outside of ±2.5 standard deviations of the mean of their group. Inspection of the participants revealed no obvious differences between the excluded participants and the rest of the sample. Figure 4.21 shows the averaged profile of cross-correlation between groups. Because the main peak in the cross-correlation profiles had an average width (measured at half-prominence) of around 1400ms, we chose a window of integration of [-1500 to -50ms] on the left and [50 to 1500ms] on the right of lag zero to measure skewness.

Planned independent samples t-tests were conducted to compare peak characteristics and skewness of the curve between groups on the remaining 13 ASD and 12 TD participants. Results are reported in table 4.9.

	ASD (n=13)	TD (n=12)	t-test		Cohen's d
			t	p	
Number of peaks					
mean *(SD)*	**1.47** *(1.18)*	**1.75** *(1.24)*	-.664	.512	.231
Range (min:max)	0:4	1:5			
Peak correlation coef					
mean *(SD)*	**.327** *(.103)*	**.315** *(.074)*	.334	.742	.134
Range (min:max)	0.163:0.445	0.188:0.424			
Peak lag					
mean (ms) *(SD)*	**519** *(1618)*	**-296** *(1098)*	1.461	.158	.375
Range (min:max)	-1850:3300	-3300:1500			
Peak width					
mean (ms) *(SD)*	**1433** *(1015)*	**1386** *(474)*	.144	.887	.059
Range (min:max)	440:4057	781:2175			
Peak prominence					
mean *(SD)*	**.071** *(.028)*	**.087** *(.027)*	-1.389	.178	.582
Range (min:max)	0.029:0.124	0.048:0.143			
Skew (Area R - Area L)	($x10^{-3}$)	($x10^{-3}$)			
mean *(SD)*	**7.778** *(24.047)*	**-4.419** *(27.979)*	1.172	.253	.468
Range (min:max)	-45.540:50.362	-57.919:39.151			

Table 4.9 Statistics of independent samples t-tests conducted to compare the characteristic of the main peak in the cross-correlation profile between motion energy and pitch.

Results showed no significant group difference between any of the cross-correlation peak characteristics or profile skewness between motion energy and pitch.

However, in the ASD group, we found that the peak lag correlated significantly with ADOS communication score ($r(13) = .678$, $p = .011$), specifically participants for whom the lag was more positive (indicating that ME was generally leading pitch) also tended to have a higher ADOS score in the communication domain (see figure 4.21). In addition, results showed that correlation at peak correlated marginally with ADOS restricted and repetitive behaviour (RRB) score ($r(13) = .526$, $p = .065$) and peak prominence correlated marginally with total ADOS score ($r(13) = .478$, $p =$

.99). Both marginal results suggest that a higher degree of coupling between motion energy and pitch was associated with greater difficulties as measured by the ADOS.

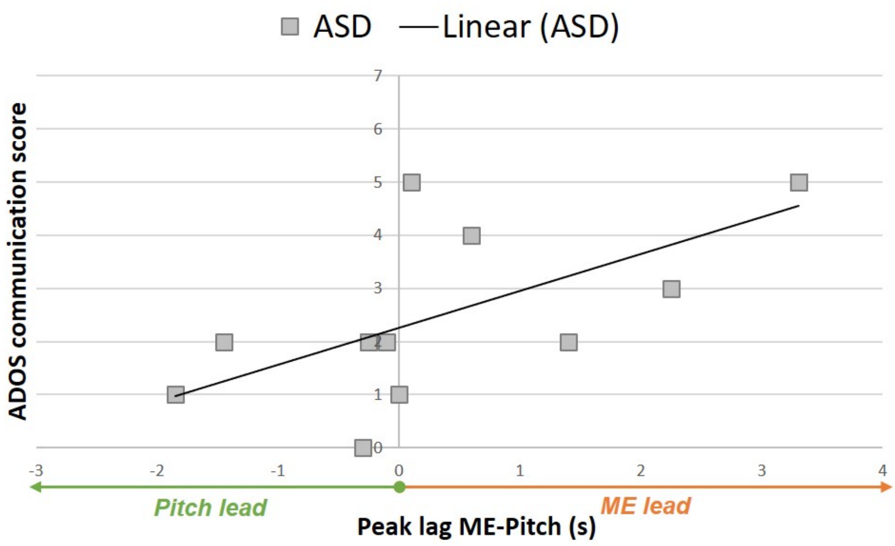

Figure 4.22 Relationship between peak lag of the cross-correlation profile between motion energy and pitch (in seconds) and ADOS communication score.

Interestingly, we found that peak prominence was negatively associated with VIQ (r(25) = -.434, p = .030), PIQ (r(25) = -.424, p = .035) and FIQ (r(25) = -.440, p = .028, see figure 4.23), suggesting that a higher degree of correlation between motion energy and pitch was associated with lower cognitive skills. There were no significant correlations between peak characteristics and quality of communication ratings.

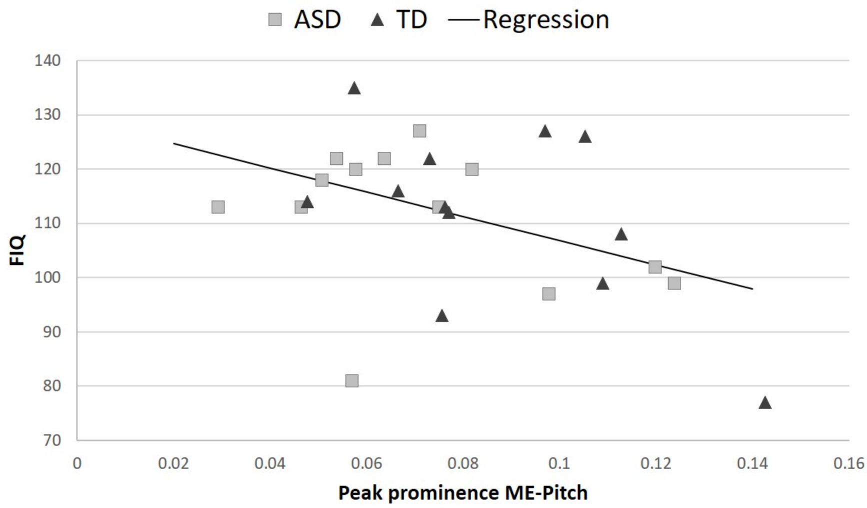

Figure 4.23 Relationship between peak prominence of the cross-correlation profile between motion energy and pitch and FIQ.
The correlation was computed across groups, datapoints for ASD (light grey squares) and TD (dark grey triangles) participants are differentiated for illustration purposes.

4.21.2.3 Cross-correlation between loudness and motion energy

Peak analysis on the cross-correlation profiles of ME and loudness detected a single peak in all participants except 1 ASD and 2 TD participants for which no peak was detected, and 1 ASD participant and 1 TD participants showing 2 peaks. Closer inspection of the participants showing no peaks revealed that one of them was the ASD participant whose hands were partly hidden behind a table during the recording. the other two participants showed few gestures during the recording. The ASD participant showing a 2-peak profile also produced a very small amount of gestures and their peak characteristics data were outside ±2.5 standard deviation of the mean of their group, and therefore excluded. The TD participant showing a 2-peak profile also produced few gestures overall. Finally, one TD participant for whom peak data were outside ±2.5 standard deviation of the mean of their group was excluded. Overall 15 ASD and 13 TD participants were included in the analysis. Figure 4.23 shows the averaged profile of cross-correlation between groups. Because the main peak in the cross-correlation profiles had an average width (measured at half-prominence) of around 2000ms, we chose a window of integration of [-2000 to -50ms] on the left and [50 to 2000ms] on the right of lag zero to measure skewness.

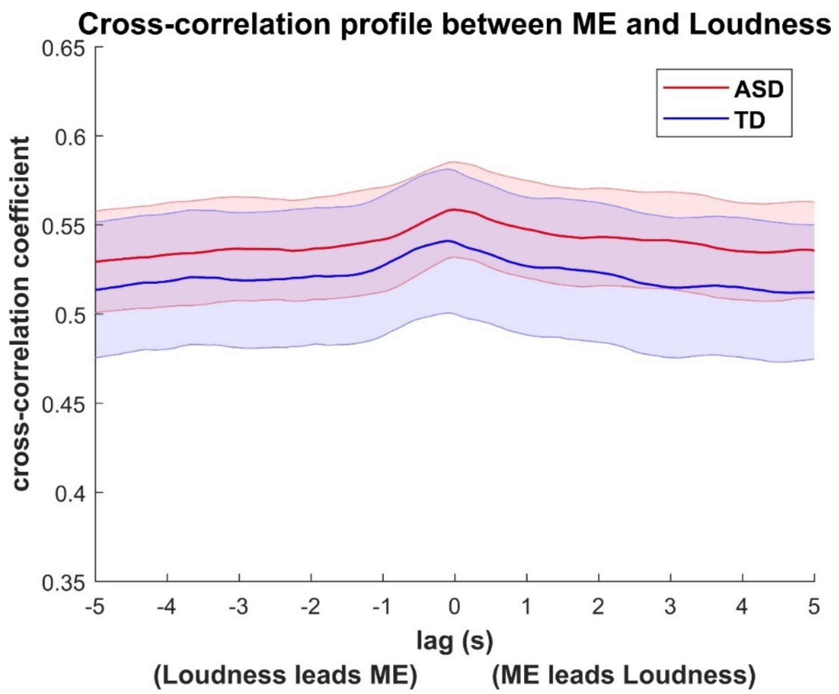

Figure 4.24 Average profile of cross-correlation between motion energy and loudness between the ASD group (in red) and the TD group (in blue).
Shadow areas represent the standard error of the mean.

Planned independent samples t-tests were conducted to compare peak characteristics and skewness of the curve between groups on the remaining 15 ASD and 13 TD participants. Results are reported in table 4.10. Comparison results claimed no significant group differences in the characteristics of the cross-correlation peak between motion energy and loudness.

	ASD (n=15)	TD (n=13)	t-test		Cohen's d
			t	p	
Number of peaks					
mean *(SD)*	**1.06** *(.43)*	**0.94** *(.44)*	.800	.430	.276
Range (min:max)	0:2	0:2			
Peak correlation coef					
mean *(SD)*	**.566** *(.107)*	**.534** *(.152)*	.666	.511	.243
Range (min:max)	0.331:0.720	0.264:0.713			
Peak lag					
mean (ms) *(SD)*	**23** *(696)*	**-58** *(276)*	.393	.698	.153
Range (min:max)	-950:2150	-450:650			
Peak width					
mean (ms) *(SD)*	**2279** *(1004)*	**1861** *(628)*	1.297	.206	.499
Range (min:max)	1084:4382	1259:3583			
Peak prominence					
mean *(SD)*	**.027** *(.013)*	**.032** *(.012)*	-1.045	.306	.400
Range (min:max)	0.012:0.058	0.013:0.059			
Skew (Area R - Area L)	($x10^{-3}$)	($x10^{-3}$)			
mean *(SD)*	**3.270** *(11.129)*	**-0.060** *(13.888)*	.704	.488	.265
Range (min:max)	-10.245:31.332	-35.464:19.339			

Table 4.10 Statistics of independent samples t-tests conducted to compare the characteristic of the main peak in the cross-correlation profile between motion energy and loudness.

However in the ASD group we found a significant correlation between peak prominence and ADOS communication (r(15) = .571, p = .026), reciprocal social interaction (r(15) = .664, p = .007) and total (r(15) = .774, p = .001, see figure 4.24) scores, indicating that a higher degree of coupling between motion energy and speech loudness was associated with higher clinical scores, and in particular with greater difficulties in the social domain (communication and social interaction).

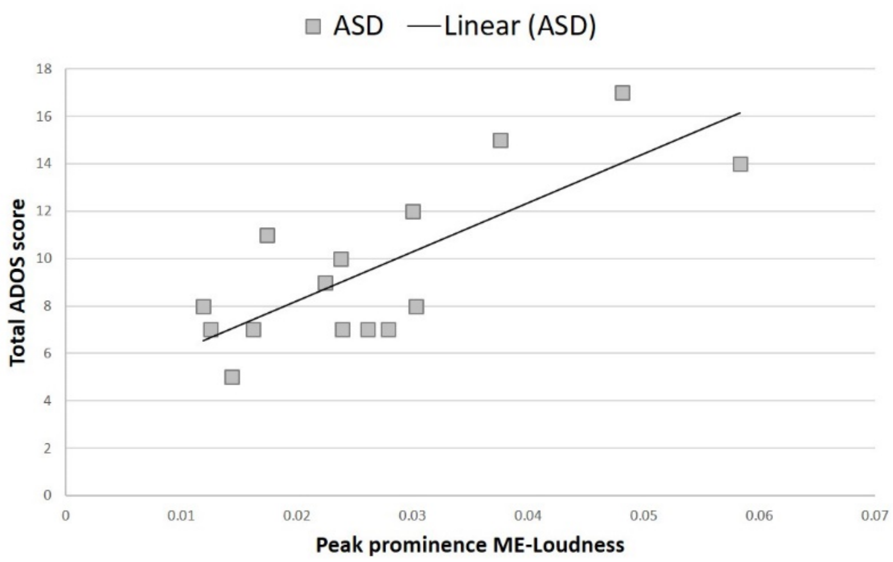

Figure 4.25 Relationship between peak prominence of the ME-loudness cross-correlation profile and total ADOS score for ASD participants.

Surprisingly, we found a marginal correlation between peak width and visual gain (r(27) = -.376. p = .053) and a significant correlation between the skew of the cross-correlation profile and visual gain (r(27) = -.402, p = .038), indicating that longer coupling interval and greater proportion of motion energy-lead were associated to smaller visual gain, although figure 4.25 where the relationship between skew and visual gain is plotted suggest that the correlation may be accentuated by the position of one particular TD datapoint.

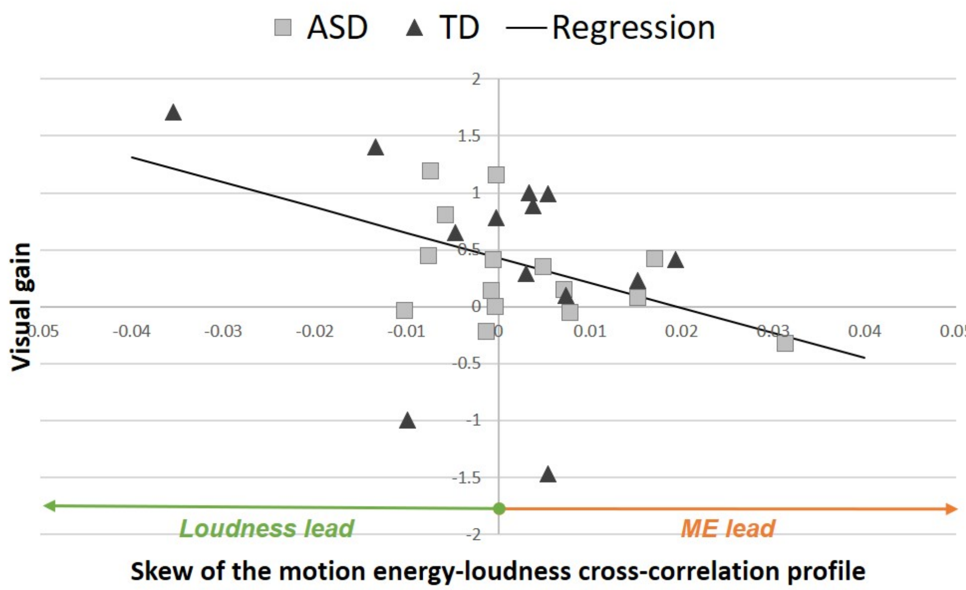

Figure 4.26 Relationship between the skew of the cross-correlation profile between motion energy and loudness and visual gain.
The correlation was computed across groups, datapoints for ASD (light grey squares) and TD (dark grey triangles) participants are differentiated for illustration purposes.

4.21.2.4 Cross-correlation between pauses and holds

Last, peak analysis of the cross-correlation profiles of pauses and holds detected a single peak in 8 ASD and 12 TD participants, two peaks in 7 ASD participants and 2 TD participants, and three peaks in 1 ASD and 2 TD participants. In addition, no peak was detected for one ASD participant (once again the participant whose hands were partly masked by a table during recording). In addition to the ASD participant who showed no peak, one TD participant was excluded from the analysis because their peak characteristic data fell outside of ±2.5 standard deviation of the mean of their group. Inspection of this participant's data did not reveal any striking differences with the rest of the group.

Overall 16 ASD and 15 TD participants were included in the analysis. Figure 4.26 shows the averaged profile of cross-correlation between groups. Because the main peak in the cross-correlation profiles had an average width (measured at half-prominence) of around 2000ms, we chose a window of integration of [-2000 to -50ms] on the left and [50 to 2000ms] on the right of lag zero to measure skewness.

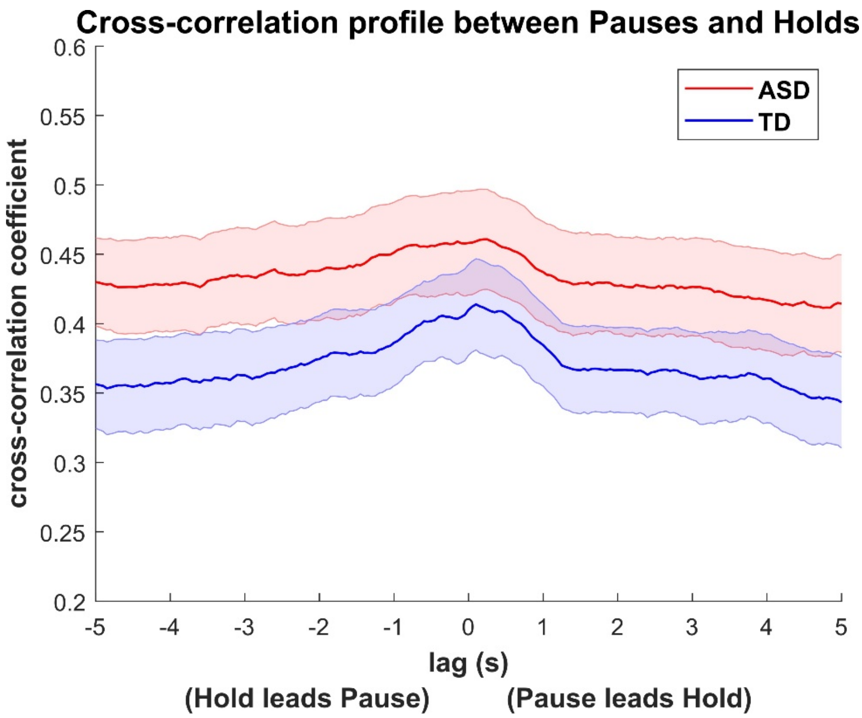

Figure 4.27 Average profile of cross-correlation between speech pauses and motion holds between the ASD group (in red) and the TD group (in blue).
Shadow areas represent the standard error of the mean.

Planned independent samples t-tests were conducted to compare peak characteristics and skewness of the curve between groups on the remaining 16 ASD and 15 TD participants. Results are reported in table 4.11.

Group comparison revealed that the correlation coefficient at peak was marginally greater in the ASD compared to the TD group, with a medium to large effect size. This suggests again that autistic participants seemed to demonstrate a greater degree of coupling between voice and gesture, or in this case their "negative", the pause and hold behaviour.

	ASD (n=16)	TD (n=15)	t-test		Cohen's d
			t	p	
Number of peaks					
mean *(SD)*	**1.47** *(.72)*	**1.38** *(.72)*	.382	.705	.125
Range (min:max)	0:3	1:3			
Peak correlation coef					
mean *(SD)*	**.485** *(.131)*	**.410** *(.080)*	1.923	.064	.691
Range (min:max)	0.225:0.656	0.262:0.544			
Peak lag					
mean (ms) *(SD)*	**-78** *(1412)*	**-23** *(375)*	-.145	.885	.053
Range (min:max)	-3350:2800	-700:550			
Peak width					
mean (ms) *(SD)*	**2021** *(1067)*	**1925** *(1048)*	.252	.803	.091
Range (min:max)	584:4087	782:4209			
Peak prominence					
mean *(SD)*	**.063** *(.040)*	**.084** *(.038)*	-1.506	.143	.538
Range (min:max)	0.021:0.157	0.033:0.133			
Skew (Area R - Area L)	($x10^{-3}$)	($x10^{-3}$)			
mean *(SD)*	**-9.795** *(31.290)*	**-6.321** *(25.586)*	-.337	.738	.122
Range (min:max)	-84.150:33.607	-48.981:60.097			

Table 4.11 Statistics of independent samples t-tests conducted to compare the characteristic of the main peak in the cross-correlation profile between pause behaviour and hold behaviour.

Correlations between pause-hold cross-correlation data and ADOS scores in the ASD group brought further information. We found that the number of peaks in the profile correlated negatively with the ADOS communication score (r(17) = -.673, p = .003). In addition, the correlation coefficient at peak correlated positively with the ADOS RRB score (r(16) = .557, p = .025). We also found that peak width was associated with ADOS communication (r(16) = .636, p = .008) and RRB (r(16) = .608. p = .013) scores. Finally peak prominence was related to total ADOS score (r(16) = .523, p = .038).

Across groups, we found a significant correlation between peak width and VIQ ($r(31) = .363$, $p = .045$).

Finally, we found that peak width was positively associated with both audio-only ($r(30) = .405$, $p = .026$) and audiovisual ($r(30) = .427$, $p = .019$) quality of communication ratings, suggesting that a wider interval of coupling between pause and holds was related to increased quality of communication in both modalities.

4.21.3 Summary and provisional discussion

Cross-correlational analysis was conducted as an attempt to replace manual coding analysis with a fast, automatic, unbiased procedure. Here we first summarise the results obtained from cross-correlational data before comparing it to the previous temporal lock analysis data and our initial predictions.

The general cross-correlation profiles and the number of peaks provided some intuitions regarding the coupling of acoustic and kinematic data. As expected, the profile of speech pitch and loudness showed a unique, narrow peak centred on lag zero, with correlation coefficients ranging in a fairly high range (.55-.6) indicating a tight near-synchronous coupling between the two acoustic signals. The cross-correlation profiles between motion energy and acoustic features showed somewhat flatter curves. The coupling of motion energy and loudness was roughly centred on lag zero and correlation coefficients ranged fairly high (.5-.55), whilst the coupling of motion energy and pitch was more variable across participants and the lag between ME and pitch took values which were typically away from lag zero by a few hundred of milliseconds. Correlation coefficients for motion energy and pitch also averaged lower than those for motion energy and loudness (approximately .3). ME and pitch also produced more complex patterns of peaks whereas ME and loudness produced a majority of single-peak profiles. Altogether, this suggests that motion energy and loudness are more tightly coupled in time and might serve functions which are more similar compared to motion energy and pitch. Finally the cross-correlation profile between pause and hold behaviour suggested a moderate degree of coupling: profiles were characterised by one to three peaks, with moderate correlation coefficient (ranging approximately between .35 and .45). The profiles were quite flat but peaks were on average centred near lag zero, suggesting that overall pauses and holds were coupled in time.

Peak analysis revealed few differences in the direct comparison of ASD and TD participants. The pitch-loudness cross-correlation profiles of ASD participants demonstrated a skew towards loudness-lead, whereas profiles of TD participants were skewed towards pitch lead. In light of the results from the descriptive analysis which found that recordings were louder in the ASD group compared to the TD, we tentatively hypothesise that ASD individuals might use loudness as a more dominant prosodic cue than pitch, whereas TD individuals might use pitch dominantly compared to loudness. This is highly speculative, but it could be tested by removing either pitch or loudness information from recordings of ASD and TD participants and evaluate the impact on comprehension and quality of communication for the listeners (using an approach similar to that of Scherer, Feldstein, Bond, & Rosenthal, 1985). The second direct difference between groups consisted in a greater correlation coefficient at peak between pauses and holds in the ASD compared to the TD group. This result suggest that the temporal pattern of pauses and holds were more tightly coupled in ASD.

The most prominent result perhaps in this section is the accumulation of correlational evidence towards the idea that overall, clinical features of ASD (as measured by the ADOS scores) are associated with a greater degree of coupling between all signals. Although in the context of multiple tests, single correlations did not provide strong evidence for any effect and should be interpreted with caution, the repeated correlations which suggest a similar trend provide convincing evidence. In particular, the correlation coefficient at peak and prominence were associated with higher ADOS scores in the communication (in pitch-loudness, ME-loudness and pause-holds profiles), reciprocal social interaction (in ME-loudness profile) and restrictive and repetitive behaviour domains (in ME-pitch and pause-hold profiles) and with higher total ADOS scores (in all profiles). This is at odds with our prediction that, following de Marchena and Eigsti's (2010) interpretation, temporal coupling between speech and gestures would be less tightly coupled in the ASD compared to the TD group. There are many methodological reasons why results could differ: the type of data (detailed manual annotations vs automatically extracted features), the amount of information (gesture segments vs the entire recording), the choice of analysis, but perhaps the most important is the definition of time-coupling in others and our study. Previous studies (de Marchena & Eigsti, 2010; Morett et al., 2016) based their conclusions about how tightly time-locked gestures and speech were on the absolute temporal lag between speech and gesture. However, whilst the lead-lag relationship between two time series can be characterised by the mean lag between them, the degree of coupling is related to the variability of the lag (or in our case the degree of correlation). For instance, we would expect that the departure of a train from station A and its arrival an hour later in station B are more tightly coupled in time that the departure of the same train and the departure of a bus scheduled 15 minutes later. Although the lag between events is shorter in the second example, in the first example the two events are causally linked, whereas in the second example the events simply happen to occur in close proximity. Specifically, a delay in the train departure at A should result in a delay at arrival in B (conserving the coupling), whereas a delay in train departure should not directly affect the departure of the bus (decreasing the degree of coupling). We therefore argue that the main difference between previous and our study resides in the characteristic time-delay and the lead-lag relationships between signals (which were discussed in the previous section on temporal lock and will be elaborated on in the main discussion), but that our data provide evidence showing that temporal lock between speech and gesture, or at least acoustic and kinematic features of communication, are more tightly coupled in time in ASD. This conclusion is comforted by similar results obtained in the previous section in the temporal lock analysis, where we found that TD participants showed a significantly higher variability of the lag between speech and gesture for both iconic and metaphoric gestures (which represented the majority of gestures in the recordings).

Another trend provided by the correlational data is that a higher degree of coupling between signals (as measured by the peak prominence) was also associated with lower IQ scores. This was the case for pitch-loudness, ME-pitch and ME-loudness cross-correlation profiles, and suggests that cognitive abilities at large are engaged to produce acoustic and kinematic signals with some degree of variability. Whilst we know temporal coupling is necessary in order to produce and integrate speech and gesture successfully (Kendon, 1980; Leonard & Cummins, 2011; McNeill, 1992; Wachsmuth, 1999; Wiltshire, 2007), it appears that an excessive degree of coupling might affect communication phenotype. More research is needed to establish whether this relates to enhanced predictability of communication behaviour (higher degree of coupling means that one signal better predicts the other temporally), or reduced information overall (temporal information contained in one signal is repeated in the other signal).

Surprisingly, but in line with the previous section on temporal lock, the coupling between speech and gesture signals also showed little relation to quality of communication outcome. The most convincing association was that wider peak width in pauses-holds profiles were associated with greater quality of communication ratings in both audio-only and audiovisual modalities. This suggests that quality of communication benefits from some variability in the delay within which coupled pauses and holds actually occur. The very limited literature on pause and holds coupling only proposes the notion that pauses and holds co-occur and communication and serve similar purposes in relation to speech for pauses and gestures for holds (Esposito & Esposito, 2011; Esposito & Marinaro, 2007). Our data suggest that pauses and holds might be functionally (rather than incidentally) coupled, and contribute to the overall efficiency of communication.

Overall, cross-correlation analysis fulfilled its objective to provide a fast, automatic procedure to quantify temporal coordination of speech and gesture and characterise the coupling of motion energy, pitch, loudness, pause and hold behaviour. However where we gained information in terms of coupling characteristics between speech and gesture, we lost the fine-grained analysis information provided by manual annotation of gestures, in particular the different dynamics presented by different types of gestures. Importantly for this work, it proved difficult to validate the cross-correlation analysis as the results obtained from temporal lock analysis and cross-correlation analysis showed different strengths: temporal lock was more informative with regards to the characterisation of specific types of gestures and their relationship to quality of communication, whereas cross-correlational data provided information about the degree of coupling between signals and their relation to diagnostic features. The main common results between the two analysis is a lack of evidence that gesture precede speech, and that the time delay between speech and gesture did not differ overall between groups, both at odds with results from previous studies (de Marchena & Eigsti, 2010; Morett et al., 2016; Morrel-Samuels & Krauss, 1992). Thus firm conclusions are undermined by the possibility that the study is altogether underpowered. We believe that cross-correlation analysis has potential to provide a fine-grained, automatic characterisation of speech and gesture dynamics, but sample size must be significantly increased in future research in order to reduce the need to use manually coded information to select the data and minimise the effect of noise in the data.

Reflecting on the results so far, the scarcity of group differences suggest that temporal processing differences between groups might not be best described by the ability to produce events within a set interval. Characteristic delays for the temporal coordination of pitch, volume and motion energy span between a few milliseconds and approximately two seconds, which is a range of durations which time perception studies have flagged as a potential domain of impairment or reduced precision in autism. The results so far suggest that whilst performance in psychophysics tasks may be atypical in ASD, it does not seem to directly affect the ability to produce such intervals between coupled events in the context of communicative behaviour with respect to the coordination of speech and gesture. Instead, variability analysis and correlational data suggest that the coupling between speech and gesture is tighter in ASD, and relates to clinical (ADOS scores) cognitive (IQ) and subjective measures (quality ratings) of communication skills, and could therefore play a role both in the autistic phenotype and in the outcome of clinical observations such as the ADOS. This point will be discussed more in depth in the main discussion.

Part E – Recurrence patterns of speech and gesture

4.22 Introduction

Despite some informative characterisation of lead-follow relationships and degree of coupling between speech and gestures provided in the previous section (part D), the results failed to entirely capture the qualitative atypicalities perceived in autistic speech and gesture such as the fact that gestures did not contribute as much to the quality of communication in ASD compared to TD speakers (Chapter 4, part B, section 3). Notably, correlational evidence showed that the lag between iconic gestures and pitch was related to visual gain, but it did not differ between groups. Evidence from the cross-correlation analysis (Chapter 4, part D, section 1.2) was somewhat more successful in relating temporal dynamics of speech and gesture to clinical scores and quality of communication outcome. Results suggest that tighter coupling between acoustic and kinematic signals is associated with lower quality of communication and greater autistic traits as measured by the ADOS, however the degree of coupling did not differ between groups overall. Yet tools like the ADOS provide a clear indication that trained observers detect qualitative abnormalities in communicative behaviours in ASD. Therefore either the ratings on the ADOS are the result of biases, or the methods used this far are not sensitive enough to capture the qualitative differences that observers are sensitive to.

An important aspect of communication which is not captured by descriptive statistics, temporal lock or cross-correlation measures is recurrence, in other words how much some fragments of information are repeated (partially or completely) in communication. Social interaction is based on how well we can predict the other's behaviour, and how much is new or unexpected (Vesper, Van Der Wel, Knoblich, & Sebanz, 2011). By using recurrence in communication, we make it easier for our interlocutor to predict and follow what we are conveying (Scarborough, Cortese, & Scarborough, 1977), but to introduce texture, intonation and new ideas, and adapt to the context, we need to break out of recurring patterns, which is also important for allowing us to adapt flexibly to the patterns of communication of others (Fusaroli & Tylén, 2015).

4.22.1 Inter-individual recurrence

One of the most intuitive forms of recurrence is the alignment or interactional synchrony between conversational partners (inter-individual recurrence): partner A produces a certain behaviour (e.g., using a particular word to describe the topic of the conversation) and conversational partner B aligns with A by repeating or using the same behaviour (e.g., using the same word which is now established as common ground). Such interactional synchrony applies to various facets of communication including vocabulary and syntax (Brennan & Clark, 1996; Garrod & Pickering, 2009; Pickering & Garrod, 2004), laughter and yawn (Hatfield, Cacioppo, & Rapson, 1994), posture (Shockley, Santana, & Fowler, 2003) and eye gaze patterns (Richardson & Dale, 2005). Various degrees of recurrence are used in social interaction to establish common ground and the degree of alignment predicts empathy with and affiliation with the interactional partner (Chartrand & Bargh, 1999; Hove & Risen, 2009) as well as successful comprehension and better cooperation between conversational partners (Fusaroli et al., 2012; Garrod & Pickering, 2009; Richardson & Dale, 2005). Alignment or synchronisation between interactional partners is believed to be particularly important during early development as evidence indicates that coordinated movements between infants and their caretakers predict later social, cognitive and communicative skills (Jaffe et al., 2001) and attachment style (Isabella, Belsky, & von Eye, 1989). For instance, the behaviour of mothers interacting with their 8 to 9 weeks old infants via a camera display was significantly different when they watched a video of their child in real-time (live condition), compared to when they watched (unknowingly) a delayed video recording of their child (replay condition) (Murray & Trevarthen, 1986). In particular, in the live video condition, mothers used shorter utterances, more repetitions and references to the infants activity, whereas in the replay condition they used more complex utterances, less repetitive and more utterances about themselves. The authors interpret these results by arguing that in the live condition, the infant's motor behaviour responds directly (and crucially, in time) to their mother's, which directly affects the quality of rapport and interaction between them. More recently, 14 months old infants were found to engage more in pro-social behaviour (helping to pick up an object) towards an experimenter when they had been bounced in synchrony with them rather than asynchronously (Cirelli, Einarson, & Trainor, 2014). Four to six years old children show a similar effect: children were more likely to engage in spontaneous helping behaviour after playing synchronously than asynchronously with their peers, and showed greater amounts of mutual smiling and eye contact (Tunçgenç & Cohen, 2018).

Research indicates that inter-individual recurrence may be disrupted in autism: one of two monozygotic twins who was later diagnosed with autism showed disrupted synchrony of social behaviour in infancy (11 months) compared to her sister, which lead to frustrating attempts from her caretakers to support her and resulted in reduced opportunities for mutual attention and joint activity (Trevarthen & Daniel, 2005). In adults, typical listeners synchronised their eye blinks with a speakers' breakpoints in speech (speech pauses) when they can see their face (Nakano & Kitazawa, 2010), but autistic listeners did not (Nakano, Kato, & Kitazawa, 2011), suggesting that autistic listeners do not spontaneously align to their interactional partner in time. Together, evidence suggests that inter-individual recurrence may be reduced in autism, with possible consequences on communication skills.

4.22.2 Intra-individual recurrence

Recurrence also occurs at the intra-individual level and makes our behaviour consistent and more predictable for our interaction partners. Using a task in which two participants were asked to coordinate their motor responses, Vesper and colleagues (2011) showed that each individual reduced the temporal variability of their own responses in the context of joint action compared to when the action was performed individually, in order to make it more predictable for their partner and facilitate joint behaviour. Self-consistency also applies between modalities of communication like speech and gesture. For instance, De Jonge-Hoekstra, Van der Steen, Van Geert, and Cox (2016) investigated recurrence patterns in speech and gesture produced by children who were engaged in a hands-on science exercise which required them to explore a device (an air-cannon) by themselves to deduct how it worked. Both speech and gesture time series were coded so as to indicate which level of skill was expressed in a particular utterance or gesture (e.g., the degree of abstraction displayed in understanding the task). Specifically, verbal and non verbal productions were coded according to whether they related to a descriptive, predictive or explanatory aspect of the device, on a scale of 1 (sensory-motor: “it has a long tube”) to 7 (an abstraction beyond the device itself, e.g., a statement about air pressure in general). Results showed that gestures were more regularly and more rigidly staying within the same tier of skill level (or “trapped”) compared to speech. The coupling between speech and gestures showing higher levels of understanding was also found to be more stable for older children, suggesting that conceptual understanding and the ability to communicate in a self-consistent way both increase with age. In this case, recurrence of behaviour is directly linked to the content of communication and supports the conceptual complexity of the message. These results illustrate how recurrence patterns can reflect cognitive processes and support successful comprehension between interactional partners.

There are also reasons to believe that temporal recurrence should differ at the intra-individual level in autism. First, evidence from the literature on time perception in ASD and from the first part of this work suggests that subtle differences in temporal processing might impact self-consistency in producing behaviours, especially for individuals with less cognitive skills. For instance, whilst typical individuals tend to increase self-consistency (reduce temporal variability) in the context of joint action (Vesper et al., 2011), this may prove difficult for autistic individuals whose show difficulties with processing short durations. In this case, the prediction is that intra-individual recurrence in communication should be lower in ASD, and as a result comprehension could be diminished for their conversational partner, or require a greater effort as the behaviour of the autistic individual would be less predictable and less easy to synchronise with to make themselves better understood (**hypothesis 1**). Note that, because under this assumption autistic individuals would show difficulties in controlling the timing of their behaviour, we would predict that they would also show reduced adaptation to their partner’s communication style, and therefore expect inter-individual recurrence to be lower than in TD individuals (this prediction will not be tested in the context of this study). Second, cross-correlation data in the previous section suggested that autistic individuals produced a higher degree of coupling (or reduced temporal variability) in speech and gesture behaviour. In addition, autistic individuals clinically present a higher rate of restricted and repetitive behaviours (RRBs) at a macroscopic level. Although the mechanisms behind RRBs are not fully understood, some evidence suggests that some RBBs have a soothing function which is highly correlated to anxiety. In a situation where individuals feel challenged, stressed or anxious, they show a greater desire for sameness and control of the environment (Leekam, Prior, & Uljarevic, 2011), which leads them to produce repetitive, highly predictable behaviours (Wigham, Rodgers, South, McConachie, & Freeston, 2015). Assuming that this tendency extends to microscopic behaviours, the prediction is that autistic individuals should produce highly recurrent patterns in their intra-individual communication behaviour. If this is the case, their behaviour should be more predictable and comprehension should be therefore be facilitated for their listener (**hypothesis 2**). Although inter-personal recurrence is not addressed in this study, in the context of reciprocal interaction, the tendency to produce highly repetitive behaviour would be expected to interfere with inter-personal recurrence, making it more difficult for autistic speakers to provide new information, flexibly adapt to their partner's communication style and repair disfluencies in the conversation.

In this section, we quantify the recurrence patterns within acoustic and kinematic features of speech and gesture, and the dynamic coupling between these systems. To that end, we use the concept of "**structural organisation**" introduced by Fusaroli and Tylén (2015), which consists of (a) the extension of stable patterns in behaviour and (b) the complexity of their structure. The extension defines the length of a given repeated pattern, for instance the reiteration of a prosodic pattern (for speech) or the repetition of a gestural segment (for movement). Because repetitions are rarely identical, we also measure the complexity of the recurrent structures as the variability presented in these recurrences.

4.23 Recurrence and cross-recurrence quantification analysis

Structural organisation can be quantified using recurrence or cross-recurrence quantification analysis (RQA or CRQA respectively), which is a method for non-linear data that is based on the theory of coupled dynamical systems (Marwan, Carmen Romano, Thiel, & Kurths, 2007). RQA is a particularly powerful method for comparative analysis as it provides indices of regularity and complexity in a dynamic system, within and across individuals. We speak of RQA when quantifying the recurrence within the same signal and CRQA when quantifying the recurrence between two distinct signals. Here we use the definitions proposed by Marwan et al. (2007), who provide further details in their paper. Recurrence $R_{i,j}$ at coordinates i and j can be expressed as:

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \| \vec{x_i} - \vec{x_j} \|), \qquad i, j = 1, \dots, N$$

Where N is the number of measured points $\vec{x_i}$, Θ it the Heaviside step function which takes the values 0 or 1 when its argument is negative (<0) or positive (≥0), respectively, and ε is a threshold distance.

For states that are in an ε-neighbourhood (ε-recurrent states):

$$\vec{x_i} \approx \vec{x_j} \iff R_{i,j} \equiv 1$$

Methodologically speaking, RQA assesses the points in time when two systems enter similar states, and provides both a quantification of recurrence between two systems and a graphical representation of the dynamics of coupled systems: at coordinate (i,j), a black dot is plotted if $R_{i,j} \equiv 1$, and white dot is plotted if $R_{i,j} \equiv 0$. Figure 4.28 (reproduced with permission from Richardson & Dale, 2005) illustrates how RQA operates. The data is based on a task where a speaker and a listener were discussing characters of a popular TV series whilst looking at pictures of them on a screen. The two systems consist of the eye-movement data of the speaker (system one) and the listener (system two). System states are defined as the fixations on predefined areas of interest on the screen, specifically the six pictures of the TV series characters. The states are compared at each time t. If the systems are in the same state (i.e., the speaker and the listener are looking at the same picture at time t), the cross-recurrence is 1; if they are in different states the cross-recurrence is 0 (figure 4.28, left panel). The same procedure is repeated when the two signals are aligned with various time lags between the two signals, evaluating whether the two systems are coupled with a delay (figure 4.28, right panel). A cross-recurrence plot can then be drawn representing the moments the systems are coupled for each lag (figure 4.28, middle panel).

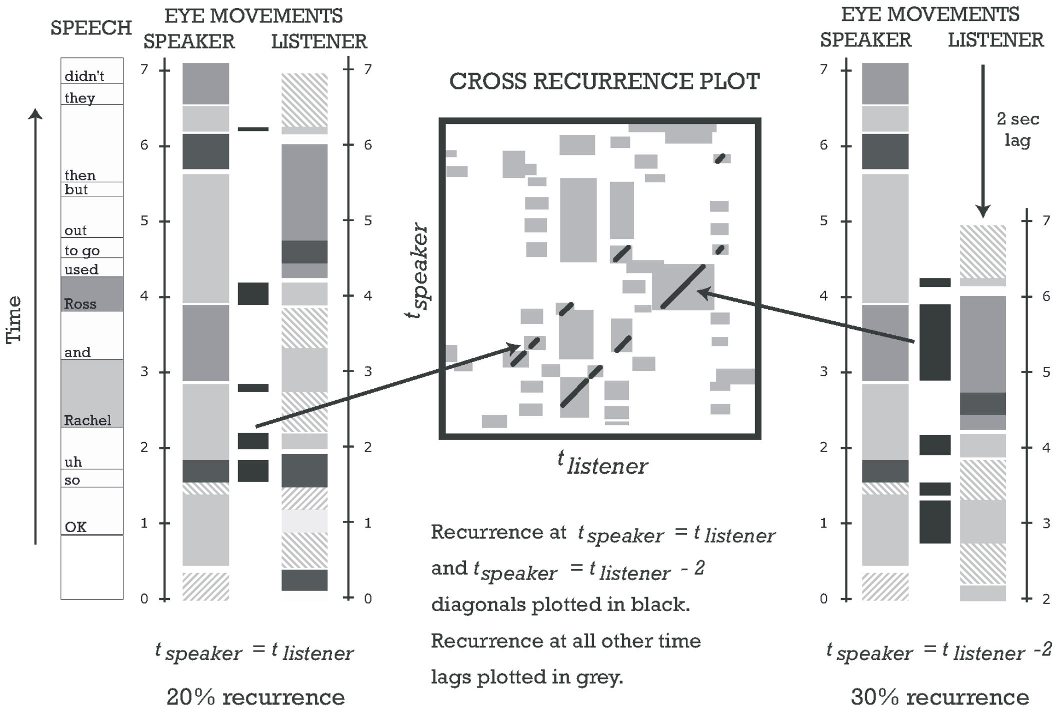

Figure 4.28 Diagram illustrating how cross-recurrence is established and a cross-recurrence plot is built.
Reproduced with permission from Richardson and Dale (2005). In this example, the two systems consist of eye-movement data from a speaker and a listener discussing characters of a popular TV series whilst looking at pictures of them. On the left, recurrence is quantified when the two systems are aligned in time. On the right, recurrence is quantified when the signals are aligned with a lag of 2 seconds. Recurrence at each time lag form a diagonal line in the cross-recurrence plot (middle section).

Cross-recurrence plots represent in an instant, visual way the coupling between systems, and each of their characteristics can be extracted as a quantitative variable to further characterise the relation between the systems. The density of the plot (proportion of "dark") or percentage of recurrence points on the plot is a raw measure of the degree to which the systems are coupled, and is called the **recurrence rate (RR)**:

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}(\varepsilon)$$

In the example above, the recurrence rate computes the amount of time when speaker and listener look at the same picture on the screen (with or without a delay). Further information is derived from the patterns formed by the recurrence points. Diagonals indicate that the two systems share the same paths (e.g., if the speaker looks at picture A then B then D, and that the listener look at the same sequence of pictures with a delay δ, a diagonal will form on the plot parallel to the main diagonal, with a distance δ to the main diagonal):

$R_{i+k,j+k} \equiv 1 \begin{cases} l-1 \\ k=0 \end{cases}$ defined as $\left(1 - R_{i-1,j-1}\right)\left(1 - R_{i+l,j+1}\right) \prod_{k=0}^{l-1} R_{i+k,j+k} \equiv 1$

where l is the length of the diagonal line.

Determinism (DET) captures the rate of recurrence points forming diagonals:

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)}$$

where P(ε,l), abbreviated as P(l), is the histogram of diagonal lines of length l:

$$P(\varepsilon, l) = \sum_{l,j=1}^{N} \left(1 - R_{i-1,j-1}(\varepsilon)\right)\left(1 - R_{i+l,j+l}(\varepsilon)\right) \prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon)$$

The **length of diagonal (L)** characterises the time during which the two systems stay coupled and therefore the stability of the coupling:

$$L = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}$$

The **length of the longest diagonal (Lmax)** reflects the longest uninterrupted time the two systems stay attuned:

$$L_{max} = \max\left(\{l_i\}_{i=1}^{N_l}\right)$$

where N_l is the total number of diagonal lines:

$N_l = \sum_{l \geq l_{min}} P(l)$A more abstract measure, **entropy (ENTR)** reflects how complex or organised the structure is in the system (a random display would be characterised by high entropy, whereas a highly organised display would be characterised by low entropy). ENTR is defined as the Shannon entropy of the probability p(l)=P(l)/N_l to find a diagonal line of exactly length l in the recurrence plot:

$$ENTR = -\sum_{l=l_{min}}^{N} p(l) \ln p(l)$$

The propensity of a system to stay in the same region (i.e., repeat the same value) is quantified in the vertical lines. For instance, if the listener looks at picture A at time t and looks away at time t+1, but the speaker looks at picture t during time t to time t+10, the recurrence plot shows a vertical line of length 10. A vertical line therefore indicates that the system on the y axis, after punctually matching the system on the x-axis, persists in the same state or changes slowly:

$R_{i,j+k} \equiv 1 \begin{cases} v-1 \\ k=0 \end{cases}$ defined as $\left(1-R_{i,j-1}\right)\left(1-R_{i,j+v}\right)\prod_{k=0}^{v-1} R_{i,j+k} \equiv 1$

with v the length of the vertical line. **Laminarity (LAM)** is the percentage of recurrence points which form vertical lines:

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}$$

where P(v) is the histogram of vertical lines of length v:

$$P(v) = \sum_{l,j=1}^{N} \left(1-R_{i,j}\right)\left(1-R_{i,j+l}\right)\prod_{k=0}^{v-1} R_{i,j+k}$$

The longer a vertical line, the longer we say one system gets "trapped" in one state (e.g., the listener keeps looking at picture A whilst the speaker looks at different pictures), which is quantified as **trapping time (TT)**, the average length of vertical lines:

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=vmin}^{N} P(v)}$$

The **length of the longest vertical line (Vmax)** indicates the longest uninterrupted time the system gets trapped in one state:

$$V_{max} = \max\left(\{v_l\}_{l=1}^{N_v}\right)$$

RQA and CRQA have been used successfully in the context of interpersonal coordination (see Fusaroli, Konvalinka, & Wallot, 2014, for review). For instance, the heart rates of participants involved in a collaborative task (building with LEGO bricks) showed greater recurrence stability (L) than virtual pairs (matching participants who did the task but not within the same group) (Fusaroli, Bjørndahl, Roepstorff, & Tylén, 2016). In a more dramatic context, the heart rates of participants (fire-walkers) during a fire-walking ritual showed higher recurrence with the heart rates of spectators who were their friends or relatives compared to the heart rates of nonrelated spectators, as measured by DET Lmax, ENTR and LAM (Dimitris, Konvalinka, Bulbulia, & Roepstorff, 2011). Using an interesting experimental setup, Ramenzoni, Riley, Shockley, and Baker (2012) combined PCA and CRQA analyses to examine both intra- and interpersonal motor coordination. The task required one person to hold a ring and the other person to hold their index finger in the middle without touching the ring. They found increase recurrence rate (RR) and stability (L) in the experimental task compared to the control task (where the participant held their finger in the middle of a fixed disc, still facing the other participant holding their disc). The authors also show that RQA measures are sensitive to the nature of the task performed and the constraints on joint and single performance. Similarly, Shockley et al. (2003) applied CRQA to postural sway analysis and found that participants who were conversing with one another whilst solving a puzzle showed greater coupling (RR) than participants who conversed with someone else. Finally, susaroli and Tylén (2015) were the first to apply RQA to the study of interpersonal conversational interaction. By comparing the stability L and structural organisation ENTR of recurrence between linguistic (lexicon) and prosodic (pitch, speech/pause patterns) time series, within individuals, between task-paired individuals, and between randomly-paired individuals, they were able to argue in favour of an interpersonal synergy model of conversation. Together, evidence suggest that RQA methods are a powerful method to detect task-sensitive fluctuations in coupling between systems across a range of physiological measures (eye-gaze, heart rate, postural sway, linguistic and prosodic features).

Visual inspection provides a fairly intuitive way to perceive the structure of recurrence between two systems. To pursue the example mentioned above, figure 4.29 (reproduced with permission from Richardson & Dale, 2005) illustrates three scenarios of eye-pattern coupling between a speaker and a listener. The denser the plot, and the longer the diagonals, the more the systems are coupled and for longer durations at a time. A “good” listener’s eye gaze pattern would be expected to match the speakers quite often and the coupling would be sustained in time, showing a checked plot such as the one illustrated in the left panel. A “bad” listener’s eye pattern state would only occasional coincide with the speaker’s state but this coupling would be expected to be sustained only for a short time, showing a more scattered pattern as illustrated in the middle panel, closer to a random pattern (right panel).

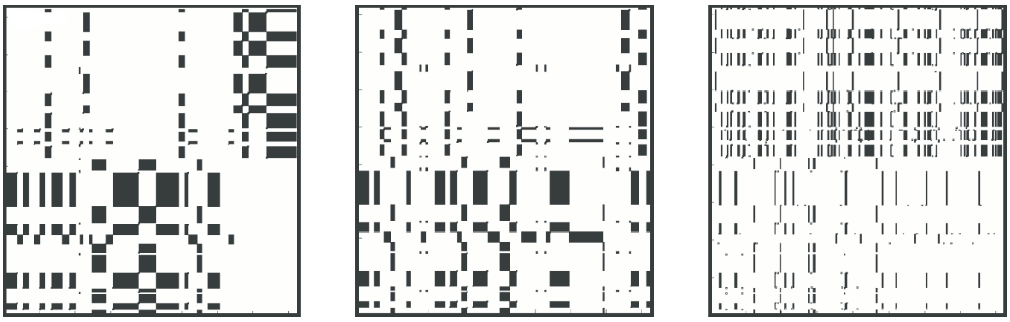

Figure 4.29 Three examples of cross-recurrence plots showing various degrees of coupling between a speaker and a listener's eye-movement data during an interaction.
Reproduced with permission from Richardson and Dale (2005). On the left, a "good listener's" profile shows a dense, highly organised pattern with wide blocks illustrating a high degree of coupling. In the middle, a "bad listener's" profile shows a scarcer, less clearly organised profile with narrower blocks. On the right is the profile of a "randomised listener" with his or her eye-movement data shuffled in a random order, showing no particular pattern.

In the context of this study, we propose to use RQA to quantify the recurrence patterns of acoustic and kinematic features of communication in ASD and TD participants. We specify our two alternative hypotheses as follow:

Hypothesis 1: due to increased variability in timing behaviour, systems relating to the acoustic and kinematic features will not stay in the same ε-recurrent state for similar amounts of time during the recordings for autistic individuals, resulting in a smaller amount of diagonal lines of length l for each l and therefore a diminished value of P(ε,l). This will result in reduced recurrence stability (L) and recurrence organisation and complexity (ENTR) of acoustic and kinematic features overall compared to TD individuals.

Hypothesis 2: because autism is clinically associated with repetitive behaviours, we expect a greater amount of repeated states $R_{i,j}$ overall, resulting in a higher recurrence rate (RR), and that systems relating to acoustic and kinematic will show repeated coupling (diagonals) with similar lengths, resulting in a greater P(ε,l) . This will result in greater stability (L) and lower entropy (ENTR). Finally, we also expect that systems will get trapped into specific states, increasing the total number of vertical lines P(v) and a greater trapping time (TT).

We will therefore focus our analysis on RR, L, ENTR and TT measures.

4.24 Analysis

In our set of data, we set out to quantify the recurrence within and cross-recurrence between acoustic and kinematic aspects of speech and gestures. Pauses, pitch and loudness were selected as acoustic features and holds and motion energy were selected as kinematic features. Pauses and holds were treated as categorical data (taking a value of 0 for pause/hold or 1 for speech/motion) whilst pitch, loudness and motion energy were continuous variables. Recurrence quantification analysis was performed in Matlab using the CRP toolbox (Marwan, N.: Cross Recurrence Plot Toolbox for MATLAB®, Ver. 5.22 (R32.1), http://tocsy.pik-potsdam.de/CRPtoolbox/, Marwan, Carmen Romano, Thiel, & Kurths, 2007).

RQA was performed on each single variable (pause, motion energy pause, pitch, loudness and motion energy) and CRQA was performed to quantify the recurrence between two variables. Categorical data were analysed without normalisation (keeping them as arrays of 0 and 1) whilst continuous data were normalised. So as not to generate spurious recurrence in the continuous data analyses, pauses were blanked out: the states in each system were attributed a different, out of range value, e.g., -2000 in system 1 and -3000 in system 2 (see Rothwell, 2018). However for analyses involving both acoustic and kinematic continuous features (ME x pitch and ME x loudness), the following procedure was adapted from Rothwell (2018): (a) to bring the data in a comparable space, we took the derivatives of the data time series (Matlab function *diff*), which means that we quantified the recurrence of changes between the two time series rather than their absolute values; (b) we normalised the data time series (mean = 0, standard deviation = 1); (c) we masked the pauses in each time series by attributing very large, distinct values (-2000 for ME, -3000 for pitch and loudness); (d) we blanked out the intersection of holds and pauses.

RQA involves 2 steps: (a) reconstructing of the phase space underlying the time series and (2) computing a cross recurrence plot. The n-dimensional phase space underlying two time series represents all the possible combined states of the two systems, which makes it possible to trace the shared trajectories of the two systems' behaviours. The phase space was reconstructed using the time delay method and recurrence plots were computed following Fusaroli and Tylén (2015), clarified and confirmed in personal communications with R. Fusaroli.

In the time delay method, we define:

$$\widehat{\vec{x_i}} = \sum_{j=1}^{m} u_i + (j-1)\tau\vec{e}_j$$

where u_i is a discrete time series and e_i are unit vectors. The parameters m (embedding dimension) and τ (delay) are embedding parameters which have to be chosen appropriately. The time delay τ was estimated as the "elbow" of the mutual average information function of the time series, and the embedding dimension m was estimated using the false nearest neighbours algorithm, where the parameter was increased until the recruitment of nearest neighbours did not significantly decrease. Finally, we set the threshold ε which is the radius of the neighbourhood in which recurrent states are identified. The choice of ε varied between the categorical and continuous data analyses. For the categorical data, ε was set to 0. This meant that for nominal data, only an exact match was counted as a recurrence. For continuous data, ε was chosen separately for each dataset to elicit a fixed recurrence rate of about 4% (Marwan et al., 2007). Parameters for each analysis are reported in table 4.12. Figure 4.29 shows examples of recurrence plots for one TD and one ASD recording.

	Categorical data			**Continuous data**					
	Pauses	Holds	Pauses x Holds	Pitch	Loudness	ME	Pitch x Loudness	ME x Pitch	ME x Loudness
τ	3	7	3	3	2	4	3	4	4
m	2	3	2	7	5	6	7	5	4
ε	0	0	0	1	1.3	0.6	1.6	6	4.2

Table 4.12 Parameter values for the Recurrence Quantification Analyses.
τ = time delay; m = embedding dimension; ε = threshold (or radius).

We performed uncorrected, planned independent sample t-tests on RQA/CRQA output measure, contrasting ASD versus TD group.

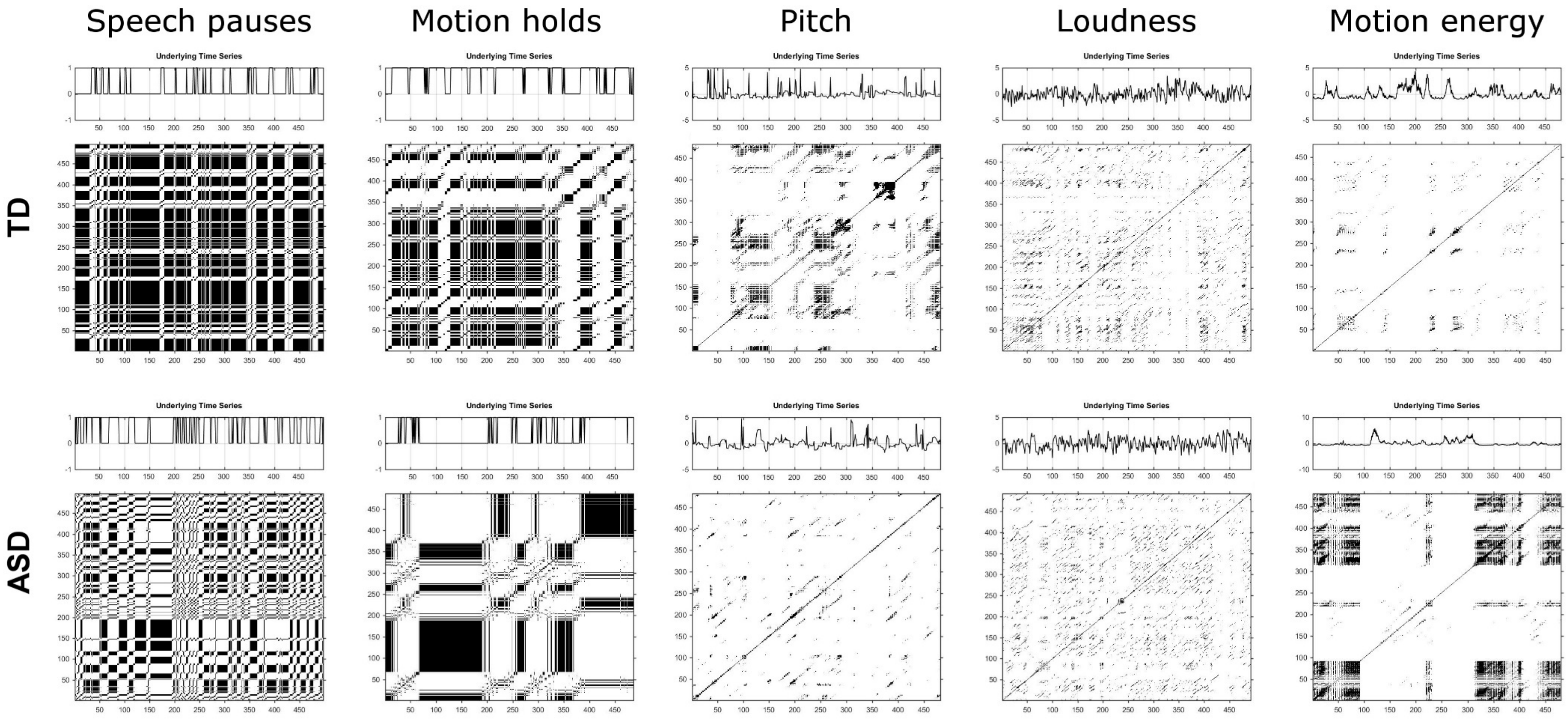

Figure 4.30 Comparison of recurrence plot for Pauses, Holds, Pitch, Loudness and Motion Energy within one TD (top) and one ASD (bottom) participants recordings.
For each plot, the top section illustrates the underlying time series, whilst the bottom section shows the recurrence plot. Plots were obtained using the CRP toolbox in Matlab.

4.25 Results: Recurrence quantification analysis

Participants were excluded of any particular analysis if their RQA measures were outside ±3 standard deviation of the mean of their group. One ASD participant was excluded from the pauses RQA, the motion energy x loudness CRQA and the motion energy x pitch CRQA. This was the participant with the lowest IQ scores across groups (FIQ = 81). Another ASD participant was excluded from the holds RQA and from the pauses x holds CRQA. This participant had produced very few gestures, and hardly any movement, which would have rendered their motion holds recurrence patterns artificially high. Two different ASD and one TD participants were excluded from the motion energy x pitch CRQA with no obvious differences in their characteristics apart from some noise in their recordings (some chair spinning, self-touching movement such as scratching). Another TD participant was excluded from the Pitch RQA. One ASD and one TD participants, different again, were excluded from the motion energy RQA. Both participants had produced very few gestures overall, which would explain why their motion energy recurrence values were atypical. Finally another one ASD and one TD participants were excluded from the pitch x loudness CRQA who did not show noticeable unique characteristics in their profiles or recordings.

Shapiro-Wilk tests revealed multiple violations of the normality assumption for the CRQA data in both groups. In addition, Levene's test of homogeneity indicated that variance was not equal between groups for all scores. Group differences for RQA and CRQA outputs were therefore explored with Mann-Whitney U tests, corrected for multiple comparison using the Benjamini-Hochberg procedure. The results are reported in Table 4.13.

		ASD median	TD median	z	p	r
Pauses	**RR**	0.362	0.408	**2.261**	**0.023***	**0.400**
$n_{ASD} = 16$	**L**	5.835	6.737	1.658	0.102	0.293
$n_{TD} = 16$	**ENTR**	2.124	2.400	**1.885**	**0.061***	**0.333**
	TT	8.285	9.717	**1.922**	**0.056***	**0.340**
Holds	**RR**	0.315	0.341	0.339	0.752	0.060
$n_{ASD} = 16$	**L**	7.246	7.819	1.319	0.196	0.233
$n_{TD} = 16$	**ENTR**	2.312	2.464	1.696	0.094	0.300
	TT	12.345	13.353	1.281	0.21	0.226
Pitch	**RR**	0.036	0.026	-0.321	0.766	-0.057
$n_{ASD} = 17$	**L**	2.637	2.581	-0.85	0.411	-0.150
$n_{TD} = 15$	**ENTR**	0.763	0.693	-0.963	0.35	-0.170
	TT	3.158	3.180	-0.359	0.737	-0.063
Loudness	**RR**	0.031	0.036	**2.413**	**0.015***	**0.427**
$n_{ASD} = 17$	**L**	3.249	3.397	**2.161**	**0.031***	**0.382**
$n_{TD} = 16$	**ENTR**	1.533	1.621	**2.053**	**0.041***	**0.363**
	TT	2.297	2.478	**1.837**	**0.068***	**0.325**
Motion energy	**RR**	0.055	0.036	**-2.135**	**0.033***	**-0.377**
$n_{ASD} = 16$	**L**	4.643	3.503	-1.739	0.086	-0.307
$n_{TD} = 15$	**ENTR**	1.698	1.453	-1.225	0.232	-0.217
	TT	6.096	4.148	**-1.897**	**0.06***	**-0.335**
Pauses x Holds	**RR**	0.298	0.349	0.98	0.341	0.173
$n_{ASD} = 16$	**L**	5.624	6.632	1.394	0.171	0.246
$n_{TD} = 16$	**ENTR**	2.137	2.362	1.658	0.102	0.293
	TT	12.699	16.444	**2.299**	**0.021***	**0.406**
Pitch	**RR**	0.031	0.036	1.542	0.129	0.273
x Loudness	**L**	2.189	2.328	1.186	0.247	0.210
$n_{ASD} = 16$	**ENTR**	0.352	0.510	0.87	0.401	0.154
$n_{TD} = 15$	**TT**	2.226	2.412	1.265	0.216	0.224
ME x Pitch	**RR**	0.044	0.031	-0.655	0.533	-0.116
$n_{ASD} = 14$	**L**	3.261	3.300	0.698	0.505	0.123
$n_{TD} = 15$	**ENTR**	1.323	1.357	0.436	0.683	0.077
	TT	3.833	4.453	1.615	0.112	0.285
ME x Loudness	**RR**	0.057	0.039	-1.131	0.27	-0.200
$n_{ASD} = 16$	**L**	3.024	3.104	-0.188	0.867	-0.033
$n_{TD} = 16$	**ENTR**	1.216	1.258	-0.339	0.752	-0.060
	TT	3.467	3.947	0.829	0.423	0.147

Table 4.13 Summary statistics of the RQA and CRQA of acoustic and kinematic features.
RR = Recurrence Rate; L = Mean diagonal length; ENTR = Entropy; TT = Trapping Time. z-scores are reported from Mann-Whitney tests. p-values identified as significant after the Benjamini-Hochberg procedure are marked with an asterisk. Effect size r are reported.

Overall, across features, TD participants tended to show a greater degree of recurrence in their acoustic features compared to ASD participants. Starting with single features, speech pauses showed a significantly higher amount of recurrence points (RR), the recurrence structure was more “organised” or complex (ENTR) and there was a greater tendency to get trapped into a certain state (TT) in the TD compared to the ASD group. This is in line with our previous finding that TD participants produce more long speech pauses and less short pauses than their ASD counterparts. Similarly, for speech loudness, degree of recurrence (RR), stability (L), complexity/organisation (ENTR) and the propensity to repeat (TT) were greater in TD than ASD participants. Although this results do not directly relate to the absolute difference in mean loudness between ASD and TD group, this is a second pointer indicating that speech loudness presented distinctive patterns in the ASD and TD groups. There were no differences in recurrence patterns of **speech pitch** or **holds** between TD and ASD groups. Interestingly, RQA measures for **motion energy** showed a greater degree of recurrence (RR) overall and a greater tendency to repeat (TT) in the ASD compared to the TD group, indicating that the same levels of ME are revisited more often and for longer by ASD participants.

Differences followed a similar trend in CRQA results, but only significantly so for the trapping time for **pause x holds**, showing that TD participants showed greater tendency to repeat (TT) in the **pauses x holds** CRQA compared to ASD participants. There were no differences between groups for any of the variables in the **pitch x loudness**, **motion energy x pitch** or **motion energy x loudness CRQA**.

4.26 Results: correlations with diagnosis and communication measures

In order to test the relationship between recurrence patterns and both perceived communication skills and clinical traits, we computed Pearson or Spearman correlation coefficients as appropriate between the four RQA measures (RR, L, ENTR, TT) on the one hand and ADOS scores, AQ score and the total visual gain in quality of communication on the other hand. The Benjamini-Hochberg procedure was applied to control for multiple tests.

4.26.1 Clinical severity scores: ADOS and AQ

In the **pauses recurrence data**, the ADOS communication score appeared to be positively associated with L ($r(16) = .547$, $p = .028$) and TT ($r(16) = .530$, $p = .035$), as illustrated on Figure 4.31, but these associations did not survive the family-wise error rate correction.

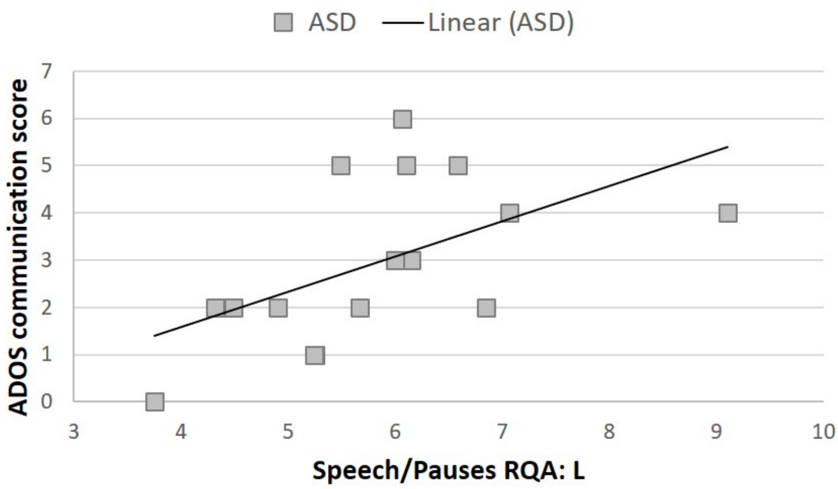

Figure 4.31 Relationship between the stability L of Speech/Pauses recurrence patterns and ADOS communication for ASD participants.

In the **holds recurrence data**, in contrast, the ADOS communication score correlated negatively with L, ($r_s(16) = -.566$, $p = .022$), ENTR (r_s $(16) = -.598$, $p = .014$) and TT (r_s $(16) = -.573$, $p = .02$), suggesting that a greater degree of recurrence in gesture holds was associated with lesser difficulties in the communication domain (see figure 4.32). This is consistent with the view that predictable patterns of holds are a means to improve communication.

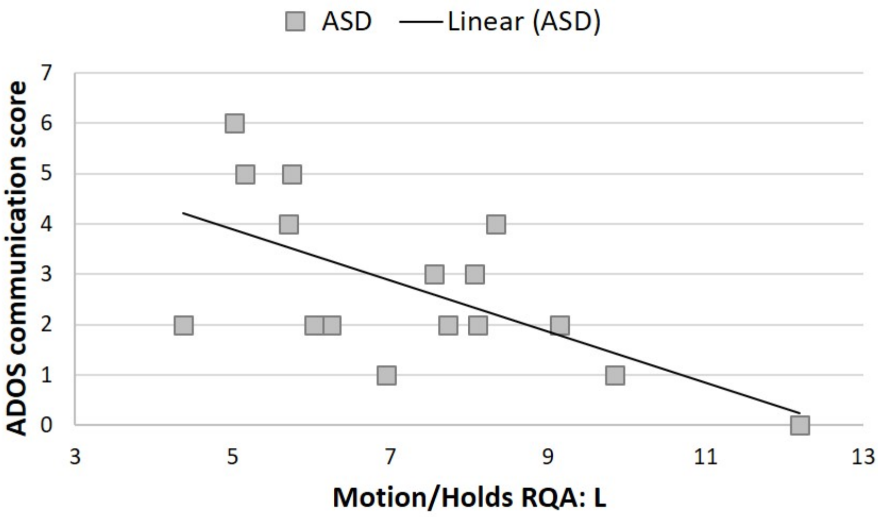

Figure 4.32 Relationship between the stability L of Motion/Holds recurrence patterns and ADOS communication for ASD participants.

In the cross-recurrence between **pauses and holds data**, the ADOS RRB score was negatively associated with RR (r(16) = -.443, p = .086), L (r(16) = -.633, p = .008), ENTR (r(16) = -.567, p = .022) and TT (r(16) = -.522, p = .038), as illustrated on Figure 4.33, but these correlations did not survive family-wise error rate correction.

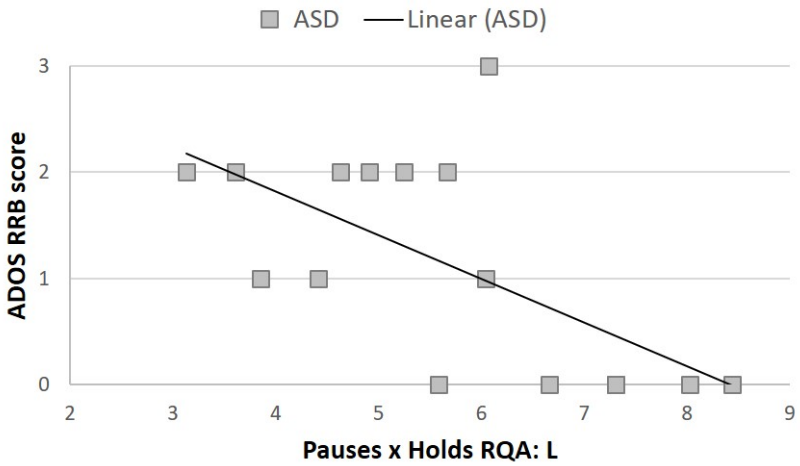

Figure 4.33 Relationship between the stability L of Pauses/Holds cross-recurrence patterns and ADOS RRB score for ASD participants.

In the **pitch recurrence data**, we found no significant associations between any of the RQA variables and severity scores.

In the cross-recurrence between **motion energy and pitch data**, we also found no significant associations between any of the RQA variables and severity.

In the cross-recurrence between **motion energy and loudness data**, the ADOS RRB score appeared to correlate negatively with ENTR ($r(16) = -.574$, $p = .020$) and TT ($r(16) = -.559$, $p = .024$), suggesting that complexity and repetition of recurrence states were associated with less restricted and repetitive behaviours (see figure 4.33), but none of these correlations did not survive correction for multiple tests.

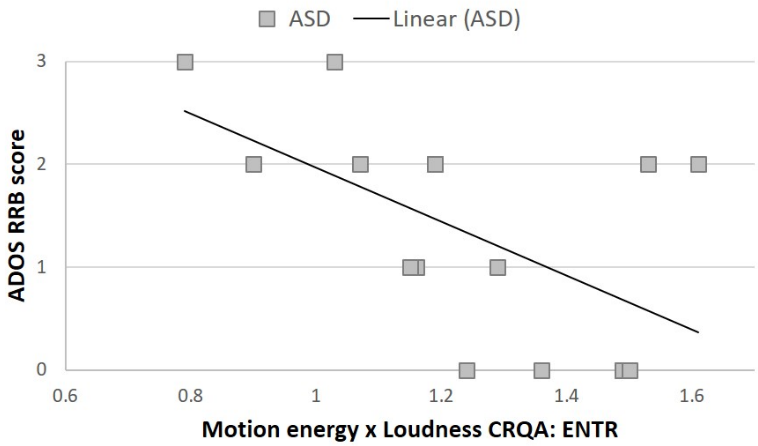

Figure 4.34 Relationship between the entropy ENTR of motion energy x loudness cross-recurrence patterns and ADOS RRB score for ASD participants.

No correlations were found between ADOS and AQ scores and the recurrence patterns of motion energy, loudness and pitch x loudness.

4.26.2 Correlations with quality of communication ratings

None of the RQA outcome variables were significantly associated with quality of communication ratings, suggesting that the structure of acoustic and kinematic features does not directly translate to the subjective perception of a person's ability to communicate.

4.27 Summary and preliminary discussion

Recurrence quantification analysis provided us with two lines of evidence into the differences between ASD and TD communication. The first was gained from a direct comparison of the degree of recurrence between groups which showed that overall, acoustic features were generally characterised by a greater degree of recurrence in TD participants' compared to ASD participants' recordings. The second line of evidence was provided by correlational data between RQA indices and measures of clinical severity (ADOS and AQ scores) and quality of communication ratings.

First, the patterns of **speech/pauses** were more stable in the TD compared to the ASD group, which means that TD participants were either speaking or not speaking for longer consecutive intervals. In comparison, verbal productions showed a more interrupted, “stop-start” structure in the ASD group. This complements our finding that ASD participants produce more brief pauses (< 200ms) and less long pauses (> 500ms), whilst TD participants produced less brief and more long pauses (chapter 4, part C, section 2). Interestingly, in a previous section, a greater rate of holds in the ASD group was associated with greater difficulties as measured by the ADOS communication and total scores (chapter 4, part C, section 3). We found a similar association here where lower degree and organisation of recurrence in holds (L, ENTR and TT) were all associated with greater difficulties in the communication domain and lower quality of communication ratings, suggesting that not only a greater amount of holds but also diminished organisation of pauses are related to poorer communication skills in autism. In addition, the cross-recurrence patterns of pause and holds were found to have a greater tendency to stay in the same states (repeat the same values) in the TD compared to the ASD group. This indicates that pauses and holds were not only more consistent individually, but they also resembled and coordinated with each other more in the TD group. The predicted consequence of this would be that in the autistic group, the pattern of motion/holds was less informative to help decode the speech signals as it did not predict the speech/pause pattern as closely as in the TD group. We postulate that in autistic individuals, behaviours such as self-stimulatory gestures (Tantam, Holmes, and Cordess, 1993) interfere with efficient use of gesture and their coordination with speech.

These results contributes to a growing amount of evidence that speech/pause patterns are atypical in autism, although previous literature more often points to longer pauses in autistic verbal productions (Mitchell, 2015; Morett et al., 2016). In line with the discussion in chapter 4, part C, we propose that the duration of pauses in autistic individuals are less task-dependent than in TD individuals, which leads them to appear shorter or longer depending on the context. Our results provide additional insight that not only the raw duration but also the segmentation of speech/pause behaviour and its coordination to motion/hold patterns may be atypical in autism, with consequences on communication skills.

Moving onto **acoustic features**, recurrence quantification analysis highlighted no differences between the degree of recurrence in pitch between the groups. In contrast, loudness was found to show a greater degree and stability of recurrence, greater complexity and repetitive patterns in the TD compared to the ASD group. Although the derivative data is somewhat difficult to interpret in linguistic terms, it suggests that typical participants modulated the loudness of their speech more often and for longer intervals of time, repeated similar sequences of loudness modulation and showed more predictable patterns of loudness change than ASD participants. This result is the second instance showing a strong distinction between loudness characteristics in the ASD compared to the TD group (the first one being overall louder speech in the ASD group, chapter 4, part C, section 1), which has to our knowledge not been reported before in the literature (see Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017, for review). Contrary to the absolute value of loudness, the RQA results on the derivative of loudness are unlikely to reflect differential orientation to the microphone or size of participants. The novelty of this effect in our data combined with the novelty of the task in which the data were acquired possibly indicates that atypical speech loudness is task- or context-dependent. Finally the cross-recurrence patterns between pitch and loudness also showed greater degree of recurrence overall in the TD compared to the ASD group, indicating that in typical individuals, pitch and loudness visited more similar spaces than in autistic individuals. In other words, if we imagine that pitch and loudness are on a common scale, we would find that TD individuals modulate pitch and loudness in coordination more often than ASD individuals. The likely consequence of this alignment is that information in one time-series can "translate" more easily and be a better predictor for the other time-series. Again, this shows a more integrated coordination of speech features in the TD group. Interestingly, a greater degree of recurrence between pitch and loudness was associated with increased visual gain, which suggests that when pitch and loudness share more common information, gestures make a greater difference to quality of communication. Future studies should explore the conditions under which differences in acoustic features do or do not appear, which will hopefully help to disentangle the somewhat mixed results in the literature on pitch in autism, and our current results for loudness.

In contrast with the results so far, recurrence quantification of **motion energy** showed a greater degree of recurrence and greater tendency to repeat the same level of motion in the ASD compared to the TD group. This is a likely consequence of the greater amount of holds in the ASD group (chapter 4, part C, section 3): if autistic individuals produce more holds, this would be expected to "boost" the amount of recurrence points (null values aligning with null values) and the amount of repetition of the same value (null motion energy). Supporting this interpretation, we found no difference in the stability (L) or organisation (ENTR) of the recurrence patterns of motion energy between groups. In addition, the cross recurrence patterns between motion energy and pitch showed a greater tendency to repeat (get trapped) in the same values in the TD group compared to the ASD group however, whilst no group differences were found in the cross-correlations between motion energy and loudness. These results are a little more difficult to interpret as they might combine increased coordination of speech and gesture patterns in the TD group and increased amount of holds in the ASD group.

Taken as a whole, recurrence data therefore favours hypothesis 1 which predicts that because ASD individuals demonstrate greater variability in the timing of their behaviour, they should show reduced stability and organisation of recurrence over acoustic and kinematic features in communication. Autistic participants showed lower stability and complexity than their typical counterparts within acoustic features (speech/pauses behaviour, loudness and pauses x holds). As a result, autistic verbal communication is expected to be less predictable and therefore more difficult to follow and adapt to for their conversational partner. This was confirmed by correlational data which found overall that lower degrees of recurrences were associated with greater difficulties in the communication domain.

As an additional insight, trends in the correlational data suggest that restricted and repetitive behaviours may be an important disruptive factor in speech and gesture coordination in autism. Whilst several authors have proposed that repetitive patterns of behaviour might be a strategy to parse time and support temporal processing in autism (Allman et al., 2011, 2014; Boucher, 2001; Lewis & Miall, 2003), in the context of communication our results suggest that the rigidity of repetitive behaviours are not facilitating the temporal coordination of the different streams of communication but on the contrary impeding it. There may be a question of scale here: repetitive behaviours such as finger flicking and self-stimulating are likely to produce regular units of time which are longer than the few dozens to hundreds of milliseconds required to efficiently coordinate speech and gesture. Parsing time using repetitive behaviours might therefore occur for larger time scales. These correlations however did not survive corrections for multiple tests and should be taken with great caution. Testing the relationship between restricted and repetitive behaviours and the organisation of gestural behaviour could be the object of future, dedicated research.

All in all, crucially, recurrence results provided a successful quantification of qualitative aspects of communication in autism which are detected by clinical assessments such as the ADOS.

Part F – Predicting clinical features of ASD

4.28 Rationale

The study so far has focused on characterising temporal aspects of speech and gesture communication in ASD with the aim to identify quantifiable variables that capture some of the qualitative differences reported by those who interact with autistic individuals. In particular, diagnostic tools such as the ADOS use subjective ratings such as the 'quality' of gestures or the overall integration of vocalisation, gesture and eye movements to inform a diagnosis of ASD. Results in previous sections found some success in establishing relationships between quantifiable acoustic and kinematic aspects of the data and qualitative scores, including ratings of quality of communication and ADOS scores. Very broadly, diminished rate, increased duration and increased repetition of gestures were generally associated with greater ADOS scores and diminished quality of communication ratings. In addition, a higher degree of temporal coupling between speech and gesture was found to relate to higher ADOS scores. Finally, higher recurrence in voice patterns (loudness in particular) seemed to be associated with higher ADOS scores and lower quality of communication ratings, whereas higher recurrence of gesture patterns tended to be associated with lower ADOS scores and higher quality of communication ratings.

The emerging question, however, is whether these associations have any predictive value, in other words whether measuring quantifiable temporal aspects of speech and gesture could accurately predict a diagnosis of ASD, severity scores, or qualitative ratings. ASD is a multifaceted developmental disorder which varies greatly from individual to individual and there is currently no biomarker to confirm a diagnosis. As a result a diagnosis of ASD is based on subjective ratings and often resource-intensive assessments. The hope of identifying quantifiable variables which reliably predict a diagnosis could have huge clinical implications, if they apply to autistic individuals across the spectrum. To start answering this question, we turn to a methodological approach which has greatly gained popularity in many domains over the past two decades or so: machine learning. Machine learning involves building algorithms that are capable of learning from examples. It is used to solve two types of tasks: classification (determining which group a datapoint belongs to) and score prediction (predicting the value of a variable under given conditions). Machine learning has proven valuable for tasks such as weather forecasting (Xingjian, Zhourong, Hao, & Yeung, 2015), financial advising (Yu, Miche, Séverin, & Lendasse, 2014), medical imaging (Erickson, Korfiatis, Akkus, & Kline, 2017), spam filtering (Guzella & Caminhas, 2009) and many more. More to the point, machine learning has recently been used successfully in 14 different studies to predict an individual's diagnostic group (ASD or TD) based on voice features, with an accuracy of well over 70%, and up to 90%, which is well above the accuracy rate claimed by univariate studies of around 61-64% (see Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017 for a meta-analysis).

Whilst traditional univariate analyses answer the question of whether there is a statistically significant difference between two populations, multivariate machine learning (ML) algorithms reverse the problem and answer the question of whether there is enough information in the data to accurately separate two (or more) populations. Specifically, ML can deal with large numbers of features and multiple statistical models. Multivariate ML studies typically follow a procedure of (1) feature extraction, (2) feature selection, (3) classification (e.g., ASD vs TD) or score prediction (e.g., total ADOS score) and (4) validation. **Feature extraction** can produce a large number of variables, usually because the systems investigated are complex and because variations are derived from the variables (e.g., derivative, log) to augment the chances to capture the data accurately. **Feature selection** ensures that only a limited number of features identified as the most informative are selected. **Classification** or **score prediction** processes use the selected features to construct a statistical model which either distinguishes between populations of interest or predicts a score accurately, respectively. Finally, the aim of ML is not so much to explain the data at hand, but rather to construct a model that generalises to new data. To that end, the process undergoes **validation** (or cross-validation), which usually involves dividing the data into two datasets: one is the training dataset to which the model is fitted (the "learning" part), the other is the test dataset on which the accuracy of the model and its predictive power are assessed (Rodriguez, Perez, & Lozano, 2010).

Previous studies investigating whether voice patterns could predict a diagnosis of ASD used a variety of features: most use a measure of pitch and its variability, some used measures of the quality of voice such as shimmer and jitter, spectral and cepstral (non linearly transformed) features of speech, energy features, recurrence measures, and few studies used measures of loudness and duration (Asgari, Bayestehtashk, & Shafran, 2013; Bone, Black, & Lee, 2012; Bonneh, Levanon, Dean-Pardo, Lossos, & Adini, 2011; Fusaroli, Bang, & Weed, 2013; Kakihara, Takiguchi, Ariki, & Nakai, 2015; Kiss, van Santen, Prud'hommeaux, & Black, 2012; Marchi et al., 2015; Oller et al., 2010; Santos et al., 2013). Crucially, the exact measures or methods used to measure particular features differed between studies, making it difficult to identify a particularly informative set of features. All studies constructed a classifier algorithm which successfully identified ASD participants with an accuracy of 71% and above, across a wide range of ages (18 months to 62 years old) and languages (British and American English, Danish, Swedish, Hebrew and Japanese). In addition, Fusaroli et al. (2014) were able to predict AQ scores with an accuracy of 80% and Bone et al. (2014) predicted ADOS severity scores with 80% accuracy.

4.29 Aims and predictions

In this section, we test whether a combination of acoustic and kinematic features of speech and gesture quantified in the previous sections (descriptive, temporal lock, cross-correlation and recurrence) contain enough information to (1) classify participants by diagnosis group; (2) predict severity of symptoms as measured by ADOS and AQ scores; (3) predict quality of communication ratings. In line with the literature, we predicted that voice patterns alone would allow us to classify participants by diagnostic group with high accuracy. We hypothesised that the addition of kinematic features would improve the accuracy of the classification. In addition, we predicted that acoustic features alone would predict ADOS and AQ scores with high accuracy, and hypothesised that kinematic features would again improve the accuracy of prediction. Finally, we tested whether acoustic and kinematic features of speech and gesture could predict the subjective ratings of quality of communication generated as a phenotypical marker of autistic communication.

4.30 Methods

The analysis was implemented in Matlab using scripts provided by Riccardo Fusaroli and adapted with permission. The procedure closely followed the one employed in (Fusaroli et al., 2013; Fusaroli, Lambrechts, Yarrow, Maras, & Gaigg, 2016).

One ASD participant was excluded from the sample because he was excluded from most of the other analyses and as a result was missing data for most features. This was also the participant with the lowest IQ across groups. The final analysis included the remaining 16 ASD and 16 TD participants.

Features containing missing data were excluded from the analysis, which included all features from the temporal lock analysis and most features generated from gesture annotation. In addition, features containing too many zeros or repeating the same value (e.g., minimum value of motion energy) were also excluded. The features included in the final analysis are summarised in table 4.14.

Descriptive	Pitch	mean, standard deviation, min, max
	Loudness	mean, standard deviation, min, max
	Motion energy	mean, standard deviation, max
	Pauses Holds	percentage, frequency, mean duration, max duration, standard deviation duration, percentage short duration, percentage medium duration, percentage long duration
Gesture annotation	Iconic Metaphoric Deictic Beat	frequency
Cross-correlation data	Pitch-Loudness	number of peaks, main peak correlation coefficient, main peak width, main peak prominence, skewness
	ME-Pitch ME-Loudness Pauses-Holds	number of peaks, skewness
Recurrence quantification	Pauses Holds Pitch Loudness Motion energy Pitch-Loudness ME-Pitch ME-Loudness Pauses-Holds	RR, DET, L, Lmax, ENTR, LAM, TT, Vmax

Table 4.14 Initial features included to construct both classification and score prediction algorithms.

Because the number of variables (our features) was larger than the number of examples (individual participants data), we used ElasticNet (Zou & Hastie, 2005) for feature selection. ElasticNet is a regularisation method with penalises some of the features and eventually selects only the group of features with the most predictive power for model construction. This sped up the computational time, but also reduced the risk of overfitting the data. Diagnosis was predicted using a 5-fold cross-validation and the accuracy was balanced using Variational Bayesian mixed effects inference (Brodersen et al., 2013; Brodersen, Ong, Stephan, & Buhmann, 2010). This method provides a valid measure of accuracy appropriate for a binary classifier, and allows to derive confidence intervals (CI) based on the cross-validation, taking both within- and between-subject variability into account. Analyses were iterated 100 times to test for stability of results and 95% CI were calculated. This means that on each of 100 iterations, the whole dataset was partitioned into five shares, one of which was selected as the test dataset and the others as the training dataset, providing a robust validation of the result. Several models were compared for best fit including logistic regression, discriminant functions (linear, diaglinear, quadratic, diagquadratic and mahalanobis), naïve Bayesian classifier and support vector machine models. Analyses were run using (1) acoustic features only and (2) acoustic and kinematic features to test whether results were similar to previous studies using acoustic patterns of voice on the one hand, and test whether the addition of kinematic patterns would improve the accuracy of the model. ADOS total score and individual factors scores (communication, reciprocal social interaction, imagination and creativity, and restricted and repetitive behaviours) were predicted using a 5-fold cross-validation multiple linear regression. Again analyses were iterated 100 times to test for stability of results and 95% CI were calculated. We report balanced accuracy (the proportion of participants classified correctly), sensitivity (the proportion of ASD participants classified correctly) and specificity (the proportion of TD participants classified correctly).

4.31 Results

4.31.1 Diagnostic group classification

4.31.1.1 Acoustic features only

We found that acoustic features on their own accurately separated ASD from TD participants. The best fit was obtained with a linear discriminant function model which yielded a balanced accuracy of 81.63% (CI: 68.62 – 91.85%, $p = 0.00423$) with a sensitivity of 83.48% and specificity of 87.54%. The selected features consisted exclusively of loudness features: mean, standard deviation, and the maximum length of diagonal in recurrence plots of loudness.

Although the full model cannot be easily plotted, a distribution of the two strongest predictors (mean and standard deviation of speech loudness, see figure 4.36) shows a clear demarcation between groups.

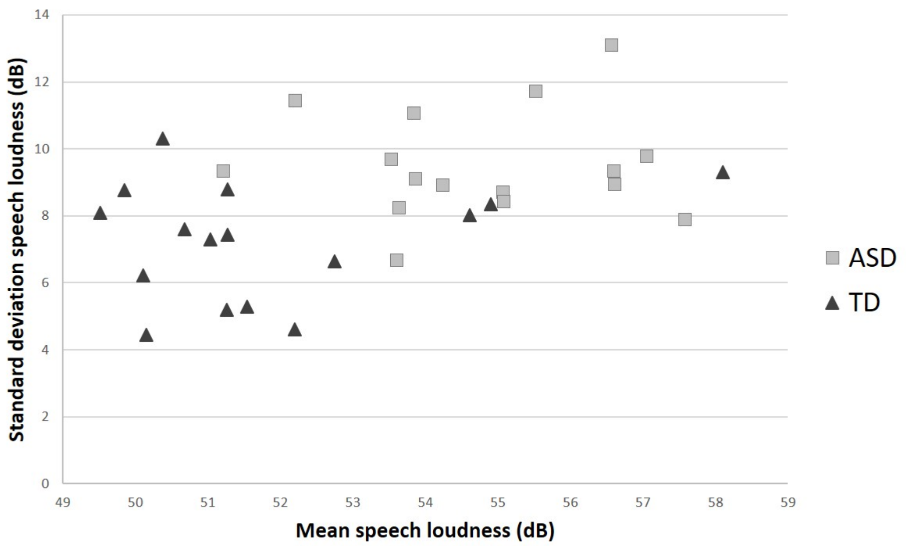

Figure 4.35 Distribution of the mean (on the x-axis) and standard deviation (on the y-axis) of speech loudness per group.

4.31.1.2 Combined acoustic and kinematic features

We found that acoustic features on their own accurately separated ASD from TD participants. The best fit was obtained with a diaglinear discriminant function model which yielded a balanced accuracy of 82.37% (CI: 69.73 – 92.18%, $p = 6.9 \times 10^{-5}$) with a sensitivity of 85.78% and specificity of 86.63%. The selected features consisted descriptive features of loudness (mean, standard deviation, maximum), the maximum length of diagonal in recurrence plots of loudness, and the maximum length of diagonal in recurrence plots of motion energy, and the skewness of the curves of cross-correlation between motion energy and pitch, and between pitch and loudness.

Again, whilst the full model cannot be easily plotted, a distribution of the two strongest predictors (motion energy Lmax and mean speech loudness, see figure 4.37) shows a clear demarcation between groups.

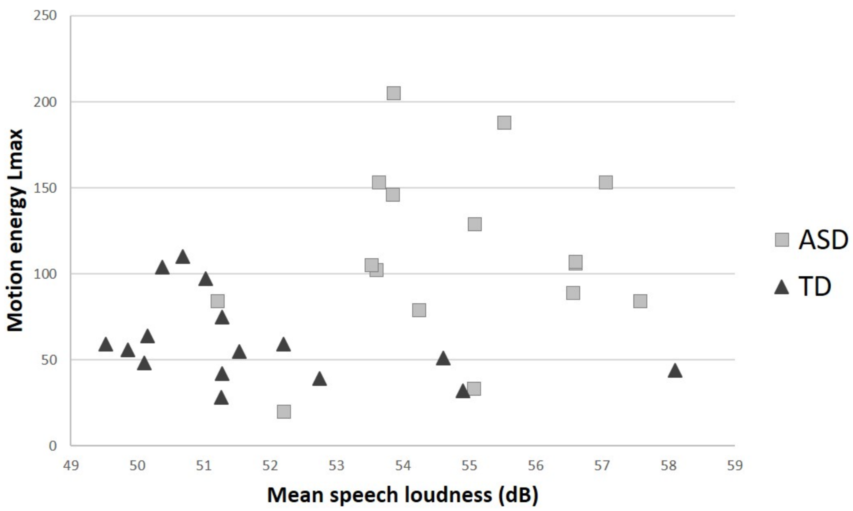

Figure 4.36 Distribution of the mean speech loudness (on the x-axis) and mean motion energy Lmax (on the y-axis) per group.

4.31.2 Scores prediction

We found that neither acoustic features alone nor a combination of acoustic and kinematic features could predict any of the ADOS scores, AQ or quality of communication scores accurately.

4.32 Summary and provisional discussion

In this section, we assessed the power of acoustic and kinematic features of speech to classify participants according to diagnostic group on the one hand, and predict symptom severity, autistic traits and quality of communication scores as measured by the ADOS, AQ and our subjective ratings of quality of communication respectively.

In line with previous studies (see Fusaroli et al., 2017 for review), we found that acoustic features of speech predicted diagnostic group with high accuracy, sensitivity and specificity (all over 80%). Overwhelmingly, all predictor variables in the acoustic-only condition were derived from speech loudness, specifically the mean and standard deviation of loudness and the maximum length of diagonals (Lmax) in the loudness recurrence plots (showing longer sequences of loudness recurrence) were selected in the final model. In our sample, autistic participants were characterised by speech that was louder overall, showed greater variability and was more repetitively louder. Three TD participants stood out as having louder speech than their group peers. Inspection of individuals characteristics did not however reveal any obvious difference in terms of gender, age or IQ.

Contrary to prediction, the addition of kinematic features did not improve the overall fit significantly but instead produced a model which yielded similar accuracy, sensitivity and specificity to the acoustic-feature-only model (once more all over 80%). This is likely because of the high amount of recurrence patterns between speech and gesture, and supports the idea that communication is a highly redundant system in which the same information is available through various channels and repeated, at least in terms of prosody and speech/gesture alignment. The strongest predictor identified was the maximum length of diagonals (a measure of recurrence stability) in the motion energy recurrence plots, with longer diagonals characterising ASD participants. As mentioned in the discussion of the previous section (chapter 4, section E, section 6) however, the greater degree of recurrence in motion energy in ASD participants was likely driven by a greater percentage of holds in the ASD group (chapter 4, part C, section 3) rather than a greater amount of recurrence within gestures so a better interpretation is probably that gesture/hold behaviour pattern was a predictor of diagnostic group. Speech loudness was again well represented in the selected predictive factors (five out of seven features involved loudness), including the same features as in the acoustic-only model, plus the maximum loudness level during the recording and the skewness of the cross-correlation profile between pitch and loudness. Finally, the skewness of the cross-correlation profile between motion energy and pitch was also selected in the final model.

Our result that speech loudness is a main predictor of diagnosis is very surprising in light of the literature, which does not flag speech loudness a potential marker of autism. In fact, a majority of studies report no difference in loudness between groups (Diehl & Paul, 2013; Filipe, Frota, Castro, & Vicente, 2014; Grossman, Bemis, Plesa Skwerer, & Tager-Flusberg, 2010; Hubbard & Trauner, 2007; Quigley, McNally, & Lawson, 2016) and one study even reported that loudness was overall lower in the ASD group (Scharfstein, Beidel, Sims, & Rendon Finnell, 2011). First of all, it is important to repeat that the original study was not set up to measure prosodic aspects of speech, therefore the precise distance between the participant and the microphone was not measured and controlled for. For this reason, it is not possible to exclude the possibility that the difference between groups was an artefact. However, as noted previously, all recordings were made with the same device and in the same room, and the set-up was extremely similar between participants. Visual examination of the recordings did not reveal any apparent differences in the propensity of autistic participants to direct their speech towards the camera more compared to non autistic participants. In addition, examination of the data (see figure 4.36 for example) indicates a clear demarcation between the groups, showing that the trend is not affected by outliers. Finally, each model constructed using the two machine learning algorithm was cross-validated so that on each of 100 iterations, the whole dataset

was partitioned into five shares, one of which was selected as the test dataset and the others as the training dataset, providing a robust validation of the result. Another limitation is that, because microphones are commonly equipped with a gain-control algorithm, it is possible that greater variability in loudness (rather than greater loudness *per se*) was responsible for the difference observed between the groups. Unfortunately information about the built-in microphone was not available to confirm this hypothesis. If this is the case, however, it remains that autistic participants in this situation showed atypical speech loudness dynamics, but it sheds some uncertainty as to the nature of this difference.

A number of methodological differences could account for the discrepancy between ours and previous results. Our sample consisted of adult participants aged 25-62, whereas previous studies investigated prosody in toddlers, children, adolescents and young adults (the oldest participant age was 21, in Hubbard & Trauner, 2007). In addition, most studies looked at the production of single words or short sentences, often in a constrained environment. For instance, in Filipe et al. (2014) and Hubbard and Trauner (2007), participants were repeating single words, and in Diehl and Paul (2013) and Grossman et al. (2010) children and adolescents generated short answers in the context of very specific tasks (e.g., complete an unfinished sentence such as "Kate calls Tom on his cell phone. When Tom doesn't answer, Kate wishes he would [pick up]," paired with an illustration of Tom picking up the phone, Grossman et al., 2010). Whilst these tasks and settings allowed the authors to control the conditions of the recording well, the prosody displayed might have differed from a task in which the speaker talks more spontaneously, naturally and for a longer time. In addition, in our task, participants spoke uninterrupted for a few minutes, which means that the influence of their conversational partner's feed-back (in our case the interviewer) was limited. Quigley et al. (2016) looked at spontaneous interaction but they consist of mother-infant vocalisation in a sample of infants and toddlers at risk or not at risk of ASD, which is very different to both our sample and task. Finally, Scharfstein et al. (2011) used acoustic samples recorded during a structured role-play session with children aged 7-13. The authors report as a limitation of their data that the script used to guide the interaction with the children was quite rigid and sometimes lead to ineffective responses. For instance in the role-play scenario, the first social prompt was always followed by a second scripted line, regardless of the child's response to the first prompt, which sometimes lead to awkward responses. In comparison with previous studies, our data provides naturalistic, spontaneous recordings of fluent adults. It also provides a long sample during which (although speech is addressed to a listener), the speaker is uninterrupted, which we reason is more likely to bring out prosodic idiosyncrasies than a fast turn-taking conversation. Naturally the results need to be replicated, preferably with a larger sample, to gain in strength but we argue that they offer a valuable insight into prosodic aspects of communication in autistic adults. Reflecting on its potential use as a marker of autism, unfortunately loudness is not an easy feature to measure in a controlled manner. It is highly dependent on the conditions of the recordings such as the microphone used and the room around. Its appeal however is that as a prosodic feature, loudness is a modulator of communication: it does not change the meaning of the message, but affects how the message is received.

Beyond speech loudness, the features selected in the combined acoustic and kinematic features model included (with the highest prediction power) a measure of motion energy recurrence (Lmax). It is interesting to learn that even a fairly crude measure of gesture (motion energy) is sensitive to the diagnosis of ASD. Teitelbaum, Teitelbaum, Nye, Fryman and Maurer (1998), coding movement in infants who later developed a diagnosis of ASD, suggested that early differences might be a useful predictor for autism in early age, before language develops. Although there are reports of atypical movement in autism (e.g., Cook, Blakemore, & Press, 2013), to our knowledge Teitelbaum's idea that movement may predict a diagnosis of ASD has not been directly examined. Our data suggest that recurrence aspects of motion energy specifically might be a good candidate. In addition, two features quantifying the temporal coordination of speech and gesture signals were identified as predictors:

the skewness of the cross-correlation curves between motion energy and pitch, and between pitch and loudness. Crucially, both variables involve pitch, which did not differ overall between groups. Instead, the temporal coordination of pitch and motion energy, and pitch and loudness, are a predictor of diagnosis, showing that temporal dynamics of speech and gesture should be considered when examining communication modes in individuals with and without a diagnosis of ASD.

Part G – Main discussion

The aim of chapter 4 was to evaluate whether quantifiable temporal aspects of communication differed between autistic and non autistic adults in the context of communication in autism, and whether these hypothetical differences could be related to qualitative phenotypical aspects of the disorder, such as those observed in the context of the ADOS and similar assessments, and which form the basis of a diagnosis of ASD. In this context, we provided a systematic exploration of speech and gesture produced by ASD and TD participants engaged in a naturalistic interaction, the report of a personally-experienced live event. With a view that insights from this study may in the long term be put to use to support a diagnostic process, we constrained our successive analyses increasingly to minimise the amount of both manual and subjective data processing. This aimed to make the analysis altogether feasible in a context where time and human resources are limited, and less dependent on an individual assessor's judgement. The materials used were opportunistic audiovisual recordings collected for an independent experiment ran at the Autism Research Group, which did not require any equipment which would not be available easily in research or clinical environments.

Altogether, the study was successful in identifying quantifiable indices of speech and gesture which differed between groups, and related to phenotypical aspects of ASD at various levels. Table 4.14 at the end of this section summarises the results in terms of group differences, relation between speech and gesture characteristics and clinical severity scores as measured by the ADOS and AQ scores, and relation between speech and gesture characteristics and subjective quality of communication ratings.

A main limitation of our findings was that several analyses appeared to be underpowered. In order to make valuable observations, we retained uncorrected results for multiple comparisons but interpret the findings with some caution by focusing on results which were consistent across several comparisons and/or showed a medium to large effect size.

4.33 Speech loudness: an unexpected marker of autism

One of the most unexpected results in this chapter was the fact that autistic participants were found to produce louder speech on average compared to TD participants. Better yet, using machine learning algorithms, variables derived from loudness (intensity) alone predicted a diagnosis of ASD with over 80% accuracy in our sample. Although level of loudness is not a temporal aspect of speech and gesture, we briefly discuss this result here. Loudness is the subjective perception of voice intensity, which reflects the effort required to produce speech. Most previous studies investigating prosodic aspects of voice in ASD were conducted in infants, children and young adults with autism and found either no difference between the level of loudness displayed by ASD and TD participants (e.g., Diehl & Paul, 2013; Filipe, Frota, Castro, & Vicente, 2014; Grossman, Bemis, Plesa Skwerer, & Tager-Flusberg, 2010; Quigley, McNally, & Lawson, 2016) or lower speech loudness (Scharfstein et al., 2011) in autistic compared to TD individuals (see Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017, for review). One study conducted with participants with a larger age range (10-49) found that autistic speakers produced a greater amount of utterances which were subjectively rated as “too loud” (Shriberg et al., 2001). The Schriberg study however used different materials for the ASD and TD groups: for the ASD group, videotaped recordings of ADOS interviews were used as speech samples, whereas for the TD group various samples of naturalistic conversations were sampled from other studies. Thus it is quite possible that the difference in quality and conditions of recordings should have affected the raters’ judgement differently between groups. This leaves two main explanations for why ASD participants spoke louder in our study: it could be that the respective level of loudness changes with age, and that adults with autism only tend to speak louder than their TD counterparts, or that relative levels of loudness between ASD and TD individuals are task dependent. Previous studies (apart from Shriberg and colleagues’) used single word or short sentence production tasks to study prosodic aspects of speech (or vocalisation in infants), more often than not in a way that was highly constrained by the task (e.g., answering a question or filling in the blank in a sentence). In comparison, our task required participants to retrieve information in memory about a complex and possibly anxiogenic event (a mock first aid scenario on a car crash scene), and report as many correct details as possible to an experimenter (mimicking an eye-witness interview) whilst being recorded. These conditions might be responsible for revealing differences in loudness levels between groups. In particular, emotions are known to affect various prosodic features including loudness (Banse & Scherer, 1996; Laukka et al., 2008). Autistic individuals are known to present with higher levels of anxiety overall, and manifest anxiety in atypical ways (e.g., Kerns & Kendall, 2012; Kerns et al., 2014). Different levels or manifestations of anxiety in the two groups could have resulted in the higher level of speech loudness measured in the ASD group. Our results highly advocate for a better investigation of loudness levels in naturalistic versus constrained speech productions in ASD.

4.34 Speech-accompanying gestures in autism

Until recently, much of the research on the production of speech-accompanying (or co-speech) gestures in autism has focused on infants and young children and the appearance of proto-gestures as a marker of speech development (Camaioni et al., 1997; Mundy et al., 1990, 1986; Watson et al., 2013). In spite of this, observations of "atypical" gestures (e.g., "exaggerated") are part of the clinical assessment which can lead to a diagnosis of ASD, including in adolescents and adults. More recently, a few studies have looked into co-speech gestures in adolescents with autism (de Marchena & Eigsti, 2010; Morett et al., 2016). In line with de Marchena and Eigsti, the results in this study found no difference in the rate of gestures produced by ASD and TD participants, although visual examination of the data suggest a non-significant trend that ASD individuals produced less iconic and metaphoric gestures overall (see chapter 4, Part C, section 4.18.1, figure 4.7). In contrast, Morett and colleagues found that autistic adolescents produced less metaphoric and beat gestures than TD adolescents. Observations not only of autistic but also non autistic profiles of gestures production vary quite a bit between ours and these other two studies, and warrant that more future research should explore the production of gestures in different contexts, using a common measure of frequency, to shed some light on these inconsistencies. Importantly, we found that aspects of gestures related to both clinical severity scores, and quality of communication scores. Specifically, individuals with greater difficulties in the ADOS communication domain produced longer metaphoric gestures and tended to repeat iconic, metaphoric and beat gestures more. In our design, repetition was defined as the consecutive reproduction of the same or a similar movement (e.g., rotating one hand around the other to illustrate "roll a bandage"). This suggests that although ASD individuals may produce a similar amount of gestures, those gestures tend to be more repetitive and less informative. This is supported by the observation that increased rate, duration and repetition of beat gestures were associated with poorer quality of communication ratings. Interestingly, this was the case for audiovisual but also audio-only ratings, confirming that gestures do not only enhance the listener's experience by providing a visual "aid" for them, but also improve the delivery of speech (Goldin-Meadow, 1999).

4.35 Fragmented structure of speech and gesture in autism

Combined evidence from several levels of analysis lead us to postulate that autistic communication is more fragmented than typical communication, both for speech and gesture. First, ASD participants produced more brief pauses (< 200ms) and less long pauses (> 500ms) than TD participants, interrupting speech more often for short amount of times. Mirroring this, they also produced more holds than TD participants, interrupting gestures more often. This was confirmed by recurrence data which found that TD participants showed greater recurrence stability in speech/pauses and gesture/holds patterns. This means that typical participants stayed in one state (speaking or not speaking, gesturing or not gesturing) for longer, whereas autistic participants switched more often, with a "stop-start" structure. Crucially, greater stability of gesture/hold patterns was associated with higher quality of communication scores, suggesting that fragmented communication may directly impact the subjective perception of autistic speech and gesture. In addition, more fragmented communication related to poorer communication skills (measured by the ADOS) and higher autism-related trait scores (measured by the AQ). In appearance, the finding that communication is more fragmented in autism is at odds with previous studies showing that fluent autistic children demonstrate typical phrasing and "chunk" narratives efficiently, with a typical number of grammatical and less non-grammatical pauses (Fine, Bartolucci, Ginsberg, & Szatmari, 1991; Thurber & Tager-Flusberg, 1993; see McCann & Peppé, 2003 for a review). However the number of studies are very limited and the identification of "pauses" in these studies (manually-coded speech interruptions) was different to the one we used in the current study (automatically detected interval based on the null value of pitch). In addition, to our knowledge, no study has reported evidence regarding the amount or structure of gesture holds in autism. Altogether this finding opens a new window on how autistic and typical communication might differ, in the context of a naturalistic event report.

4.36 Speech and gesture coordination: distinguishing temporal distance, degree of coupling, and mutual information

Through the various levels of analysis presented in this chapter, it became increasingly apparent that a distinction needs to be made between several aspects of what has been pooled under the umbrella term of "temporal coordination" of speech and gestures. First, the aspect which has traditionally been measured in the literature as an index of temporal coordination is the **asynchronicity** between speech and gesture, computed as the average interval between an event in one time series (e.g., onset of a gesture) and a related event in the other time series (e.g., onset of a word). Asynchronicity indicates the proximity with which two events happen in time as well as their relative temporal order. A consensual result in the literature is that gestures often precede their speech affiliates (e.g., Leonard & Cummins, 2011; Morrel-Samuels & Krauss, 1992). In ASD, de Marchena and Eigsti (2010) found that autistic adolescents produced iconic gestures less synchronously with their speech affiliates than their TD counterparts (by over 200ms), whilst Morett and colleagues (2016) reported no group difference overall, but found that autistic adolescents produced more gestures with an asynchrony over 200ms than TD adolescents. In the current study in contrast, we found that the asynchrony (temporal lock) between gesture strokes and the nearest salient peak in pitch and loudness were very close to zero and did not differ between groups, with the exception that the stroke of metaphoric gestures happened further away from the nearest pitch peak in the TD than in the ASD group (pitch leading gesture in TD but gesture leading pitch in the ASD group). In line with this null result in the temporal lock analysis, cross-correlation data also found no group difference in peak width (the interval during which speech and gesture are more likely to correlate or co-occur). Interestingly, greater asynchrony between metaphoric gestures and pitch was associated with less difficulties in the communication domain (ADOS), and with a lower severity score (AQ), indicating that better communication is not necessarily characterised by more synchronicity. In direct contradiction with de Marchena & Eigsti (2010), we also found that individuals who produced greater asynchronicity between gesture and pitch also showed greater visual gain, which means that their gestures added more value to the quality of communication than individuals showing greater synchronicity. Beyond the fact that our measure of asynchronicity was different from previous authors (discussed in chapter 4, part D, section 1.3), the discrepancy between ours and de Marchena and Eigsti's results could be accounted for by the fact that de Marchena and Eigsti calculated the absolute asynchronicity between speech and gesture, whereas we computed the relative asynchronicity (positive or negative): by averaging positive and negative lags, we obtained near-zero values, whereas de Marchena and Eigsti obtained positive values. More research is needed to clarify whether the absolute distance, or the temporal order of speech and gesture, are most relevant to quality of communication and phenotypical traits of ASD.

Second, we argue that an important descriptor of coordination is the **degree of coupling** between speech and gesture, which is preoccupied not by the average temporal interval between speech and gesture but rather by how strict this interval is. Consider scenario A in which two events are not strongly coupled: the distance between the two events is highly variable, and therefore their asynchrony would likely average near zero, but the two events are poor predictors of one another. However, in scenario B, two highly coupled events could be separated by a constant interval (e.g., 200ms) with low variability. In this case, the two events would show greater asynchrony, but each event would actually be a good predictor of the other. We argue that whilst asynchrony provides useful information about the temporal order of events, coupling is a better descriptor of the temporal coordination of speech and gesture because it distinguishes events which simply happen in close temporal proximity from events which have a strict temporal relationship between them. This idea was suggested by Leonard and Cummins (2011) who measured the variability of the lag between various time points in beat gestures and co-occurring speech and showed that the stroke of beat gestures was most closely coupled to speech. We replicated and extended this finding and showed that the stroke of all types of gestures were most closely coupled to pitch (although gestures onset and offset also showed high coupling). The pattern was similar between gesture strokes and loudness but did not reach significance. This confirms the intuitive idea that the stroke of the gesture (the most meaningful part of the gesture, often the maximum extension or maximum acceleration of the gesture) is most strictly coupled to salient events in the co-occurring speech (peaks in pitch and loudness). Interestingly, we found that ASD participants showed a higher degree of coupling between representative (iconic and metaphoric) gestures and pitch, and that it was associated with a greater amount of difficulties in the communication domain. Increased degree of coupling between deictic gestures and pitch was also associated with poorer communication skills and smaller visual gain. Although these findings should be replicated to reach a firm conclusion, this suggests that when representative gesture and speech were too closely coupled in time, they did not "add" as much information to speech. This interpretation was supported by our second measure of temporal coupling, the prominence of the cross-correlation peak between acoustic and kinematic factor. Prominence reflects how "strongly" two time series are correlated at peak compared to outside of the peak. Whilst results found no group difference in the peak prominence between motion energy and pitch, or motion energy and loudness, greater prominence (higher degree of coupling) was associated with higher clinical severity overall (ADOS total score) and with more difficulties in the communication domain specifically (ADOS communication score).

Interestingly, visual inspection of the data indicated an opposite trend for beat gestures, with beats showing looser coupling to pitch in the ASD compared to the TD group (chapter 4, part D, section 1.1.1, figure 4.11). Beat gestures are presumed to mainly carry out an emphatic function, realised by providing a rhythmical "pulse" which coincides with, and therefore draws attention to, a "pulse" in speech (Holle et al., 2012; Leonard & Cummins, 2011). In comparison to other gestures, beats can thus be described as pure temporal information. In that view, decreased temporal coupling between beats and speech would be expected to diminish the impact of beat gestures on communication. Because beats were not produced with a high frequency, the results in this study are only indicative but certainly warrant further investigation of the temporal coupling of beats and speech in autism.

The third aspect of speech and gesture coordination we identified was the **mutual information** contained in the speech and gesture streams of information. Mutual information is a mathematical term to describe how much one variable tells us about another, in other words how much knowledge of one variable reduces the uncertainty about the other variable. Here we use it to describe how much the information contained in one stream of information (e.g., gesture) reduces the uncertainty about another stream of information (e.g., speech). Taking the example of emphasis, in the sentence “I put it *right there*!”, emphasis on “right there” is expected to be marked with combined increase in pitch, loudness and motion energy (Heldner, 2001; Leonard & Cummins, 2011; Richter & Mehlhorn, 2006), showing redundancy or mutual information between all three features. In the current study, this line of evidence was provided by quantifying recurrence patterns in acoustic and kinematic features of speech and gestures. Recurrent points between two variables indicate that these variables visited similar states (i.e., took comparable values), punctually or for an interval of time. The raw amount of recurrent points and the tendency to repeat the same values for consecutive intervals (recurrence rate and trapping time in chapter 4, section E) reflect how much an individuals uses the same levels of pitch, loudness and motion energy. The organisation of recurrence points in diagonals and generally in more complex and organised structures (as opposed to random patterns, respectively the mean length of diagonal and entropy in chapter 4, section E) indicates that two variables contain more mutual information and are better predictors of each other. We found compelling evidence that typically developing participants showed greater amounts of mutual information compared to autistic participants within acoustic and kinematic features. In particular, the speech/pauses and gesture/holds patterns showed higher amounts and organisation of recurrence patterns in the TD group, as did pauses/holds patterns and loudness patterns. Crucially, higher recurrence indices were associated with fewer difficulties in the communication domain. It seems therefore that mutual information in acoustic and kinematic features in autism is lower and as a result communication patterns are less predictable, hindering general communicative skills. In contrast, the recurrence patterns between motion energy on the one hand and pitch and loudness on the other hand did not differ between groups. However correlational data strongly indicated that restricted and repetitive behaviours might interfere with these results, and that repetitive movements with no communicative intent might artificially increase the degree of recurrence between speech and gesture in ASD, without providing additional mutual information.

Altogether, and following the proposed taxonomy for the temporal coordination of speech and gesture, we found that both ASD and TD participants produced speech and gesture in close temporal synchrony, but that in ASD gestures were more rigidly coupled to speech in time. Finally, speech and gesture contained more mutual information in the TD group, making them better predictors of one another. The relation to quality of communication outcome were not always clear, but a more definite profile emerged regarding clinical severity scores: greater temporal coupling and lower levels of mutual information were associated with poorer communication skills, suggesting communication benefits from some variability in the distance between speech and gestures (particularly representational gestures), whilst maintaining a redundant structure between channels of information.

4.37 Predicting diagnosis: could speech and gesture features be a biomarker of autism?

A main motivation for investigating temporal aspects of speech and gesture in autism is the possibility that fine features of communication may underlie aspects of the autistic phenotype and contribute to the decision to give or not to give a diagnosis of autism. Identifying such markers of autism and being able to quantify them could be a very useful tool to assist the diagnosis process and make it both easier and more reliable. Using multivariate methods, chapter 4, part F evaluated whether participants' diagnostic group on the one hand, and severity scores on the other hand, could be determined using some or all of the acoustic and kinematic features that were examined in the course of the study. Supervised machine learning algorithms allowed us to test the predictive power of various models. Interestingly, we found that acoustic features on their own, and a combination of acoustic and kinematic features, predicted diagnostic group with high accuracy, sensitivity and specificity (all over 80%), showing similar validity to the gold-standard ADOS in adults (Hus & Lord, 2014). In contrast, we were not able to predict any severity scores (ADOS main and subscores, AQ) or the quality of communication ratings collected at the start of this work (chapter 4, part B). The contribution of different features and their potential as a biomarker of autism were discussed more in depth in chapter 4, part F, section 5, and much work is needed to replicate and unpack the consequences of this finding. However as a whole, successful classification of participants according to their diagnostic is an important step to help bridge the gap between quantifiable behaviours which are atypical in autism, and the qualitative perception and assessment of ASD.

QUALITY OF COMMUNICATION			
Measure	**Group differences**	**ADOS / AQ scores**	**Quality of communication ratings**
Auditory score	No	No	NA
Audiovisual score	No	No	NA
Visual gain	TD > ASD (d = .59)	No	NA
DESCRIPTIVE			
Measure	**Group differences**	**ADOS / AQ scores**	**Quality of communication ratings**
Pitch	No	No	No
Loudness	ASD > TD (d = 1.11)	No	No
Motion energy (ME)	No	Mean ME x ADOS comm (r = -.566)	No
Speech pauses			
Percentage pause	No	No	No
Rate per minute	No	No	No
Duration pause	ASD < TD (d = .81)	No	No
Percentage brief, medium, long pauses	Brief pauses: ASD > TD (d = .81) Long pauses: ASD < TD (d = .77)	No	No
Motion holds			
Percentage hold	No	No	No
Rate per minute	No	No	No
Duration hold	No	No	No
Percentage brief, medium, long holds	No	No	No
Gestures			
Rate	No	No	No
Duration	No	No	No
Repetitions	No	No	No
Handedness	No	No	No

TEMPORAL COORDINATION: TEMPORAL LOCK			
Measure	**Group differences**	**ADOS / AQ scores**	**Quality of communication ratings**
Lag variability			
IG-Pitch	TD > ASD (η_p^2 = .14)	Lag var IG-Pitch x ADOS comm (marg., r = -.470)	No
MG-Pitch	TD > ASD (η_p^2 = .20)	No	No
DG-Pitch	No	Lag var DG-Pitch x ADOS comm (r = -.621)	Lag var DG-Pitch x visual gain (r = -.424)
BG-Pitch	No	No	No
Lag variability			
IG-Loudness	No	No	No
MG-Loudness	No	No	No
DG-Loudness	No	Lag var DG-Loudness x ADOS comm (r = -.766) Lag var DG-Loudness x ADOS RRB (r = -.655)	No
BG-Loudness	No	No	No
Mean lag			
IG-Pitch	No	No	Lag IG-pitch x Visual gain (r = .471)
MG-Pitch	ASD < TD (ASD: gesture lead; TD: pitch lead (d = .803)	Lag MG-pitch x ADOS comm (marg, r = -.480) Lag MG-pitch x AQ (marg., r = .370)	No
DG-Pitch	No	No	No
BG-Pitch	No	No	No
Mean lag			
IG-Loudness	No	No	No
MG-Loudness	No	Lag MG-loudness x ADOS RSI (marg., r = .491)	No
DG-Loudness	No	No	No
BG-Loudness	No	No	No

TEMPORAL COORDINATION: CROSS CORRELATIONS			
Measure	**Group differences**	**ADOS / AQ scores**	**Quality of communication ratings**
Pitch-Loudness			
Peak correlation	No	No	No
Peak lag	No	No	No
Peak prominence	ASD > TD (marg., d = .726)	Prom x ADOS comm (marg., r = .451) Prom x ADOS tot (marg., r = .471)	No
Peak width	No	No	No
Skewness	ASD < TD (ASD loudness lead, TD pitch lead, d = 1.068)	No	No
ME-Pitch			
Peak correlation	No	Peak x ADOS RRB (marg., r = .526)	No
Peak lag	No	Lag x ADOS comm (r = .678)	No
Peak prominence	No	Prom x ADOS tot (marg., r = .478)	No
Peak width	No	No	No
Skewness	No	No	No
ME-Loudness			
Peak correlation	No	No	No
Peak lag	No	No	No
Peak prominence	No	Prom x ADOS comm (r = .571) Prom x ADOS RSI (r = .664) Prom x ADOS tot (r = .774)	No
Peak width	No	No	Width x Visual gain (marg., r = -.376)
Skewness	No	No	Skew x Visual gain (r = -.402)
Pauses-Holds			
Peak correlation	ASD > TD (marg, d = .691)	Peak x ADOS RRB (r = .557)	No
Peak lag	No	No	No
Peak prominence	No	Prom x ADOS tot (r = .523)	No
Peak width	No	Width x ADOS comm (r = .636) Width x ADOS RRB (r = .608)	Width x Aud score (r = .405) Width x AV score (r = .427)
Skewness	No	No	No

RECURRENCE			
Measure	**Group differences**	**ADOS / AQ scores**	**Quality of communication ratings**
Speech/Pauses			
RR	TD > ASD (r = .400)	No	No
L	No	No	No
ENTR	TD > ASD (r = .333)	No	No
TT	TD > ASD (r = .340)	No	No
Gesture/Holds			
RR	No	No	No
L	No	L x ADOS comm (r_s = -.566)	No
ENTR	No	ENTR x ADOS comm (r_s = -.598)	No
TT	No	TT x ADOS comm (r_s = -.573)	No
Pitch			
RR	No	No	No
L	No	No	No
ENTR	No	No	No
TT	No	No	No
Loudness			
RR	TD > ASD (r = .427)	No	No
L	TD > ASD (r = .326)	No	No
ENTR	TD > ASD (r = .363)	No	No
TT	TD > ASD (r = .325)	No	No
Motion energy			
RR	ASD > TD (r = .377)	No	No
L	No	No	No
ENTR	No	No	No
TT	ASD > TD (r = .335)	No	No
Pauses/Holds			
RR	No	No	No
L	No	No	No
ENTR	No	No	No
TT	TD > ASD (r = .406)	No	No

Pitch x Loudness			
RR	No	No	No
L	No	No	No
ENTR	No	No	No
TT	No	No	No
ME x Pitch			
RR	No	No	No
L	No	No	No
ENTR	No	No	No
TT	No	No	No
ME x Loudness			
RR	No	No	No
L	No	No	No
ENTR	No	No	No
TT	No	No	No
MACHINE LEARNING			
Measure	**Predicting diagnosis group**	**Predicting ADOS / AQ scores**	**Predicting quality of communication ratings**
Acoustic features only	Mean, standard deviation, and Lmax in recurrence plots of loudness predicted diagnosis with 81.63% accuracy	No	No
Acoustic and kinematic features	Mean, standard deviation, maximum, and Lmax for loudness, Lmax for motion energy, skewness of cross-correlation profile between motion energy x pitch, skewness of cross-correlation profile between pitch x loudness predicted diagnosis with 82.37% accuracy	No	No

Table 4.15 Summary of results for chapter 4 with effect sizes.
Groups differences, correlations to clinical scores (ADOS / AQ) and correlations to quality of communication ratings. Marg. = marginal effect (p > .05); d = Cohen's d; r = Pearson's r; ADOS comm = ADOS communication score; ADOS Im/Cr = ADOS Imagination and Creativity score; ADOS RRB = ADOS Restricted and Repetitive Behaviour score; NA = Not Applicable.

Chapter 5
Predicting a diagnosis of ASD based on acoustic features of speech: a replication study

5.1 Introduction

Chapter 4 examined atypicalities in the temporal dynamics of acoustic and kinematic features of speech and gesture in ASD. Concluding the chapter, we found that some acoustic and kinematic features showed potential as a marker of autism and predicted a diagnosis of ASD with an accuracy over 80%. In particular, prosodic variables derived from speech loudness were selected in both acoustic-only and acoustic and kinematic models.

5.1.1 Prosody in ASD

Prosody is an umbrella term which encompasses the characteristics of speech deriving from variations in the duration, amplitude and fundamental frequency of speech-sounds (Peppé, 2009). It describes aspects of speech and vocalisations which, although not semantic in nature, can directly affect the meaning of a sentence or provide paralinguistic information (for instance about the emotional state of the speaker). Peppé (2009) argued that prosody needs to be better described and understood in the context of speech-language pathologies, because it captures aspects which are only referred to as "atypical speech" or "atypical prosody" but are not quantified or comparable between pathologies or between patients. Specifically, "typical" prosody refers to the notion that any individual speaker speaks within a habitual set of parameters which make their productions idiosyncratic: a usual pitch height, pitch span, range of pitch variation, speech rate and loudness. To capture "atypical prosody", we must define the typical range of parameters and measure deviations from the norm.

In the context of ASD specifically, as noted previously, prosody has been reported to be "atypical" but exactly how prosodic patterns differ between ASD and TD individuals is unclear. Recently, quantification of prosody in ASD has been the focus of a large amount of research, because of its potential to provide a biomarker of autism. In addition, assuming differences in prosody extend to early age vocalisations, a good understanding of prosody differences in autism would constitute a quantifiable measure of autistic behaviour early in development. Pitch in particular has generated a large number of papers, most of them using univariate analyses, with mixed results. Overall, few studies have found a difference in mean pitch between groups (but see Filipe, Frota, Castro, & Vicente, 2014; Sharda et al., 2010). In contrast, a number of studies found group differences in the variability of pitch (Bonneh et al., 2011; DePape, Chen, Hall, & Trainor, 2012; Diehl, Watson, Bennetto, Mcdonough, & Gunlogson, 2009; Diehl & Paul, 2013; Filipe et al., 2014; Nadig & Shaw, 2012; Nakai, Takashima, Takiguchi, & Takada, 2014; Parish-Morris et al., 2013; Sharda et al., 2010) with generally a larger variability of pitch found in autism (but see Green & Tobin, 2009; Grossman, Bemis, Skwerer, & Tager-Flusberg, 2010; Hubbard & Trauner, 2007; Chuileann & Quigley, 2013; Scharfstein, Beidel, Sims, & Finnell, 2011). Notably, a majority of studies were conducted with children and adolescents, using tasks where single words or short sentences were elicited in highly controlled tasks. For instance, in Filipe et al. (2014) and Hubbard and Trauner (2007), participants were repeating single words, and in Diehl and Paul (2013) and Grossman et al. (2010) children and adolescents generated short answers in response to very particular tasks (e.g., complete an unfinished sentence). Only two studies have looked at infants and toddlers and found no difference in the mean or range of pitch produced by infants who were at risk of autism (Quigley et al., 2016) or were later diagnosed with ASD (Brisson, Martel, Serres, Sirois, & Adrien, 2014) and their typical counterparts. Three studies have looked at adults, one of which found no group difference in pitch (Chan & To, 2016) , one found a lower pitch range in ASD (Kaland, Swerts, & Krahmer, 2013) and the last found that pitch range was wider for higher functioning individuals and narrower for medium functioning individuals (DePape et al., 2012). A meta-analysis by Fusaroli, Lambrechts, Bang, Bowler and Gaigg (2016) concluded that overall univariate studies indicated that measures of pitch mean and range differed between ASD and TD individuals, with an accuracy of about 61–64% (i.e., the proportion of participants classified correctly). No other prosodic feature was identified as robustly different between groups in Fusaroli and colleagues' meta-analysis. Interestingly, features reflecting temporal aspects of speech were limited to syllable or vowel durations rather than referring to the dynamics of speech (e.g., temporal coordination between signals). In comparison, multivariate studies showed that acoustic features of speech reported accuracies over 70%, but the features used were too variable to allow for meta-analysis and made it hard to identify clear features which reliably predict clinical diagnosis.

5.1.2 Aim of the study

In Chapter 4, using recordings of naturalistic speech in adults with and without ASD, we found that in an acoustic-only model, loudness mean, maximum and standard deviation were sufficient to classify participants in the ASD or TD group with over 81% accuracy, 83% sensitivity and 87% specificity.

In light of the literature, the result that speech loudness was such a powerful predictor was unexpected (see chapter 4, discussion), and since our sample was relatively small (16 ASD and 16 TD participants) we recommended that results should be replicated with a larger sample in order to reach firmer conclusions. In this chapter, we set out to replicate our findings that acoustic features of speech constitute a good predictor of ASD diagnosis.

5.2 Methods

5.2.1 Context

The context of an experiment on decision-making in autism conducted by a fellow doctoral student (Dr Alida Acosta Ortiz) provided an opportunistic sample of recordings. In the original design, Dr Acosta Ortiz was interested to measure the relationship between decision making and theory of mind in a ultimatum game (UG) experiment (Güth, Schmitttberger, & Schwarze, 1982). A UG task involves two players, who have to split an allocated sum of money. The first player makes an offer, which is either accepted (in which case both players get the proposed share of the money) or rejected (in which case none of them get any of the money). UG is used to pitch two sources of motivations against each other in the decision-making balance: one is to maximise personal gain (encouraging the participant to offer as small a share as possible to the other person), the other is to act in a socially acceptable way (which would encourage the participant to make a “fair” offer). A crucial factor in this equation is the idea that the first player tries to evaluate how their offer is going to be received, and whether it is likely to be accepted. For this reason, decision-making in a UG task is believed to engage theory of mind skills, which were examined using the 'Triangles' task previously described by Castelli, Happe, Frith and Frith (2000), in which participants are asked to narrate events of 12 short videos that illustrate triangles moving either relatively randomly on a screen or in a fashion that implies intentional interactions. The audio recordings of the participant's narratives is of most interest in the current thesis.

5.2.2 Participants

32 ASD and 28 TD adults took part in the original task. Because of the nature of the current analysis (voice analysis), we excluded all female participants, and excluded a further 3 ASD participants based on age to obtain two groups matched on gender, age and IQ. The final sample consisted of 27 ASD and 17 TD participants. 16 of these participants (9 ASD, 7 TD) had also taken part in the task reported in chapter 4. All participants were recruited from the Autism Research Group's research participants database. ASD participants held a clinical diagnosis delivered by local health authorities according to DSM-IV (American Psychiatric Association, 2000) criteria for Autistic Disorder or Asperger Disorder, and diagnoses were confirmed for all participants by assessment with the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, Dilavore, & Risi, 2002). All except eight participants met the criteria for ASD on the ADOS. The participants who did not meet criteria scored 5, 5, 6, 6, 5, 3, 5 and 3 on the total ADOS score. The analysis was rerun excluding below-criteria participants but results did not differ and all participants were therefore included in the final analysis. TD participants were group matched with ASD participants on gender (all male), age, VIQ, PIQ and FIQ. Independent samples t-tests confirmed that groups did not differ significantly on any of these measures (all ps $> .1$). Autism Quotient scores (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) was collected to provide a continuous measure of autistic traits across groups. Participants characteristics and t-test statistics are summarised in table 5.1.

	ASD (n=27)		TD (n= 17)		t-test		Cohen's
	Mean *(SD)*	*Range*	Mean *(SD)*	*Range*	t	p	d
Age in years	47.5 *(12.9)*	31-73	53.1 *(10.8)*	32-71	-1.497	.142	0.470
AQ	32.2 *(7.4)*	20-46	14.2 *(5.4)*	6-24	8.635	< .001***	2.770
ADOS total	8.68 *(3.87)*	3-17					
Comm	2.88 *(1.76)*	0-6					
RSI	5.80 *(2.68)*	1-12					
Im / Crea	1.16 *(0.62)*	0-2					
Sens Behav	1.20 *(1.08)*	0-3					
VIQ	109.4 (20.4)	74-143	115.7 (12.6)	82-131	-1.222	.229	0.372
PIQ	102.8 (19.2)	59-134	105.5 (16.2)	75-136	-.459	.642	0.152
FIQ	107.3 (19.5)	73-138	111.6 (14.9)	77-135	-.727	.472	0.240

Table 5.1 Participant characteristics.
AQ = Autism spectrum Quotient (Baron-Cohen, 1995); ADOS = Autism Diagnosis Observation Schedule (Lord et al., 1999) ; Comm: Comunication, RSI: Reciprocal Social Interaction, Im / Crea: Imagination and Creativity, Sens Behav: Sensory behaviour; VIQ, PIQ, FIQ = Verbal, Performance and Full-scale Intellectual Quotient (The Psychological Corporation, 2000).

5.2.3 Materials

12 short videos (Castelli et al., 2000) were presented featuring a big red triangle and a small blue triangle, which were moving about on a framed white background. The triangles portrayed three types of movement sequences: random, goal-directed and theory of mind (ToM). In the random condition, the triangles were bouncing off the walls independently in a manner similar to billiard balls. In the goal-directed condition, the triangles movement portrayed physical interactions: dancing together, one chasing the other, one imitating the other, one leading the other. Finally in the ToM condition, the triangles movement portrayed interactions which implied mentalising: persuading, bluffing, mocking and surprising the other triangle. Each sequence lasted between 34 and 45s.

5.2.4 Procedure

Participants were presented with 12 videos (plus 2 practice) on a 24" desktop monitor in one of two pseudorandom orders that avoided presenting more than 2 of the same types of videos (random, goal directed, mental) consecutively. Throughout video presentation, participants were asked to provide a running commentary of the triangles movements and this commentary was digitally recorded on a Dictaphone. Although the distance of the Dictaphone to the participant was not precisely controlled, it was always at approximately the same distance as it was always placed on the desk near the keyboard. At the end of each video, they were asked to answer multiple choice questions to indicate whether the videos portrayed random, goal-directed or metalizing interactions. If participants chose the mentalizing option for the videos that portrayed mentalizing, they were additionally asked to choose which of five mental-state terms best described how the big and small triangles felt during the interaction. After choosing relevant answers, the next video was presented.

5.2.5 Data processing

5.2.5.1 Acoustic features extraction

First, the recordings were edited by cropping the beginning and end to only include the participant's commentary. Following the same procedure as in chapter 4, regularly sampled time-series of pitch (fundamental frequency f0, in Hz) and loudness (intensity, in dB) were extracted from the audio recordings every 50ms using Praat (http://www.praat.org/). Pitch was filtered at 50-700 Hz and volume was filtered at 0-75dB. Voice/pause behaviour was extracted as a binary variable every 50ms (1 for speech, 0 for no speech) using Praat.

5.2.5.2 Inter-trial gap correction

The inter-trial gap (interval between two videos) was variable, as it depended on how quickly the participant gave their answers to the multiple choice questions (which were not administered following all of the videos). Sometimes the participant answered very quickly and continued to comment on the actions so there was no gap in the commentary, other times participants took several seconds to make a decision. In order to exclude long gaps which represented a trial transition interval rather than a pause in speech, the longest 20 pauses were plotted on a bar chart, and the elbow of the plot (discriminating between pauses that were significantly longer than others on the left, and pauses which were of similar duration on the right) was identified as the number of pauses to take out. For instance, if the elbow of the plot was 8, the 8 longest pauses were segmented out of the data (across all time series)[2].

5.3 Data analysis

Because our aim was to replicate the results obtained in chapter 4, we followed the same procedure to analyse data sequentially and pool the results together to build a machine learning classification algorithm.

5.3.1 Descriptive data

For each participant, descriptive statistics were computed for speech **pitch** and **loudness**. ASD and TD groups did not differ on any of the pitch or loudness measures (see table 5.2).

	ASD (n=27)	**TD** (n= 17)	**Independent sample test** test statistic	p	**Effect size**
Pitch (Hz)					
mean *(SD)*	**118.4** *(23.75)*	**107.2** *(20.72)*	-1.868	.06	r = .28
median	110.7	103.8			
range (min:max)	89.83:206.65	82.16:160:18			
Loudness (dB)					
mean *(SD)*	**62.2** *(4.85)*	**60.6** *(3.28)*	1.187	.242	d = .39
median	62.0	59.8			
range (min:max)	54.0:72.0	56.2:66.3			

Table 5.2 Continuous acoustic and kinematic features.
Mean, standard deviation and range of pitch (in Hz) and loudness (in dB) for each group. SD = Standard Deviation of the Mean. Range = min:max.

In light of the results of chapter 4 where we found that ASD participants showed a greater level of speech loudness than TD participants, we plotted the distribution of mean loudness level by the standard deviation of loudness levels to visualise any potential outliers. Figure 5.1 illustrates that no outliers were found and that both ASD and TD participants spoke with a similar range of loudness.

[2] Although inter-trial pauses could have been excluded by segmenting the data manually, we had at heart to minimise manual processing in the view that this procedure could be replicated for larger samples and in the context of different datasets.

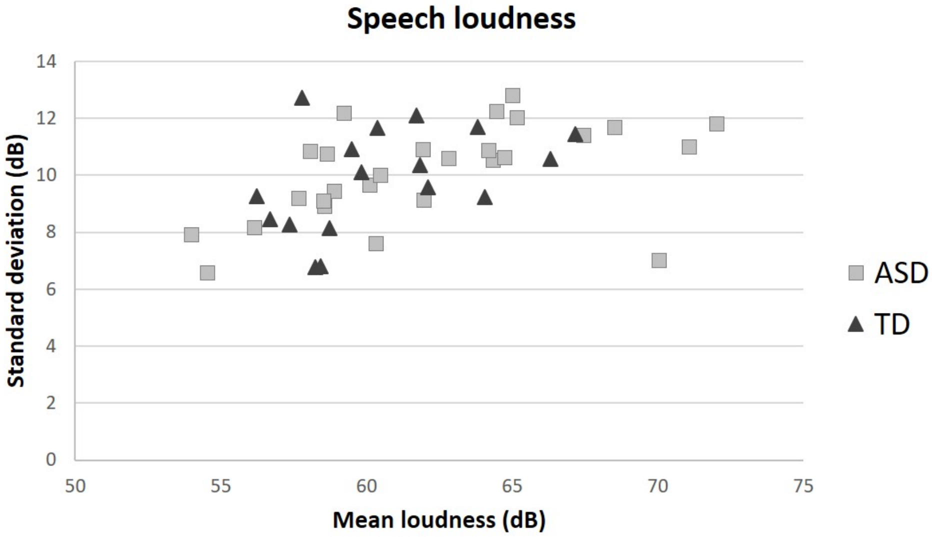

Figure 5.1 Scatter plot of the mean and standard deviation of loudness levels for the ASD (light grey squares) and TD (dark grey triangles) participants, in dB.

5.3.2 Pause analysis

For each participant, we computed the **percentage of pauses** in the analysed section of recording, the **rate per minute** and the **duration** of pauses. Following Campione and Véronis (2002), and the analytic approach of chapter 4, we computed the frequency of pauses by duration category following a **trimodal distribution**: brief (less than to 200 ms), medium (200 to 1000 ms included) and long (over 1000 ms). Bonferroni-corrected paired sample t-tests were conducted to examine group differences. ASD and TD groups did not differ on any pause measure (see table 5.3).

Speech pauses	ASD (n=27)	TD (n= 17)	t-test t	p	Cohen's d
Percentage					
mean *(SD)*	**62.4** *(12.28)*	**65.0** *(8.25)*	-.759	.452	.25
range (min:max)	42.36:88.64	48.46:75.89			
Rate (/s)					
mean *(SD)*	**1.46** *(.39)*	**1.38** *(.31)*	.716	.478	.23
range (min:max)	.81:2.16	.87:1.84			
Duration (ms)					
mean *(SD)*	**475.2** *(223.04)*	**502.5** *(170.53)*	-.415	.680	.14
range (min:max)	211.18:1100.13	305.08:868.14			
Proportion					
Brief	**.59** *(.08)*	**.57** *(.08)*	1.123	.268	.25
Medium	**.28** *(.06)*	**.30** *(.07)*	-.856	.397	.31
Long	**.12** *(.07)*	**.13** *(.05)*	-.684	.498	.14

Table 5.3 Speech pauses characteristics.
Mean percentage, rate (in pause/s), mean duration (ms) and proportion brief (< 200ms), medium ([200:1000ms]) and long (> 1000ms) duration pauses for each group. SD = Standard Deviation of the Mean. Range = min:max.

In particular, there was no difference in the proportions of brief, medium and long duration pauses (see figure 5.2).

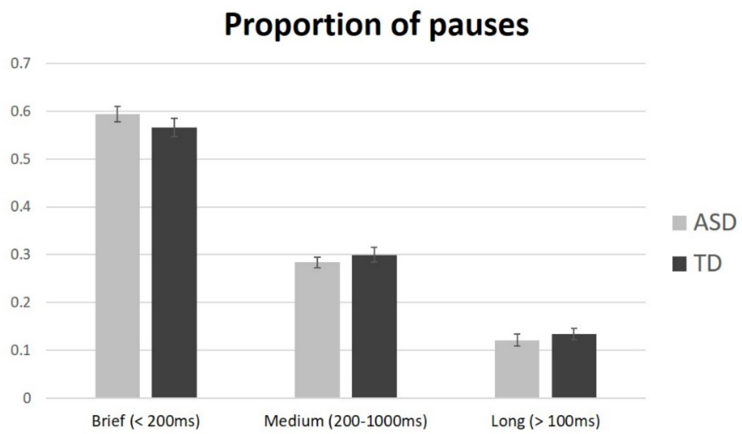

Figure 5.2 Proportion of speech pauses by duration category.
Following Campione and Véronis (2002): brief (< 200ms), medium (200 to 1000ms) and long (> 1000ms) and per group. Error bars represent the standard error of the mean.

5.3.3 Cross-correlation analysis

Cross-correlation was computed between the pitch and loudness time series. In continuity with chapter 4, we allowed a maximum lag of ±5s between signals. All recordings produced a single-peak cross-correlation profile, which was further quantified by extracting the following information: the **correlation value at peak** (i.e., the strength of the association between signals), the **peak lag** (i.e., the time-delay at which the correlation between signals is strongest), **peak prominence** (i.e., the height of the peak, indicating how the correlation at peak lag differs from baseline correlation throughout the recording) and **peak width** (i.e., the width of the peak measured at half the prominence, indicating the interval within which events in the two time series are likely to happen in proximity). Finally, in order to detect differences in lead/lag relationship between time-series, we computed the **area under the curve** on the left and the right of lag 0ms respectively. The window of integration to compute this area was defined based on the average peak width and set to 500ms. We then subtracted the area on the left from the area on the right to obtain a measure of **skewness**. A positive difference meant that the cross-correlation curve was skewed to the right, which meant that overall loudness was leading pitch. Conversely, a negative difference meant that loudness was generally lagging behind pitch. All cross-correlation values are reported in table 5.4.

	ASD (n=27)	**TD (n=17)**	**independent sample test**		**Effect size**
			est. statistic	**p**	
Peak correlation coef					
mean *(SD)*	**.679** *(.111)*	**.676** *(.067)*	199	.462	.11
Range (min:max)	.390:0.800	.590:0.800			
Peak lag (ms)					
mean *(SD)*	**-32** *(25)*	**-32** *(25)*	225.5	.908	.02
Range (min:max)	-50:0	-50:0			
Peak width (ms)					
mean *(SD)*	**578** *(202)*	**665** *(251)*	-1.260	.215	.38
Range (min:max)	250:1010	340:1210			
Peak prominence					
mean *(SD)*	**.102** *(.020)*	**.112** *(.027)*	-1.259	.181	.42
Range (min:max)	.050:0.150	.070:0.180			
Skew (Area R - Area L)					
mean *(SD)*	**-.002** *(.003)*	**-.003** *(.002)*	1.475	.148	.39
Range (min:max)	-.010:0	-.010:0			

Table 5.4 Statistics of independent samples tests conducted to compare the characteristic of the main peak in the cross-correlation profile between pitch and loudness.

Shapiro-Wilk tests revealed that correlation values at peak were not normally distributed in the ASD group (D(27)=.894, p<.05), and that the peak lag values weren't normally distributed in either group (ASD: D(27)=.614, p<.001), TD: D(17)=.611, p<.001). Mann-Whitney U tests were performed on these two variables and independent-sample t-tests were performed on the remaining variables, and all tests were Bonferroni-corrected. Results revealed no difference between groups for any of the cross-correlation indices. Figure 5.3 illustrates the cross-correlation between loudness and pitch for each group, and shows that the two profiles are extremely similar and nearly superimposed.

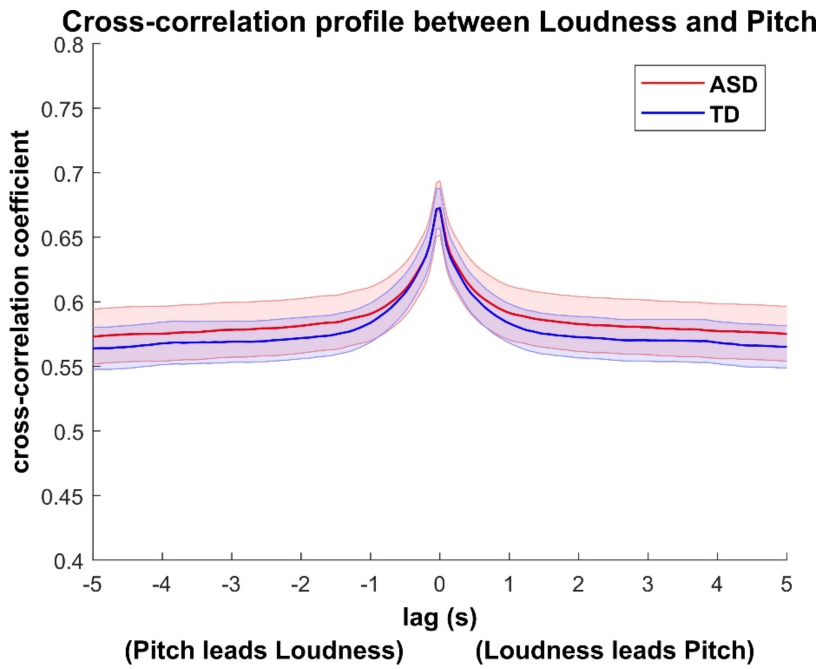

Figure 5.3 Average profile of cross-correlation between loudness and pitch between the ASD group (in red) and the TD group (in blue).
Shadowed areas represent the standard error of the mean.

5.3.4 Recurrence quantification analysis

Using the same procedure as in chapter 4, recurrence quantification analysis (RQA) was performed in Matlab using the CRP toolbox (Marwan, N.: Cross Recurrence Plot Toolbox for MATLAB®, Ver. 5.22 (R32.1), http://tocsy.pik-potsdam.de/CRPtoolbox/, Marwan, Carmen Romano, Thiel, & Kurths, 2007). RQA was performed on single variables (pitch, loudness) and cross-recurrence quantification analysis (CRQA) was performed to quantify the recurrence between pitch and loudness. All data were normalised and treated as continuous variables. So as not to generate spurious recurrence in the continuous data analyses, pauses were blanked out. Parameters for each analysis are reported in Table 5.5. and figure 5.4. shows examples of recurrence plots for one TD and one ASD recording.

	Continuous data		
	Pitch	Loudness	Pitch x Loudness
τ	2	2	2
m	6	5	5
ε	1.1	1.4	1.2

Table 5.5 Parameter values for the Recurrence Quantification Analyses.
τ = time delay; m = embedding dimension; ε = threshold (or radius).

Measures collected included **recurrence rate (RR)**, the proportion of recurrence points in the plot); **determinism (DET)**, the percentage of recurrence points which form diagonals; the mean **length of diagonal (L)** which characterises the time during which the two systems stay attuned and the **length of the longest diagonal (Lmax)** which reflects the maximum time the two systems are coupled for; **entropy**, which reflects how complex the structure is in the system; **laminarity (LAM)**, the percentage of recurrence points which form vertical lines, **trapping time (TT)**, the average length of vertical lines and the **length of the longest vertical line (Vmax)** which indicates how long the system gets trapped in one state. Shapiro-Wilk tests and Levene's homogeneity of variance tests confirmed that all variables were normally distributed and that variances were equal between the groups. For the benefit of comparing the results to those obtained in chapter 4, part E, we performed planned independent sample t-tests on RR, L, ENTR and TT, contrasting ASD versus TD group and applied a Benjamini-Rochberg procedure to control for multiple comparisons (reported in table 5.6). We found no significant group differences between any of the RQA indices. Figure 5.4 provides representative examples of recurrence plots in both groups.

		ASD mean (SD)	TD mean (SD)	t	p	Cohen's d
Loudness	**RR**	.046 (.007)	.047 (.009)	-.469	.642	.124
$n_{ASD} = 27$	**L**	3.436 (.190)	3.441 (.219)	-.082	.935	.024
$n_{TD} = 17$	**ENTR**	1.641 (.103)	1.640 (.114)	.015	.988	.009
	TT	2.611 (.255)	2.694 (.324)	-.948	.348	.285
Pitch	**RR**	.058 (.037)	.064 (.032)	-.541	.591	.173
$n_{ASD} = 27$	**L**	4.762 (.941)	4.598 (.672)	.626	.535	.201
$n_{TD} = 17$	**ENTR**	2.130 (.282)	2.098 (.222)	.406	.687	.126
	TT	4.500 (1.375)	4.226 (.879)	.732	.468	.237
Loudness	**RR**	.032 (.010)	.039 (.011)	-2.022	.050	.666
x Pitch	**L**	3.457 (.351)	3.471 (.314)	-.126	.900	.042
$n_{ASD} = 27$	**ENTR**	1.636 (.188)	1.651 (.168)	-.279	.782	.084
$n_{TD} = 17$	**TT**	3.573 (.623)	3.592 (.488)	-.109	.914	.034

Table 5.6 Summary statistics of the RQA and CRQA of acoustic features.
SD = Standard Deviation of the mean. RR = Recurrence Rate; L = Mean diagonal length; ENTR = Entropy; TT = Trapping Time. Significant p-values ($p < .05$) are marked with an asterisk. Cohen's d values indicating a medium to large effect ($d > .5$) are highlighted in bold.

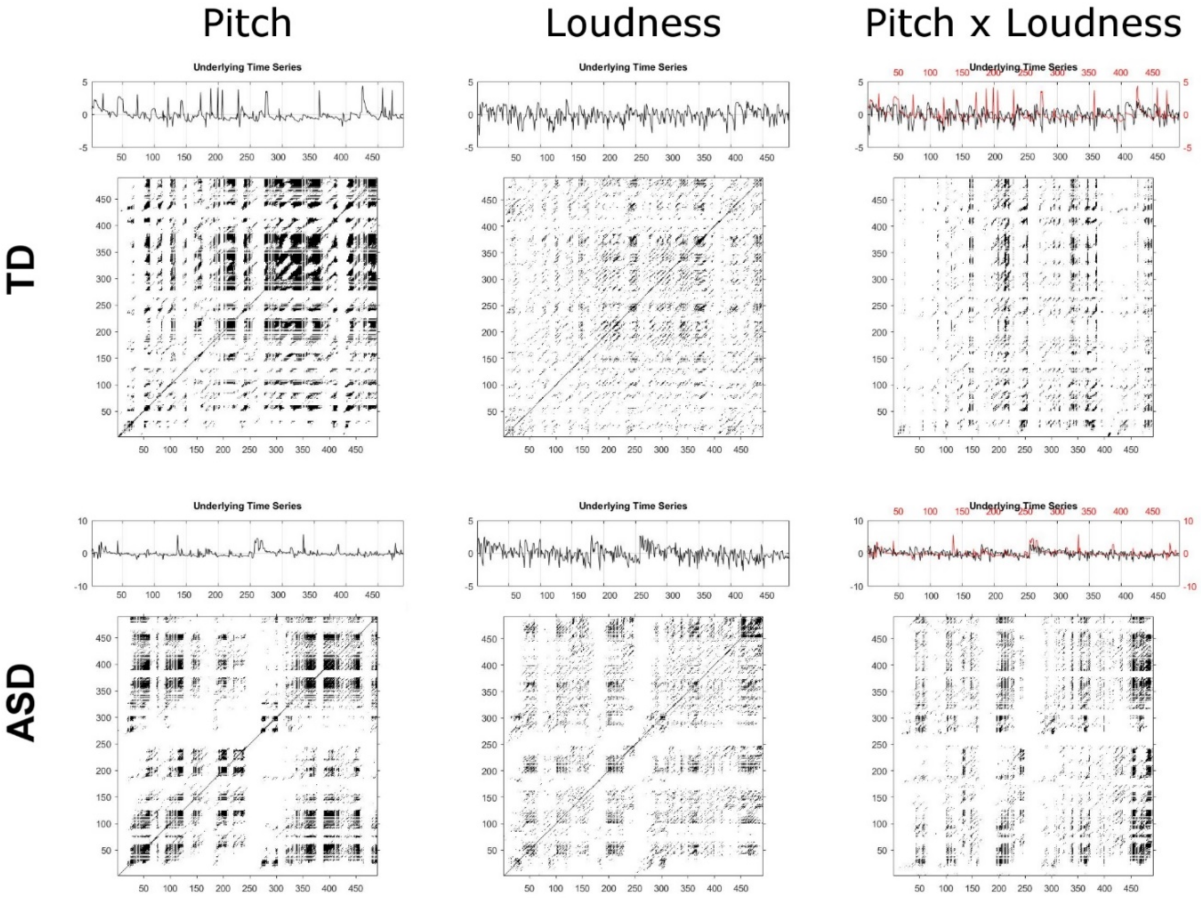

Figure 5.4 Comparison of recurrence plot for Pitch, Loudness and Pitch x Loudness within one TD (top) and one ASD (bottom) participants recordings.
For each plot, the top section illustrates the underlying time series, whilst the bottom section shows the recurrence plot. Plots were obtained using the CRP toolbox in Matlab.

5.4 Machine learning analysis

The analysis was implemented in Matlab using scripts provided by Riccardo Fusaroli and adapted with permission. All participants were included in the final analysis. Features containing too many zeros or repeating the same value (e.g., number of peaks in the cross-correlation profile) were excluded as in Chapter 4. The features included in the final analysis are summarised in table 5.7.

Descriptive	Recording	total duration
	Pitch	mean, standard deviation, min, max
	Loudness	mean, standard deviation, min, max
	Pauses	percentage, frequency, mean duration, max duration, standard deviation duration, percentage brief duration, percentage medium duration, percentage long duration
Cross-correlation data	Pitch-Loudness	correlation coefficient at peak, peak width, peak prominence, skewness
Recurrence quantification	Pauses Pitch Loudness Pitch-Loudness	RR, DET, L, Lmax, ENTR, LAM, TT, Vmax

Table 5.7 Initial features included to construct both classification and score prediction algorithms.

We used ElasticNet (Zou & Hastie, 2005) for feature selection. Diagnosis was predicted using a 5-fold cross-validation and the accuracy was balanced using Variational Bayesian mixed effects inference. Several models were compared for best fit including logistic regression, discriminant functions (linear, diaglinear, quadratic, diagquadratic and mahalanobis), naïve Bayesian classifier and support vector machine models. ADOS total scores and individual factors scores were predicted using a 5-fold cross-validation multiple linear regression. Both analyses were iterated 100 times to test for stability of results and 95% CI were calculated.

Contrary to prediction, and in stark contrast with the results in chapter 4, no features were selected in the procedure and we found no acoustic features or combination of acoustic features which contained enough information to accurately classify participants according to their diagnostic group. Figure 5.5 illustrates the distribution of speech loudness and standard deviation of speech loudness levels in both groups (mirroring figure 4.37) which shows that the clear demarcation between groups observed in chapter 4 was no longer visible in this dataset.

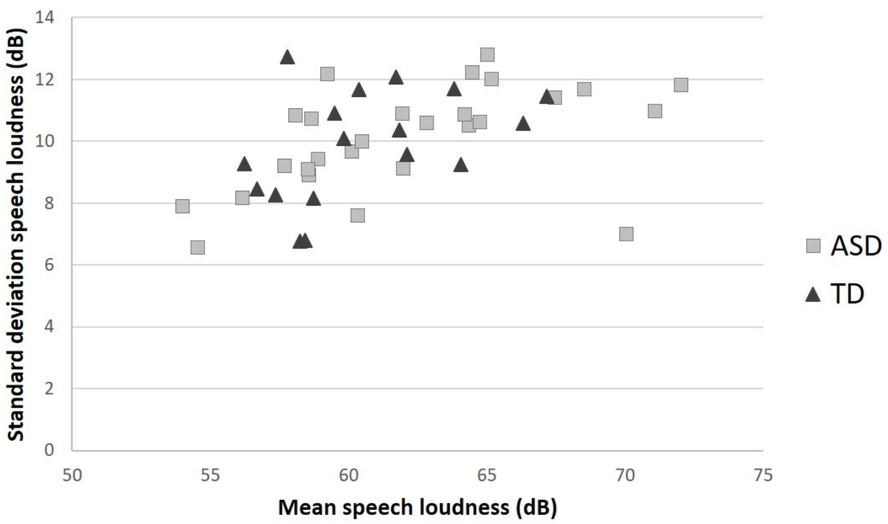

Figure 5.5 Distribution of the mean (on the x-axis) and standard deviation (on the y-axis) of speech loudness per group.

No features were selected in the procedure and we found no acoustic features or combination of acoustic features which contained enough information to predict ADOS or AQ scores accurately.

5.5 Post-hoc analysis: Overlap group

In order to better understand the differences between the results in chapter 4 and the current chapter, we examined the acoustic data of the 16 participants (9 ASD, 7 TD) who took part in both studies (the "overlap" group). The correlation between the mean loudness in chapter 4 and 5 was positive but not significant ($r(16) = .420$, $p = .106$). Surprisingly, the correlation between the mean pitch in chapter 4 and chapter 5 in the overlap group was very low ($r(16) = .073$, $p = .787$), but visual inspection of the data suggests that this may be driven by 4 outliers (2 ASD, 2 TD, see figure 5.5).

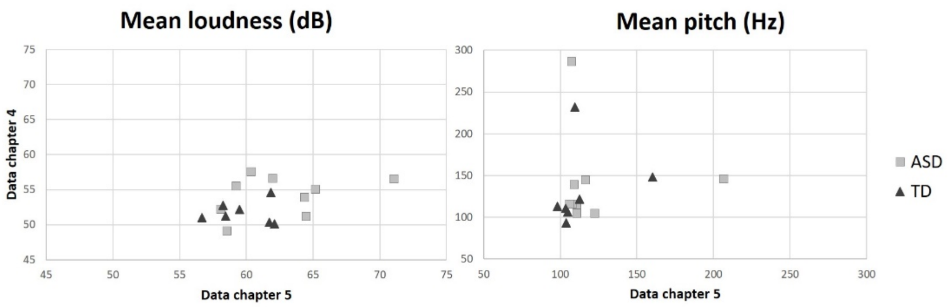

Figure 5.6 Scatter plot of the mean and standard deviation of loudness levels for the ASD (light grey squares) and TD (dark grey triangles) participants, in dB.

Altogether, the low sample size of the overlap group and the presence of outliers make it difficult to draw firm conclusions. However, the limited correlations between pitch and loudness levels between chapter 4 and chapter 5 suggest that the task performed may impact acoustic features of voice significantly.

5.6 Summary and discussion

In this study, we aimed to build on the literature on prosody in autism and our own results from chapter 4 regarding the predictive power of acoustic features to classify individuals according to diagnostic group (ASD vs TD) on the one hand, and predicting clinical severity scores on the other hand. Past studies have indicated that models can be constructed using acoustic features which predict a diagnosis of autism with high accuracy (70% and over), although the features identified as good predictors and how they are computed varies widely between studies (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017). In the previous chapter, we found that variables derived from speech loudness in particular allowed us to predict diagnostic group with an accuracy over 80%.

In the current study, we used data from an opportunistic sample of ASD and TD male adults matched on age and IQ who took part in a video-based theory of mind task (Castelli et al., 2000). Surprisingly, we found that no single feature or combination of acoustic features could classify individuals into diagnostic groups or predict severity scores (ADOS and AQ) accurately. Traditional univariate group comparisons confirmed that the two groups showed very similar voice characteristics in the task (as opposed to showing clusters or outliers driving the data).

A number of factors could explain the discrepancy between this and previously reported evidence. First, a limitation of our results is that the data was issued from an opportunistic sample rather that a specifically designed study, and the exact distance between the participant and the recorder was not specifically controlled for. However the setup was extremely similar between participants as the recorder was always placed near the keyboard that the participant used to give responses, so we argue that this factor alone is unlikely to explain the loss of predictive power.

Second, a main differences between ours and previous studies is the age of the population tested. Whilst most other studies have focused on infants, children and adolescents at risk or with a diagnosis of ASD (Asgari, Bayestehtashk, & Shafran, 2013; Bone, Black, & Lee, 2012; Bone et al., 2013; Bonneh et al., 2011; Kakihara, Takiguchi, Ariki, & Nakai, 2015; Kiss, van Santen, Prud'hommeaux, & Black, 2012; Marchi et al., 2015; Oller et al., 2010; Santos et al., 2013), our study investigated a male adult population. Considering the changes that the voice undergoes, especially in males and through adolescence (e.g., Hollien, 2012), it could be expected that differences before or during adolescence might not persist in the same form in adulthood. To our knowledge, only one other unpublished study examined voice patterns in adults (Fusaroli, Bang, & Weed, poster presented at the International Meeting For Autism Research 2013). Interestingly, the task used by Fusaroli and colleagues was the same as the one used in the current study (theory of mind test using videos of moving triangles), the participants were also recruited and tested in the UK and the features examined were very similar to ours (fundamental frequency, intensity, and speech/pause patterns, using descriptive statistics and CRQA analyses). With a relatively small sample of 10 ASD and 13 TD participants, Fusaroli and colleagues constructed a model which was able to allocate participants to ASD or TD group with 86% accuracy. In addition, they were able to reconstruct the 10 ASD participants AQ scores from acoustic features successfully (R squared = 0.800, $p < .01$). Using a very similar range of features and the same machine learning procedure (scripts were adapted from Riccardo Fusaroli's with permission), however, we did not replicate either of these findings in our somewhat larger sample of 27 ASD and 17 TD participants. Discrepancy of results between these two very similar studies warrant more research using larger samples to better describe voice patterns in autistic adults and establish whether acoustic features can discriminate "autistic" voice patterns in adults as they do in children and adolescents.

Now, comparing the results in this chapter (using the "Triangle task") to those obtained in chapter 4, it was again surprising to find that acoustic features did not allow us to classify participants accurately. In the previous chapter, using data collected during the retrieval and report of a first-aid scenario ("First Aid task"), we successfully constructed an algorithm that predicted participants' diagnostic group with over 80% accuracy. Moreover, a third of the participants in the current study had also taken part in the study reported in chapter 4 (9 ASD, 7 TD). To get a sense of the task effect, we compared the acoustic profile (loudness and pitch) of these participants between the First Aid task and the Triangle task, and found that the correlation between the two were fairly small. Although admittedly the resulting sample size was low, this suggests that the nature of the task could be responsible for changing voice patterns significantly and display – or not display – differences between groups. Specifically, the First Aid task in chapter 4 involved the retrieval in memory of a complex, emotionally-loaded, personally experienced event, with emphasis on agency details (who did which task) and the order of events. This type of task is likely to put a strain on autistic individuals' cognitive skills who are known to show difficulties with unsupported relational memory tasks (e.g., Gaigg, Bowler, Ecker, Calvo-Merino, & Murphy, 2015; Ring, Gaigg, & Bowler, 2016), temporal order retrieval (e.g., Poirier, Martin, Gaigg, & Bowler, 2011) and also difficulties reflecting on the self (Crane, Goddard, & Pring, 2009; Lind, 2010) with repercussions on episodic memory (Crane & Goddard, 2008; Klein, Chan, & Loftus, 1999). Autistic children for instance show poorer memory for the agent of remembered actions (self or other) compared to typically developing children (Russell & Jarrold, 1999). In addition, emotions are known to affect prosodic cues such as fundamental frequency and quality of voice (e.g., Bachorowski & Owren, 1994; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). In the First Aid task participants were involved in a mildly stressful situation during the scenario itself, then interviewed about everything they could remember, both of which could have generated a degree of stress or arousal, and affected the quality of their voice. For all these reasons, any differences between ASD and TD groups could have been exacerbated.

By comparison, the Triangle task in this chapter required an on-going commentary of the movements of two simple shapes. There was no demand on memory, and the task could be successfully completed with fairly simple language (describing the colour and shapes of the triangle, the simple box display on the screen, and the movements of the triangles around the screen). The videos in the current tasks were also extremely similar as they involved the same elements and only differed in the type of interaction portrayed in the movement of the triangles. A literature search for context- or task-dependent differences in acoustic-prosodic features of speech (e.g., single words vs narratives, constrained responses vs naturalistic speech) claimed no results, but we speculate that the theory of mind task (or triangle task) would encourage more monotonous, repetitive voice patterns in both groups, whereas the first-aid scenario free recall task would produce a more complex acoustic and prosodic profile in TD participants at least, and would therefore be more likely to reveal fine differences in voice (and gesture) patterns. This prediction needs to be addressed in a direct comparison study. In addition, the discrepancy between our results in chapters 4 and 5, and the discrepancy between our results and Fusaroli et al. (IMFAR 2013) warrants further research to examine the test-retest reliability of acoustic features, even if the task were held constant.

Considered on the whole, the data from chapters 4 and 5 suggest that multivariate analyses of speech and gesture show some promise for identifying acoustic and kinematic features of speech that may provide a marker for some aspects of the social-communication difficulties characteristic of autism. However the context and limitations of when acoustic features of speech and gesture have predictive value and how reliable such measures are requires more research. In particular, features extraction and analysis need to be consistent across studies, and the tasks and contexts used to generate data need to be altered and compared systematically to discriminate task and contextual effects from actual speech and gesture pattern differences between groups.

Chapter 6
General discussion

The overarching aim of the work presented in this thesis was to assess the integrity of temporal processing in Autism Spectrum Disorder (ASD), in highly controlled experimental tasks but also within naturalistic behaviour, and to address the question whether differences in temporal processing may relate to core clinical features of ASD and its phenotypical manifestations. In particular, we were interested in probing temporal aspects of communication in ASD and their impact on communicative skills. The motivation for this work stemmed from reports that individuals on the autism spectrum experience time differently (Boucher, 2001) and show evidence of atypical timing and difficulties with social timing (Wimpory et al., 2002), which has been proposed to contribute to core clinical features of the disorder (Allman, 2011).

The thesis was structured in two main parts: the first part provided a traditional, psychophysical assessment of the integrity of temporal perception in autism, whilst the second part provided a systematic analysis of temporal aspects of speech and gesture coordination in autism. In the discussion that follows, we will evaluate our initial goals, assess what we have learnt or achieved from this work, and reflect on the next steps for research.

6.1 Is temporal processing atypical in autism?

The literature review in chapter 1 offered somewhat mixed evidence about the integrity of temporal processing in autism. Whilst a majority of studies reported atypical performance in time perception (the most robust result being reduced sensitivity of temporal judgements, e.g., Falter et al., 2012; Karaminis et al., 2016; Kargas, López, Reddy, & Morris, 2014), the nature of the differences between ASD and TD performance was found to be more elusive. Some elements of response were brought forward by Maister and Plaisted-Grant (2011) who showed that autistic participants, in contrast with TD participants, did no engage their episodic memory to process long durations (30-45s), and that increased variability in the processing of short durations (under 10s) was underpinned by attentional difficulties. In support of this idea, Lambrechts, Falter-Wagner, & van Wassenhove (2017) found that autistic adults engaged less neural resources for the processing of short durations, regardless of the task instructions. Together, these two studies suggest that whilst temporal processing is not dramatically compromised in ASD, the underlying cognitive processes might differ in comparison with the typical population. Chapter 2 provided a direct evaluation of the perception of short durations in ASD using a comparison task, which was chosen over a bisection or generalisation task to reduce the task memory load. In retrospect, the comparison task lasted longer than a bisection or generalisation task would have, and might have placed a strain on attentional resources in the ASD group. We found on the one hand, that half of the autistic participants performed the task typically. In particular, the performance of autistic participants was characterised by classic effects including Vierordt's law (the tendency to construct a representation of the average duration presented during the task, resulting in underestimating long durations and overestimating short durations) and distance effect (the observation that participants respond faster when discriminating durations which are further apart, or in other words when the task is easier). The central tendency result is particularly interesting because it challenges theories of autism proposing that autistic perception is defined by piecemeal, locally-focused processing (Weak Central Coherence account, Frith & Happé, 1994) or that autistic individuals have hypo-priors (Pellicano & Burr, 2012b), which would predict that they should make temporal decisions based on single stimulus encoding rather than on a predicted (average) representation of the stimulus. On the other hand, we found that a large proportion of the ASD participants (nearly 50%) and a non-negligible proportion of TD participants (around 20%) could simply not perform the task above chance level in at least one of the sensory modalities, particularly the visual modality. Although participant exclusion on the basis of poor performance is not uncommon in either the time perception literature or the autism literature more generally, the high proportion of exclusions in the study reported in Chapter 2 needs to be addressed. Exploration of participant characteristics revealed that in the ASD group (but not the TD group), excluded participants had lower IQ scores than included participants, particularly performance and full-scale IQ. This suggests that temporal processing in autism might rely more heavily on cognitive resources than in the TD group. Together with evidence from Maister and Plaisted-Grant, and from

Lambrechts et al., we propose that in autism, attentional processes fail to prioritise temporal processing of short durations even when the task instructions focus on duration judgment. As a consequence, temporal processing competes for neural resources with co-occurring processes (e.g., sensory processes), which would arguably make performance 'noisier' (i.e., lower sensitivity, higher discrimination thresholds) and more reliant on the overall cognitive resources of an individual (as tapped into by IQ). This interpretation also predicts that depending on task difficulty and on the cognitive profile of participants, group differences may or may not reach statistical significance in different studies. To test this hypothesis, future research could compare time perception under full vs. divided attention conditions in ASD and TD groups and/or systematically comparing individuals with and without co-occurring intellectual impairments. For instance, adapting the paradigm from Lambrechts et al. (2017), participants could be required to provide duration judgements of auditory intervals only (single task condition), or both duration and pitch judgements of auditory intervals (dual task condition). If our hypothesis is correct, autistic participants would be expected to show split-attention in both conditions, and therefore no difference in the precision of temporal judgements between conditions. In contrast, we would expect TD participant to allocate more attentional resources to temporal processing in the single task, with an expected drop in performance in precision in the dual task. A second, related recommendation for future research is that investigators should monitor and report their exclusion rates and criteria, and explore the profile of excluded participants both in typically developing and clinical groups.

The idea proposed that autistic individuals depend more on cognitive for temporal processing resources has potentially huge consequences in a population where about half of all individuals have intellectual disabilities (Charman et al., 2011). Temporal processing is ubiquitous: sensory integration, motor coordination and language are examples of where disrupted timing could impact behaviour, and examples of domains in which many autistic individuals show some degree of difficulty. Yet it is still unclear whether temporal aspects in these domains – and others – really *are* atypical in ASD.

6.2 Does atypical temporal processing in autism matter?

Despite the recent surge of interest in temporal perception in autism (e.g., Allman, 2011; Brenner et al., 2015; Brodeur, Gordon Green, Flores, & Burack, 2014; Falter, Noreika, Wearden, & Bailey, 2012; Karaminis et al., 2016; Maister & Plaisted-Grant, 2011), few studies have addressed the question whether atypical timing and time perception (as investigated in the laboratory) actually matter when it comes to day-to-day behaviours in autism. One study (Kunchulia et al., 2017) found intact time-based expectancies in autistic children, in a task where the duration of the interval that preceded the display of a moving target predicted the direction in which the target would move. This result suggests that the differences in temporal processing in the context of psychophysical experiments might not be relevant for real-world instances of temporal processing. In particular, the question arises whether atypical temporal processing contributes to 1) the phenotypical profile of autism, and 2) some of the interview-based observations that form the basis of a diagnosis of autism. The second part of the work in this thesis focused on studying temporal aspects of communication in autism. Specifically, chapter 4 provided a systematic exploration of successive levels of temporality in the production and coordination of speech and gesture in autistic and non autistic adults, in the context of a mock eye witness interview. We will not repeat here the detailed results that were obtained at each stage of the analysis, and that were discussed in detail earlier. Instead, the focus here will be on the broader picture that emerged from the ensemble of results.

First, the temporal structure of speech and gesture was found to be more fragmented in autistic compared to typically developing adults. Speech and pauses, as well as gesture and holds, alternated more frequently and for shorter intervals of time in ASD participants, producing a more interrupted or “stop-start” structure. Using the analogy of music, a fragmented song would be a song using shorter notes and melodic sequences and shorter silences between them, something that might sound a little bit like the crackling noise of an old recording. Importantly, a greater degree of fragmentation was associated with lower quality of communication ratings and poorer communication skills, suggesting that fragmented structure affects phenotypical aspects of autistic communication, and that it also contributes to the clinical observations that form the basis of a diagnosis of autism. Incidentally, to our knowledge, our study is the first to report data regarding the structure of gestural holds in autism. Gestural holds and their relation to speech are also hugely under-researched in the general population (Esposito & Esposito, 2011; Esposito, Esposito, & Trojano, 2007), and our results strongly advocate for the need to research them more extensively, in order to better understand the role they play in communication (typical and atypical).

Second, the degree of coupling between acoustic and kinematic features of speech and gesture was higher in the ASD group, in other words the interval between a salient event in one time series (e.g., loudness) and the nearest salient event in another (e.g., motion energy) was more consistent in the ASD than in the TD group. To pursue the musical analogy, a song with greater coupling between acoustic features would be a song where the melody and the volume always change *together* (albeit not necessarily *synchronously*). For instance, we can imagine a song where higher notes are always followed by louder singing, something that might not sound very rich musically! Again, results suggested that stricter coupling of speech and gesture features was associated with both lower quality of communication ratings (phenotypical manifestation) and poorer communication skills (clinical observation).

Third, recurrence patterns of acoustic information were less stable and less complex in the ASD compared to the TD adults, suggesting that prosodic information was less consistent and predictable in autism within and across modalities. We tentatively interpret this finding in terms of "mutual information", in other words the degree to which a time series is predictive of itself or of a related time series. Because acoustic recurrence patterns show lower stability and complexity within and between features, this means that information about one feature (e.g., change in loudness) does not give as much information about itself (e.g., change in loudness at a different time) or about another feature (e.g., change in pitch) in autistic compared to non-autistic individuals. Using the musical analogy again, a song with poor mutual information may be a song with no choruses, and in which each verse would have a different pattern and length, something which may be quite difficult to sing along to! There was no evidence that diminished mutual information in speech and gesture features affected quality of communication ratings, but it was associated with poorer communication skills, suggesting once again that the fine temporal organisation of speech and gesture may be part of what forms a clinical opinion about an individual's diagnosis.

These three observations – more fragmented temporal structure of speech and gesture, greater degree of temporal coupling between speech and gesture, reduced mutual information in the temporal recurrence – and their relation to phenotypical and clinical aspects of ASD firmly suggest that atypical temporal processing in autism *does* matter. Yet, the temporal aspects we found to be atypical had little to do with the estimating of durations of a single event, or determining the temporal order of two events, or other skills that were tested in the psychophysical study of Chapter 2. Instead, fragmentation, coupling and mutual information relate to the general coherence of communication, or what Fusaroli and Tylén (2015) call individual *self-consistency*. The idea is that by producing regular patterns of communicative signals, such as consistent lexical choices or consistent pause timing, we reduce the cognitive cost for the receiver and make it easier to understand and respond to. This strategy in turn facilitates coordination with others, which is the foundation for social interaction. For instance, Vesper, Van Der Wel, Knoblich, and Sebanz (2011) found that pairs of participants who performed a joint motor task (but not pairs who worked next to each other on a non collaborative task) reduced the variability of their own actions to make them more predictable. As a consequence, joint coordination was higher and pairs performed better in the task. Our results suggest that autistic adults show reduced self-consistency in the structure and recurrence of their patterns of speech and gesture, which could contribute to increased cognitive cost for the receiver. However, they also demonstrated increase self-consistency in terms of coupling between speech and gesture features, presumably reducing the cognitive cost for their listener in that respect. Results from chapter 4 therefore suggest that autistic communication is characterised by a different tuning of self-consistency compared to typically developing individuals. It would be of interest in future studies to examine how autistic and non-autistic individuals adapt their self-consistency during social interactions with one another.

A limitation to these findings came directly from the results of the following chapter (chapter 5). Exploring the same temporal aspects of communication, this time for acoustic features only (pitch and loudness, and speech/pause behaviour), we found no differences in self-consistency between ASD and TD groups. Whilst the features extraction methods, data processing and analyses were the same, there were major differences between the two tasks from which the data were extracted. In chapter 4, in a mock eye-witness scenario, participants were reporting from memory a personally-experienced, potentially anxiogenic first-aid scenario in which they assisted and experimenter to carry out a series of actions on a manikin (the "First Aid task"). In chapter 5, participants were commenting online on the respective movements of two triangles in short videos presented on a screen (the "Triangle task"). A number of factors could account for the discrepancy in results between the two chapters. As discussed in chapter 5, the First Aid task relied on episodic memory and temporal order, which are both known to be atypical in ASD (Gaigg et al., 2015; Poirier et al., 2011; Ring et al., 2016), whereas the Triangle task did not. In addition, autistic individuals have been shown to experience anxiety more often than TD individuals (Kerns & Kendall, 2012) and given that emotions and arousal states are known to impact on prosodic cues such as fundamental frequency and quality of voice (e.g., Bachorowski & Owren, 1994; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009), it is possible that the First Aid task triggered different patterns of speech and gesture in ASD compared to TD, whereas the triangle task did not. If this is the case, it would mean that temporal aspects of communication such as self-consistency aspects are context-dependent, and ASD-related atypicalities more apparent in more emotionally salient settings. Exploring and understanding the contexts in which they may occur could be addressed in future research to help 1) provide optimal conditions for the detection of autistic traits; 2) support autistic individuals' communication in the situations where it is most needed.

6.3 Moving away from slow, manual, subjective observations and towards fast, automatic, objective measures

One of the motivations for the work presented in this thesis was to address the lack of quantifiable variables which are indicative of a diagnosis of autism. As mentioned previously, a diagnosis of ASD is based on subjective, clinical observations using interview-based tools such as the ADOS. In these assessments, the clinician(s) records the occurrence of atypical behaviours, usually on a qualitative scale (e.g., odd or typical), or based on a limited numeric scale (e.g., number of iconic gestures observed during a 5 minute task) (e.g., Lord et al., 2000). Subjective observations present different types of limitations: first, by definition, they rely on the observer's judgement, raising the issue of inter-rater reliability and test-retest reliability. Second, they are also costly in time and human resources: a diagnosis is usually established after conducting a number of one-to-one assessments that involve a team of specialists. Finally, these assessments do not provide a quantifiable measure that can be followed-up as a measure of severity or progress over time. In chapter 4, we quantified temporal aspects of speech and gesture coordination at different levels, progressively moving away from manually coded variables (e.g., gesture coding) to only retain fast, automatic data extraction and analysis. In addition, we used materials which are readily available in clinical (and research) contexts: audiovisual recordings of interview sessions, recorded with standard equipment. The recordings presented some noise (flickering lights, background noise, back-channelling from the experimenter) which we were able to either correct for (e.g., by introducing a minimum threshold to detect movement in order to correct for light flickering) or average out (i.e., background noise). First, our descriptive data analysis (chapter 4, part C) and temporal lock analysis (chapter 4, part D) combined manually coded (gesture timepoints, gesture type) and automatically extracted (pitch, loudness, motion energy) variables. Then, in the cross-correlation (chapter 4, part D) and recurrence quantification analyses (chapter 4, part E), we removed manually-extracted information altogether, both in the inclusion criteria and from the data itself. As a result, some fine-grained information such as the differences in temporal dynamics between various types of gestures was lost. On the positive side, however, both cross-correlation and recurrence quantification analyses were able to detect complex relationships between acoustic and kinematic features which were neither detectable nor quantifiable by an observer. Overall, the various analyses did not replicate each other's results exactly because of their intrinsic strengths and weaknesses. However, the results were coherent between analyses and allowed us to construct a rich picture of temporal patterns of speech and gesture overall. More research is needed to determine whether each type of analysis measured different aspects of the same underlying phenomenon, or whether they measured different mechanisms altogether. In this work, we proposed a framework to interpret temporal coordination of speech and gesture along 3 dimensions: asynchrony, coupling and mutual information. Overall, our data supports the idea that fast, automatic and objective processing can provide valuable information to better understand the communication style displayed by autistic individuals, and how it relates to phenotypical and clinical aspects of the disorder.

6.4 Predicting a diagnosis of autism: could acoustic and kinematic features of speech and gesture be biomarkers of ASD?

The last analysis in chapter 4 (part F) pooled together all previous variables derived from acoustic and kinematic features of speech and their coordination in time, and used a machine learning algorithm to construct a model able of predicting diagnosis group accurately. We found that either acoustic-based features on their own, or a combination of acoustic and kinematic features, could predict diagnosis with over 80% accuracy, replicating and extending recent findings that acoustic features show high predictive power in infants (Oller et al., 2010; Santos et al., 2013), children (Asgari et al., 2013; Bone et al., 2013, 2014; Bonneh et al., 2011; Kakihara et al., 2015; Kiss et al., 2012; Marchi et al., 2015) and adults (Fusaroli, Bang, & Weed, 2013) with ASD, as reviewed in Fusaroli, Lambrechts, Bang, Bowler, and Gaigg (2017). Whilst previous studies focused predominantly on prosodic aspects of speech (tone and quality of voice), our study in chapter 4 also introduced features that reflected temporal aspects of communication. In chapter 5, however, we failed to replicate the result from chapter 4 using a second, acoustic-only dataset collected with a different task, despite the fact that a percentage of the same participants were involved in both studies. Interestingly, both our studies used scripts adapted (with permission) from the scripts used by Fusaroli and colleagues (2013), and the dataset in chapter 5 was based on the same task than Fusaroli et al. Despite this, results were not consistent between the study presented in chapter 5 and Fusaroli et al.'s. This raises the question whether the predictive power of acoustic (and possibly kinematic) features is reliable. Although, as a whole, results in the literature are extremely promising, the studies published so far usually involve relatively small sample sizes for multivariate analyses, with often (as is the case in this work) opportunistic data collection involving a variety of tasks. In addition, the features selected for the analysis and the way they are computed differ widely from one study to the next. Our mixed results in chapter 4 and 5 also raise some doubts about publication bias: the tendency for null result studies not to get published may be amplifying the idea that we are close to finding a reliable, quantifiable marker of ASD when more work is needed to identify exactly under which conditions reliable classification can be performed. In addition, there is as of yet no consensus about which features are the most informative, whether the aim is to classify individuals according to their diagnosis group, predict severity scores or even follow the progress of an individual over time. Finally, a danger of complex, multivariate methods is that the (often abstract) selected features and models are challenging to portray otherwise than mathematically, and are difficult to relate to a theoretical framework. Consequently, even if high reliability can be achieved to classify individuals, at this stage multivariate methods do not provide the means to understand an individual's strengths, weaknesses and needs. More research is needed to understand the relationship between predictive features, phenotypical outcome and clinical severity for the person. Following Fusaroli et al. (2017), we advocate for larger sample sizes which can be obtain by together multicentric, anonymised datasets, promote a more open mode of research by sharing scripts and converge towards consensual features. We also recommend that future studies should be theory-driven so as to provide a more useable framework for clinical and educational applications.

References

Abbott, A. E., Nair, A., Keown, C. L., Datko, M., Jahedi, A., Fishman, I., & Müller, R.-A. (2016). Patterns of Atypical Functional Connectivity and Behavioral Links in Autism Differ Between Default, Salience, and Executive Networks. *Cerebral Cortex*, *26*(10), 4034–4045. https://doi.org/10.1093/cercor/bhv191

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., … Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060–12065. https://doi.org/10.1073/pnas.1302982110

Alcántara, J. I., Cope, T. E., Cope, W., & Weisblatt, E. J. (2012). Auditory temporal-envelope processing in high-functioning children with Autism Spectrum Disorder. *Neuropsychologia*, *50*(7), 1235–1251. https://doi.org/10.1016/j.neuropsychologia.2012.01.034

Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *31*(1), 24–31. https://doi.org/10.1037/h0081647

Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*(2), 101–117. https://doi.org/10.1016/S0376-6357(98)00043-6

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, *22*(1–2), 39–58. https://doi.org/10.1016/0023-9690(91)90016-2

Allely, C. S., Gillberg, C., & Wilson, P. (2013). Neurobiological abnormalities in the first few years of life in individuals later diagnosed with Autistic Spectrum Disorder: A review of recent data. *Behav.Neurol.*, *2014*(1875-8584 (Electronic)).

Allman, M. J. (2011). Deficits in temporal processing associated with autistic disorder. *Frontiers in Integrative Neuroscience*, *5*(March), 2. https://doi.org/10.3389/fnint.2011.00002

Allman, M. J., & DeLeon, I. G. (2009). No time like the present: time perception in autism. In A. C. Giordano & V. A. Lombardi (Eds.), *Causes and risks for autism* (pp. 65–76). Nova Science Publishers New York, NY.

Allman, M. J., DeLeon, I. G., & Wearden, J. H. (2011). Psychophysical assessment of timing in individuals with autism. *American Journal on Intellectual and Developmental Disabilities*, *116*(2), 165–178. https://doi.org/10.1352/1944-7558-116.2.165

Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: first- and second-order principles of subjective time. *Annual Review of Psychology*, *65*, 743–771. https://doi.org/10.1146/annurev-psych-010213-115117

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders, 5th edition. Arlington, VA: American Psychiatric Association.

Amos, P. (2013). Rhythm and timing in autism: learning to dance. *Frontiers in Integrative Neuroscience*, *7*(April), 27. https://doi.org/10.3389/fnint.2013.00027

Angrilli, a, Cherubini, P., Pavese, a, & Mantredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, *59*(6), 972–982. https://doi.org/10.3758/BF03205512

Anney, R. J. L. (2013). *Chapter 2.3 - Common Genetic Variants in Autism Spectrum Disorders*. *The Neuroscience of Autism Spectrum Disorders*. Elsevier. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-391924-3.00010-7

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate Vocalization, Gesture, and the Evolution of Human Language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

Artieda, J., Pastor, M. A., Lacruz, F., & Obeso, J. A. (1992). Temporal discrimination is abnormal in Parkinson's Disease. *Brain*, *115*, 199–210.

Asgari, M., Bayestehtashk, A., & Shafran, I. (2013). Robust and Accurate Features for Detecting and Diagnosing Autism Spectrum Disorders. *Interspeech*, (August), 191–194.

Asperger, H. (1944a). *Autistic psychopathy in childhood*. *Autism and Asperger syndrome*. https://doi.org/10.1017/CBO9780511526770.002

Asperger, H. (1944b). Die "Autistisehen Psychopathen" im Kindesalter. *Archiv Für Psychiatrie Und Nervenkrankheiten*, *117*(1), 76–136. https://doi.org/10.1007/BF01837709

Bachorowski, J.-A., & Owren, M. J. (1994). Vocal expression of emotion: Intensity and context. *Psychological Science*, *6*(4), 219–224.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, *6*(4), 429–451. https://doi.org/10.1016/S0163-6383(83)90241-2

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder- evidence from a British twin study, 63–77.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelright, S., & Drew, A. (2000). A Screening Instrument for Autism at 18 Months of Age: A 6-Year Follow-up Study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(6), 694–702.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet*, *368*(9531), 210–215. https://doi.org/10.1016/S0140-6736(06)69041-7

Bald, L., Berrien, F. K., Price, J. B., & Sprague, R. O. (1942). Errors in perceiving the temporal order of auditory and visual stimuli. *Journal of Applied Psychology*, *26*(3), 382.

Banse, R., & Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bares, M., Lungu, O., Liu, T., Waechter, T., Gomez, C. M., & Ashe, J. (2007). Impaired predictive motor timing in patients with cerebellar disorders. *Experimental Brain Research*, *180*(2), 355–365. https://doi.org/10.1007/s00221-007-0857-8

Bares, M., Lungu, O. V., Husarova, I., & Gescheidt, T. (2010). Predictive motor timing performance dissociates between early diseases of the cerebellum and parkinson's disease. *Cerebellum*, *9*(1), 124–135. https://doi.org/10.1007/s12311-009-0133-5

Baron-Cohen, S, Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind”? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, Simon, Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001a). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baron-Cohen, Simon, Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The Autism Spectrum Quotient : Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Bauman, M., & Kemper, T. L. (1985). Histoanatomic observations of the brain in early infantile autism. *Neurology*, *35*(6), 866.

Bauman, M L, & Kemper, T. L. (1990). Limbic and cerebellar abnormalities are also present in an autistic child of normal intelligence. *Neurology*, *40*(Suppl 1), 307.

Bauman, Margaret L, & Kemper, T. L. (1986). Developmental cerebellar abnormalities: a consistent finding in early infantile autism. *Neurology*, *36*(4), 190.

Bebko, J. M., Weiss, J. a., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(1), 88–98. https://doi.org/10.1111/j.1469-7610.2005.01443.x

Ben-Sasson, A., Hen, L., Fluss, R., Cermak, S. A., Engel-Yeger, B., & Gal, E. (2009). A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(1), 1–11. https://doi.org/10.1007/s10803-008-0593-3

Berlin, H. a, & Rolls, E. T. (2004). Time perception, impulsivity, emotionality, and personality in self-harming borderline personality disorder patients. *Journal of Personality Disorders*, *18*(4), 358–378. https://doi.org/10.1521/pedi.2004.18.4.358

Beudel, M., Renken, R., Leenders, K. L., & de Jong, B. M. (2009). Cerebral representations of space and time. *NeuroImage*, *44*(3), 1032–1040. https://doi.org/10.1016/j.neuroimage.2008.09.028

Beudel, Martijn, Galama, S., Leenders, K. L., & de Jong, B. M. (2008). Time estimation in Parkinson’s disease and degenerative cerebellar disease. *Neuroreport*, *19*(10), 1055–1058.

Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber’s law in time perception and production. *Behavioural Processes*, *71*(2–3), 201–210. https://doi.org/10.1016/j.beproc.2005.11.006

Block, R. a, & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, *4*(2), 184–197. https://doi.org/10.3758/BF03209393

Boddaert, N., Barthélémy, C., Poline, J. B., Samson, Y., Brunelle, F., & Zilbovicius, M. (2005). Autism: Functional brain mapping of exceptional calendar capacity. *British Journal of Psychiatry*, *187*(JULY), 83–86. https://doi.org/10.1192/bjp.187.1.83

Bone, D, Black, M., & Lee, C. (2012). Spontaneous-Speech Acoustic-Prosodic Features of

Children with Autism and the Interacting Psychologist. *Proc. InterSpeech …*, 3–6.

Bone, Daniel, Lee, C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (2014). The Psychologist as and Interlocutor in Autism Spectrum Disorder Assessment: Insights From a Study of Spontaneous Prosody. *Journal of Speech, Language, and Hearing Research*, *57*, 1162–1177. https://doi.org/10.1044/2014

Bone, Daniel, Lee, C., Chaspari, T., Black, M. P., Williams, M. E., Lee, S., … Narayanan, S. (2013). Acoustic-Prosodic , Turn-taking , and Language Cues in Child-Psychologist Interactions for Varying Social Demand. *Interspeech*, (August), 2400–2404.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*(January), 237. https://doi.org/10.3389/fnhum.2010.00237

Bottini, S. (2018). Social reward processing in individuals with autism spectrum disorder: A systematic review of the social motivation hypothesis. *Research in Autism Spectrum Disorders*, *45*(October 2017), 9–26. https://doi.org/10.1016/j.rasd.2017.10.001

Boucher, J. (2012). Putting theory of mind in its place: psychological explanations of the socio-emotional-communicative impairments in autistic spectrum disorder. *Autism*, *16*(3), 226–246. https://doi.org/10.1177/1362361311430403

Boucher, Jill. (2001). Lost in a sea of time: Time parsing and autism. In T. McCormack & C. Hoerl (Eds.), *Time and memory: issues in Philosophy and Psychology* (Oxford Uni). Oxford: Oxford University Press.

Boucher, Jill, Lewis, V., & Collis, G. (2000). Voice processing abilities in children with autism, children with specific language impairments and young typically developing children. https://doi.org/10.1111/1469-7610.00672

Boucher, Jill, Pons, F., Lind, S., & Williams, D. (2007). Temporal cognition in children with autistic spectrum disorders: Tests of diachronic thinking. *Journal of Autism and Developmental Disorders*, *37*(8), 1413–1429. https://doi.org/10.1007/s10803-006-0285-9

Boulter, C., Freeston, M., South, M., & Rodgers, J. (2014). Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(6), 1391–1402. https://doi.org/10.1007/s10803-013-2001-x

Bowler, D M, Gardiner, J. M., Grice, S., & Saavalainen, P. (2000). Memory illusions: false recall and recognition in adults with Asperger's syndrome. *Journal of Abnormal Psychology*, *109*(4), 663–672.

Bowler, Dermot M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger's syndrome. *Consciousness and Cognition*, *16*(1), 124–143. https://doi.org/10.1016/j.concog.2005.12.001

Bowler, Dermot M., Gardiner, J. M., Grice, S., & Saavalainen, P. (2000). Memory illusions: False recall and recognition in adults with Asperger's syndrome. *Journal of Abnormal Psychology*, *109*(4), 663–672. https://doi.org/10.1037//0021-843X.109.4.663

Bowler, Dermot M., Poirier, M., Martin, J. S., & Gaigg, S. B. (2016). Nonverbal short-term serial order memory in autism spectrum disorder. *Journal of Abnormal Psychology*, *125*(7), 886–893. https://doi.org/10.1037/abn0000203

Brennan, S. E., & Williams, M. (1995). The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1995.1017

Brennan, Susan E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brenner, L. A., Shih, V. H., Colich, N. L., Sugar, C. A., Bearden, C. E., & Dapretto, M. (2015). Time reproduction performance is associated with age and working memory in high-functioning youth with autism spectrum disorder. *Autism Research*, *8*(1), 29–37. https://doi.org/10.1002/aur.1401

Brett, D., Warnell, F., McConachie, H., & Parr, J. R. (2016). Factors Affecting Age at ASD Diagnosis in UK: No Evidence that Diagnosis Age has Decreased Between 2004 and 2014. *Journal of Autism and Developmental Disorders*, *46*(6), 1974–1984. https://doi.org/10.1007/s10803-016-2716-6

Brisson, J., Martel, K., Serres, J., Sirois, S., & Adrien, J.-L. (2014). Acoutsic analysis of oral productions of infants later diagnosed with autism and their mother. *Infant Mental Health Journal*, *35*(3), 285–295. https://doi.org/10.1002/imhj.

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *16*(12), 573–574. https://doi.org/10.1016/j.tics.2012.10.005

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Development and Psychopathology*, *14*(2), 209–224. https://doi.org/10.1017/S0954579402002018

Brodersen, Kay H, Daunizeau, J., Mathys, C., Chumbley, J. R., Buhmann, J. M., & Stephan, K. E. (2013). Variational Bayesian mixed-effects inference for classification studies. *Neuroimage*, *76*, 345–361.

Brodersen, Kay Henning, Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition* (pp. 3121–3124).

Brodeur, D. a., Gordon Green, C., Flores, H., & Burack, J. a. (2014). Time Estimation Among Low-Functioning Individuals With Autism Spectrum Disorders: Evidence of Poor Sensitivity to Variability of Short Durations. *Autism Research*, *7*(2), 237–244. https://doi.org/10.1002/aur.1364

Brondino, N., Fusar-Poli, L., Panisi, C., Damiani, S., Barale, F., & Politi, P. (2016). Pharmacological Modulation of GABA Function in Autism Spectrum Disorders: A Systematic Review of Human Studies. *Journal of Autism and Developmental Disorders*, *46*(3), 825–839. https://doi.org/10.1007/s10803-015-2619-y

Brosy, J., Bangerter, A., & Mayor, E. (2016). Disfluent Responses to Job Interview Questions and What They Entail. *Discourse Processes*, *53*(5–6), 371–391. https://doi.org/10.1080/0163853X.2016.1150769

Brown, J. F. (1931). On time perception in visual movement fields. *Psychologische Forschung*, *14*(1), 233–248.

Brown, S. W. (1997). Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118–1140. https://doi.org/10.3758/BF03205526

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(3), 169–175. https://doi.org/10.1002/mrdd.20036

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Buhusi, C. V., & Meck, W. H. (2009). Relativity theory and time perception: Single or multiple clocks? *PLoS ONE*, *4*(7). https://doi.org/10.1371/journal.pone.0006268

Buhusi, C. V, & Meck, W. H. (2002). Differential effects of methamphetamine and haloperidol on the control of an internal clock. *Behavioral Neuroscience*, *116*(2), 291–297. https://doi.org/10.1037/0735-7044.116.2.291

Buhusi, C. V, & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews. Neuroscience*, *6*(10), 755–765. https://doi.org/10.1038/nrn1764

Buonomano, D. V. (2000). Decoding Temporal Information: A Model Based on Short-Term Synaptic Plasticity. *Journal of Neuroscience*, *20*(3), 1129–1141.

Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, *125*(3), 409–434. https://doi.org/10.1037/rev0000091

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for internal clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 195–205.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In *Recent advances in the psychology of language* (pp. 347–360). Springer.

Buxhoeveden, D. P., Semendeferi, K., Buckwalter, J., Schenker, N., Switzer, R., & Courchesne, E. (2006). Reduced minicolumns in the frontal cortex of patients with autism. *Neuropathology and Applied Neurobiology*, *32*(5), 483–491. https://doi.org/10.1111/j.1365-2990.2006.00745.x

Camaioni, L., Perucchini, P., Muratori, F., & Milone, a. (1997). Brief report: a longitudinal examination of the communicative gestures deficit in young children with autism. *Journal of Autism and Developmental Disorders*, *27*(6), 715–725. https://doi.org/10.1023/A:1025858917000

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of pause duration. *Speech Prosody 2002. Proceedings of The1st International Conference on Speech Prosody*, 199–202. https://doi.org/10.1.1.12.561

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational Abilities Among Children with Autism and Children with Developmental Delays. *Autism*, *2*(4), 325–344. https://doi.org/10.1177/1362361398024002

Caron, M. J., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain*, *129*(7), 1789–1802. https://doi.org/10.1093/brain/awl072

Casanova, M. F., van Kooten, I. a J., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch,

H. W. M., … Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathologica*, *112*(3), 287–303. https://doi.org/10.1007/s00401-006-0085-5

Cassell, J. (1998). A Framework for Gesture Generation and Interpretation. *Computer Vision in Human-Machine Interaction*, 1–19. https://doi.org/CBO9780511569937.013

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*, 314–325. https://doi.org/10.1006/nimg.2000.0612

Cerasa, A., Hagberg, G. E., Peppe, A., Bianciardi, M., Gioia, M. C., Costa, A., … Sabatini, U. (2006). Functional changes in the activity of cerebellum and frontostriatal regions during externally and internally timed movement in Parkinson ' s disease, *71*, 259–269. https://doi.org/10.1016/j.brainresbull.2006.09.014

Chan, K. K. L., & To, C. K. S. (2016). Do Individuals with High-Functioning Autism Who Speak a Tone Language Show Intonation Deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(3), 619–627. https://doi.org/10.1017/S0033291710000991

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon Effect The Perception-Behavior Link and Social Interaction.pdf. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Cheng, R.-K., Tipples, J., Narayanan, N. S., & Meck, W. H. (2016). Clock Speed as a Window into Dopaminergic Control of Emotion and Time Perception. *Timing & Time Perception*, *4*(1), 99–122. https://doi.org/10.1163/22134468-00002064

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–238. https://doi.org/10.1016/j.tics.2012.02.007

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*(6), 871–887. https://doi.org/10.1016/j.pragma.2004.10.016

Church, R. M. (1984). Properties of the Internal Clock. *Annals of the New York Academy of Sciences*, *423*(1), 566–582.

Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*(2), 165.

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*(6), 1003–1011. https://doi.org/10.1111/desc.12193

Clark, C. R., & Geffen, G. M. (1990). Hemispheric Equivalence for Simultaneity Judgements of Somatosensory Stimuli. *Cognitive Neuropsychology*, *7*(4), 311–327. https://doi.org/10.1080/02643299008253446

Colgan, S. E., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., & Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *12*(4–

5), 307–319. https://doi.org/10.1080/09297040600701360

Condon, W. S., & Sander, L. W. (1974). Neonate Movement Is Synchronized with Adult Speech : Interactional Participation and Language Acquisition. *Science*, *183*(4120), 99–101.

Cook, J. L., Blakemore, S. J., & Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain*, *136*(9), 2816–2824. https://doi.org/10.1093/brain/awt208

Corballis, M. C. (2003). From mouth to hand: gesture, speech, and the evolution, *26*, 199–208.

Coull, Jennifer T, Cheng, R.-K., & Meck, W. H. (2010). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(1), 3–25. https://doi.org/10.1038/npp.2010.113

Coull, Jennifer T, Vidal, F., Nazarian, B., & Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science (New York, N.Y.)*, *303*(5663), 1506–1508. https://doi.org/10.1126/science.1091573

Coull, JT T., & Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology*, *18*(2), 137–144. https://doi.org/10.1016/j.conb.2008.07.011

Courchesne, E. (2004). Brain development in autism: Early overgrowth followed by premature arrest of growth. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(2), 106–111. https://doi.org/10.1002/mrdd.20020

Courchesne, E., Hesselink, J. R., Jernigan, T. L., & Yeung-Courchesne, R. (1987). Abnormal Neuroanatomy in a Nonretarded Person With Autism. *Archives of Neurology*, *44*(3), 335–341. https://doi.org/10.1016/j.virol.2015.08.001

Courchesne, E., Pierce, K., Schumann, C. M., Redcay, E., Buckwalter, J. a, Kennedy, D. P., & Morgan, J. (2007). Mapping early brain development in autism. *Neuron*, *56*(2), 399–413. https://doi.org/10.1016/j.neuron.2007.10.016

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A. J., … Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience*, *108*(5), 848–865. https://doi.org/10.1037/0735-7044.108.5.848

Couturier, J. L., Speechley, K. N., Steele, M., Norman, R., Stringer, B., & Nicolson, R. (2005). Parental perception of sleep problems in children of normal intelligence with pervasive developmental disorders: prevalence, severity, and pattern. *J Am. Acad. Child Adolesc. Psychiatry*. https://doi.org/S0890-8567(09)61743-3 [pii]\r10.1097/01.chi.0000166377.22651.87

Craig, A. D. (Bud). (2009). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1933–1942. https://doi.org/10.1098/rstb.2009.0008

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, *12*, 1191–1202. https://doi.org/10.2147/NDT.S104620

Crane, L., & Goddard, L. (2008). Episodic and semantic autobiographical memory in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(3), 498–506. https://doi.org/10.1007/s10803-007-0420-2

Crane, L., Goddard, L., & Pring, L. (2009). Specific and general autobiographical knowledge in adults with autism spectrum disorders: the role of personal goals. *Memory (Hove, England)*, *17*(5), 557–576. https://doi.org/10.1080/09658210902960211

Crespi, B., Leach, E., Dinsdale, N., Mokkonen, M., & Hurd, P. (2016). Imagination in human social cognition, autism, and psychotic-affective conditions. *Cognition*, *150*, 181–199. https://doi.org/10.1016/j.cognition.2016.02.001

Damarla, S. R., Keller, T. A., Kana, R. K., Cherkassky, V. L., Williams, D. L., Minshew, N. J., & Just, M. A. (2010). Cortical underconnectivity coupled with preserved visuospatial cognition in autism: Evidence from an fMRI study of an embedded figures task. *Autism Research*, *3*(5), 273–279. https://doi.org/10.1002/aur.153

Danckert, J., Ferber, S., Pun, C., Broderick, C., Striemer, C., Rock, S., & Stewart, D. (2007). Neglected time: impaired temporal perception of multisecond intervals in unilateral neglect. *Journal of Cognitive Neuroscience*, *19*(10), 1706–1720. https://doi.org/10.1162/jocn.2007.19.10.1706

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004a). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004b). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science*, *7*(3), 340–359. https://doi.org/10.1111/j.1467-7687.2004.00352.x

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424. https://doi.org/10.1207/s15326942dn2703_6

De Jonge-Hoekstra, L., Van der Steen, S., Van Geert, P., & Cox, R. F. A. (2016). Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding. *Frontiers in Psychology*, *7*(MAR), 1–19. https://doi.org/10.3389/fpsyg.2016.00473

de Marchena, A., & Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research*, *3*(6), 311–322. https://doi.org/10.1002/aur.159

de Ruiter, J.-P., Mitterer, H., & Enfield, N. J. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

DeGroot, T., & Motowidlo, S. J. (1999). Why Visual and Vocal Interview Cues Can Affect Interviewers ' Judgments and Predict Job Performance Why Visual and Vocal Interview Cues Can Affect Interviewers ' Judgments and Predict Job Performance. *Journal of Applied Psychology*, *84*(6), 986–993. https://doi.org/10.1037/0021-9010.84.6.986

DePape, A. M. R., Chen, A., Hall, G. B. C., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with Autism in relation to language ability. *Frontiers in Psychology*, *3*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2012.00072

Di Martino, A., Zuo, X.-N., Kelly, C., Grzadzinski, R., Mennes, M., Schvarcz, A., … Milham, M. P. (2013). Shared and Distinct Intrinsic Functional Network Centrality in Autism and Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *74*(8), 623–632. https://doi.org/10.1016/j.biopsych.2013.02.011.

Diehl, Joshua J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(03), 385. https://doi.org/10.1017/S0142716409090201

Diehl, Joshua John, & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics*, *34*, 135–161. https://doi.org/10.1017/S0142716411000646

Dimitris, X., Konvalinka, I., Bulbulia, J., & Roepstorff, A. (2011). Quantifying collective effervescence. Heart-rate dynamics at a fire-walking ritual. *Communicative & Integrative Biology*, *0889*, 735–738. https://doi.org/10.1073/pnas.1016955108

Dimitrova, N., Özçalışkan, Ş., & Adamson, L. B. (2017). Do Verbal Children with Autism Comprehend Gesture as Readily as Typically Developing Children? *Journal of Autism and Developmental Disorders*, *47*(10), 3267–3280. https://doi.org/10.1007/s10803-017-3243-9

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*(4), 478–484. https://doi.org/10.1016/0010-0285(79)90021-5

Droit-Volet, S., Brunot, S., & Niedenthal, P. M. (2004). Perception of the duration of emotional events. *Cognition and Emotion*, *18*(6), 849–858.

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our perception of time. *Trends in Cognitive Sciences*, *11*(12), 504–513. https://doi.org/10.1016/j.tics.2007.09.008

Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244–250. https://doi.org/10.1016/j.beproc.2006.09.012

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: evidence from effects of stimulus order and trial sequence. *Attention Perception Psychophysics*, *74*(8), 1819–1841. https://doi.org/10.3758/s13414-012-0362-4

Edgar, R. S., Green, E. W., Zhao, Y., Van Ooijen, G., Olmedo, M., Qin, X., … Reddy, A. B. (2012). Edinburgh Research Explorer Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, *485*(7399), 459–464.

https://doi.org/10.1038/nature11088.Peroxiredoxins

Eigsti, I.-M., & Shapiro, T. (2003). A systems neuroscience approach to autism: biological, cognitive, and clinical perspectives. *Mental Retardation and Developmental Disabilities Research Reviews*, *9*, 205–215. https://doi.org/10.1002/mrdd.10081

Eigsti, I. M., De Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. https://doi.org/10.1016/j.rasd.2010.09.001

Ellinghaus, R., Ulrich, R., & Bausenhart, K. M. (2018). Effects of stimulus order on comparative judgments across stimulus attributes and sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 7–12. https://doi.org/10.1037/xhp0000495

Engel, G. R., & Dougherty, W. G. (1971). Visual-auditory distance constancy. *Nature*, *234*(5327), 308.

Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine Learning for Medical Imaging. *Radiographics*, *March-Apri*, 505–515. https://doi.org/10.1148/rg.2017160130

Esposito, A., & Esposito, A. M. (2011). On speech and gestures synchrony. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6800 LNCS*, 252–272. https://doi.org/10.1007/978-3-642-25775-9_25

Esposito, A., Esposito, D., & Trojano, L. (2007). Gestures , Pauses , and Aphasia : Four Study Cases, 59–63.

Esposito, A., & Marinaro, M. (2007). What pauses can tell us about speech and gesture partnership. *Volume 18: Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, *18*, 45–57.

Esposito, A., McCullough, K. E., & Quek, F. (2001). Disfluencies in gesture: Gestural correlates to filled and unfilled speech pauses. *IEEE International Workshop on Cues in Communication “Cues 2001”* , (August), On CD-Rom.

Exner, S. (1874). Experimentelle Untersuchung der einfachsten psychischen Processe. *Pfl??Ger, Archiv f??R Die Gesammte Physiologie Des Menschen Und Der Thiere*, *8*(1), 526–537. https://doi.org/10.1007/BF01612312

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders*, *5*(1), 28. https://doi.org/10.1186/1866-1955-5-28

Falter, C. M., & Noreika, V. (2011). Interval timing deficits and abnormal cognitive development. *Frontiers in Integrative Neuroscience*, *5*(June), 26. https://doi.org/10.3389/fnint.2011.00026

Falter, C. M., Noreika, V., Wearden, J. H., & Bailey, A. J. (2012). More consistent, yet less sensitive: Interval timing in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2093–2107. https://doi.org/10.1080/17470218.2012.690770

Feldman, R. (2007a). Infant Biological Foundations Synchrony and Developmental Outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345.

https://doi.org/10.1111/j.1467-8721.2007.00532.x

Feldman, R. (2007b). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(3–4), 329–354. https://doi.org/10.1111/j.1469-7610.2006.01701.x

Feldstein, S., Konstantareas, M., Oxman, J., & Webster, C. D. (1982). The chronography of interactions with autistic speakers: An initial report. *Journal of Communication Disorders*, *15*(6), 451–460. https://doi.org/10.1016/j.cnd.2009.05.001

Fernell, E., & Gillberg, C. (2010). Autism spectrum disorder diagnoses in Stockholm preschoolers. *Research in Developmental Disabilities*, *31*(3), 680–685. https://doi.org/10.1016/j.ridd.2010.01.007

Fierro, B., Palermo, A., Puma, A., Francolini, M., Panetta, M. L., Daniele, O., & Brighina, F. (2007). Role of the cerebellum in time perception: a TMS study in normal subjects. *Journal of the Neurological Sciences*, *263*(1), 107–112.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical Prosody in Asperger Syndrome: Perceptual and Acoustic Measurements. *Journal of Autism and Developmental Disorders*, 1–10. https://doi.org/10.1007/s10803-014-2073-2

Filippopoulos, P. C., Hallworth, P., Lee, S., & Wearden, J. H. (2013). Interference between auditory and visual duration judgements suggests a common code for time. *Psychological Research*, *77*(6), 708–715. https://doi.org/10.1007/s00426-012-0464-6

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *32*(5), 771–782. https://doi.org/10.1111/j.1469-7610.1991.tb01901.x

Flagg, E. J., Cardy, J. E. O., Roberts, W., & Roberts, T. P. L. (2005). Language lateralization development in children with autism: Insights from the late field magnetoencephalogram. *Neuroscience Letters*, *386*(2), 82–87. https://doi.org/10.1016/j.neulet.2005.05.037

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591–598.

Fortin, C., & Rousseau, R. (1987). Time estimation as an index of processing demand in memory search. *Perception & Psychophysics*, *42*(4), 377–382. https://doi.org/10.3758/BF03203095

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, *203*(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4

Fournier, K. a., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism and Developmental Disorders*, *40*(10), 1227–1240. https://doi.org/10.1007/s10803-010-0981-3

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., & Griffiths, T. D. (2003). Absence of auditory "global interference" in autism. *Brain*, *126*(12), 2703–2709. https://doi.org/10.1093/brain/awg274

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, *35*, 1–36.

Frith, U. (1989). *Asperger and his syndrome*.

Frith, U, & Happé, F. (1994). Autism: beyond "theory of mind". *Cognition*, *50*(1–3), 115–132.

Frith, Uta. (2012). Why we need cognitive explanations of autism. *Quarterly Journal of Experimental Psychology*, *65*(11), 2073–2092. https://doi.org/10.1080/17470218.2012.697178

Frith, Uta, & Happé, F. (1994). Autism : beyond " theory of mind ," *50*, 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, a., Rees, G., Frith, C., & Tylen, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816

Fusaroli, R, Bang, D., & Weed, E. (2013). Non-Linear Analyses of Speech and Prosody in Asperger's Syndrome. International Meeting For Autism Research, San Sebastian, Spain.

Fusaroli, R, Lambrechts, A., Yarrow, K., Maras, K., & Gaigg, S. (2016). Conversational voice patterns in adult English speakers with ASD. *Paper Presented at the IMFAR 2016, Baltimore, United States*.

Fusaroli, Riccardo, Bjørndahl, J. S., Roepstorff, A., & Tylén, K. (2016). A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1297.

Fusaroli, Riccardo, Konvalinka, I., & Wallot, S. (2014). Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis. In *Translational recurrences* (pp. 137–155). Springer.

Fusaroli, Riccardo, Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2016). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, 1–24. https://doi.org/10.1002/aur.1678

Fusaroli, Riccardo, Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Fusaroli, Riccardo, & Tylén, K. (2015). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Gabriele, S., Sacco, R., & Persico, A. M. (2014). Blood serotonin levels in autism spectrum disorder: A systematic review and meta-analysis. *European Neuropsychopharmacology*, *24*(6), 919–929. https://doi.org/10.1016/j.euroneuro.2014.02.004

Gadow, K. D., Pinsonneault, J. K., Perlman, G., & Sadee, W. (2014). Association of dopamine gene variants, emotion dysregulation and ADHD in autism spectrum disorder.

Research in Developmental Disabilities, *35*(7), 1658–1665. https://doi.org/10.1016/j.ridd.2014.04.007

Gaigg, S. B. (2012). The Interplay between Emotion and Cognition in Autism Spectrum Disorder: Implications for Developmental Theory. *Frontiers in Integrative Neuroscience*, *6*(December), 113. https://doi.org/10.3389/fnint.2012.00113

Gaigg, S. B., Bowler, D. M., Ecker, C., Calvo-Merino, B., & Murphy, D. G. (2015). Episodic Recollection Difficulties in ASD Result from Atypical Relational Encoding: Behavioral and Neural Evidence. *Autism Research*, *8*(3), 317–327. https://doi.org/10.1002/aur.1448

Gaigg, S. B., Bowler, D. M., & Gardiner, J. M. (2013). Episodic but not semantic order memory difficulties in autism spectrum disorder: Evidence from the Historical Figures Task. *Memory*, (July), 37–41. https://doi.org/10.1080/09658211.2013.811256

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65. https://doi.org/10.1016/S1364-6613(99)01424-2

García-Pérez, R. M., Lee, A., & Hobson, R. P. (2007). On intersubjective engagement in autism: A controlled study of nonverbal aspects of conversation. *Journal of Autism and Developmental Disorders*, *37*(7), 1310–1322. https://doi.org/10.1007/s10803-006-0276-x

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: a first-person approach. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1413), 1351–1361. https://doi.org/10.1098/rstb.2001.0955

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Geurts, H. M., Verte, S., Oosterlaan, J., Roeyers, H., Sergeant, J. A., Verté, S., … Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(4), 836–854. https://doi.org/10.1111/j.1469-7610.2004.00276.x

Gibbon, J, Church, R., & Meck, W. (1984). Scalar timing in memory. *Of the New York Academy of …, 1984*. https://doi.org/10.1111/j.1749-6632.1984.tb23417

Gibbon, John. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Gil, S., Chambres, P., Hyvert, C., Fanget, M., & Droit-Volet, S. (2012). Children with Autism Spectrum Disorders Have "The Working Raw Material" for Time Perception. *PLoS ONE*, *7*(11), 1–10. https://doi.org/10.1371/journal.pone.0049116

Gil, S., & Droit-Volet, S. (2009). Time perception, depression and sadness. *Behavioural Processes*, *80*(2), 169–176. https://doi.org/10.1016/j.beproc.2008.11.012

Goldberg, W. A., Jarvis, K. L., Osann, K., Laulhere, T. M., Straub, C., Thomas, E., … Spence, M. A. (2005). Brief report: Early social communication behaviors in the younger siblings of children with autism. *Journal of Autism and Developmental Disorders*, *35*(5), 657–664. https://doi.org/10.1007/s10803-005-0009-6

Goldfarb, J. L., & Goldstone, S. (1963). Time judgment: A comparison of filled and unfilled durations. *Perceptual and Motor Skills*.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

Goldstone, S., & Lhamon, W. T. (1974). Studies of auditory-visual differences in human time judgment: I. Sounds are judged longer than lights. *Perceptual and Motor Skills*.

Gooch, C. M., Wiener, M., Wencil, E. B., & Coslett, H. B. (2010). Interval timing disruptions in subjects with cerebellar lesions. *Neuropsychologia*, *48*(4), 1022–1031.

Gowen, E., & Miall, R. C. (2005). Behavioural aspects of cerebellar function in adults with Asperger syndrome. *Cerebellum (London, England)*, *4*(4), 279–289. https://doi.org/10.1080/14734220500355332

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, *26*(2), 129–143.

Grammer, K., Honda, M., Juette, A., & Schmitt, A. (1999). Fuzziness of nonverbal courtship communication unblurred by motion energy detection. *Journal of Personality and Social Psychology*, *77*(3), 487–508.

Grandin, T. (1992). An inside view of autism. *High Functioning Individuals with Autism*, 105–126.

Grandin, T., & Scariano, M. M. (1986). *Emergence, labeled autistic.* (Arena Pres). Novato, CA: Warner Books.

Grant, D., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., & Baird, G. (2009). Impairment in movement skills of children with autistic spectrum disorders. *Developmental Medicine and Child Neurology*, *51*(4), 311–316. https://doi.org/10.1111/j.1469-8749.2008.03242.x

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult ... but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315. https://doi.org/10.1080/17549500903003060

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology*, *40*(3), 213–225. https://doi.org/10.1002/dev.10028

Grondin, S. (2005). Overloading temporal memory. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(5), 869–879. https://doi.org/10.1037/0096-1523.31.5.869

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, *72*(3), 561–582.

Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: cognitive, emotional, and timing substrates. *Psychological Review*, *113*(3), 483–525. https://doi.org/10.1037/0033-295X.113.3.483

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research : JSLHR*, *53*(3), 778–793. https://doi.org/10.1044/1092-4388(2009/08-0127)

Güth, W., Schmitttberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*(4), 367–388.

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to Spam filtering. *Expert Systems with Applications*, *36*(7), 10206–10222. https://doi.org/10.1016/j.eswa.2009.02.037

Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Experimental Brain Research*, *166*(3–4), 474–480. https://doi.org/10.1007/s00221-005-2387-6

Hairston, W. D., Hodges, D. a, Burdette, J. H., & Wallace, M. T. (2006). Auditory enhancement of visual temporal order judgment. *Neuroreport*, *17*(8), 791–795. https://doi.org/10.1097/01.wnr.0000220141.29413.b4

Hallmayer, J., Cleveland, S., Torres, a., Phillips, J., Cohen, B., Torigoe, T., … Risch, N. (2011). Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Archives of General Psychiatry*, *68*(11), 1095–1102. https://doi.org/10.1001/archgenpsychiatry.2011.76

Happé, F. (1999). Autism: cognitive deficit or cogntive style? *Trends in Cogntive Science*, *3*(6), 216–222. https://doi.org/10.1016/S1364-6613(99)01318-2

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Happé, F. G. E. (1994). Wechsler IQ Profile and Theory of Mind in Autism: A Research Note. *Journal of Child Psychology and Psychiatry*, *35*(8), 1461–1471. https://doi.org/10.1111/j.1469-7610.1994.tb01287.x

Happé, F. G. E. (1995). The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism Author (s): Francesca G . E . Happé Published by : Wiley on behalf of the Society for Research in Child Development Stable URL : http://www.jstor.org/stabl. *Child Development*, *66*(3), 843–855.

Happé, F. G. E. (1996). Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *37*(7), 873–877. https://doi.org/10.1111/j.1469-7610.1996.tb01483.x

Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, *15*(1), 1–12. https://doi.org/10.1111/j.2044-835X.1997.tb00721.x

Happé, F., & Ronald, A. (2008). The “fractionable autism triad”: A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review*, *18*(4), 287–304. https://doi.org/10.1007/s11065-008-9076-8

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220. https://doi.org/10.1038/nn1770

Harrington, D L, Haaland, K. Y., & Hermanowicz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology*, *12*(1), 3–12.

Harrington, Deborah L., Castillo, G. N., Greenberg, P. A., Song, D. D., Lessig, S., Lee, R. R., & Rao, S. M. (2011). Neurobehavioral mechanisms of temporal processing deficits in Parkinson’s disease. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0017461

Harrington, Deborah L., Lee, R. R., Boyd, L. A., Rapcsak, S. Z., & Knight, R. T. (2004). Does the representation of time depend on the cerebellum? Effect of cerebellar stroke. *Brain*, *127*(3), 561–574. https://doi.org/10.1093/brain/awh065

Hashimoto, T., Tayama, M., Murakawa, K., Yoshimoto, T., Miyazaki, M., Harada, M., & Kuroda, Y. (1995). Development of the Brain-Stem and Cerebellum in Autistic Patients. *Journal of Autism and Developmental Disorders*, *25*(1), 1–18. https://doi.org/10.1007/bf02178163

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion: Studies in emotion & social interaction*. *Current Directions in Psychological Science* (Vol. 2). Cambridge University Press.

Heaton, P., Pring, L., & Hermelin, B. (1999). A pseudo-savant: A case of exceptional musical splinter skills. *Neurocase: Case Studies in Neuropsychology, Neuropsychiatry, & Behavioural Neurology*, *5*(6), 503–509. https://doi.org/10.1080/13554799908402745

Heaton, Pamela, & Hermelin, B. (1998). Autism and Pitch Processing : A Precursor for Savant Musical Ability ? Author (s): Pamela Heaton , Beate Hermelin and Linda Pring Published by : University of California Press Stable URL : http://www.jstor.org/stable/40285769 REFERENCES Linked reference, *15*(3), 291–305.

Heeman, P. a, Lunsford, R., Selfridge, E., Black, L., & Santen, J. Van. (2010). Autism and Interactional Aspects of Dialogue. *Proceedings of SIGDIAL 2010: The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 249–252.

Heldner, M. (2001). Spectral Emphasis as an Additional Source of Information in Accent Detection. *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, 57–60.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, *97*(1), 35–61. https://doi.org/10.1037/0033-2909.97.1.35

Herbert, M. R. (2004). Neuroimaging in Disorders of Social and Emotional Functioning : What Is the Question ? *Journal of Child Neurology*, *19*, 772–784.

Hermelin, B. (2001). *Bright splinters of the mind: A personal story of research with autistic savants*. Jessica Kingsley Publishers.

Hicks, R. E., Miller, G. W., Gaes, G., & Beirman, K. (1977). Concurrent Processing Demands and the Experience of Time-in-Passing. *The American Journal of Psychology*, *90*(3), 431–446.

Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome : Patterns of performance in a multiple case series Executive processes in Asperger syndrome : Patterns of performance in a multiple case series. *Neuropsychologia*, *44*, 2822–2835. https://doi.org/10.1016/j.neuropsychologia.2006.06.007

Hirsh, I.-J., & Fraisse, P. (1964). Simultanéité et succession de stimuli hétérogènes. *L'Année Psychologique*, *64*(1), 1–19.

Hirsh, I. J., & Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*(5), 423.

Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British Journal of Psychology*, *79*(4), 441–453. https://doi.org/http://dx.doi.org/10.1111/j.2044-8295.1988.tb02745.x

Hodgson, A. R., Freeston, M. H., Honey, E., & Rodgers, J. (2017). Facing the Unknown: Intolerance of Uncertainty in Children with Autism Spectrum Disorder. *Journal of Applied Research in Intellectual Disabilities*, *30*(2), 336–344. https://doi.org/10.1111/jar.12245

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*(MAR), 1–12. https://doi.org/10.3389/fpsyg.2012.00074

Hollien, H. (2012). On pubescent voice change in males. *Journal of Voice*, *26*(2), e29–e40. https://doi.org/10.1016/j.jvoice.2011.01.007

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–961.

Hubbard, A. L., Mcnealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., & Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and Behavior*, *2*(5), 606–619. https://doi.org/10.1002/brb3.81

Hubbard, K., & Trauner, D. a. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173. https://doi.org/10.1007/s10936-006-9037-4

Hus, V., & Lord, C. (2014). The Autism Diagnostic Observation Schedule, Module 4: Revised Algorithm and Standardized Severity Scores. *Journal of Autism and Developmental Disorders*, 1–17. https://doi.org/10.1007/s10803-014-2080-3

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, *36*(1), 77–90. https://doi.org/10.1007/s10803-005-0044-3

Isabella, R. A., Belsky, J., & von Eye, A. (1989). Origins of Infant-Mother Attachment: An Examination of Interactional Synchrony During the Infant's First Year. *Developmental Psychology*, *25*(1), 12–21. https://doi.org/10.1037/0012-1649.25.1.12

Iverson, J. M., Northrup, J. B., Leezenbaum, N. B., Parladé, M. V., Koterba, E. A., & West, K. L. (2017). Early Gesture and Vocabulary Development in Infant Siblings of Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(1), 55–71. https://doi.org/10.1007/s10803-017-3297-8

Ivry, R.B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, *1*(2), 136–152. https://doi.org/10.1162/jocn.1989.1.2.136

Ivry, Richard B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*(7), 273–280. https://doi.org/10.1016/j.tics.2008.04.002

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of Dialogue in Infancy: Coordinated Timing in Development. Wiley. https://doi.org/10.1037/10522-163

Jahanshahi, M., Brown, R. G., & Marsden, C. D. (1992). The effect of withdrawal of dopaminergic medication on simple and choice reaction time and the use of advance information in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(12), 1168–1176. https://doi.org/10.1136/jnnp.55.12.1168

Jahanshahi, M., Brown, R. G., & Marsden, C. D. (1993). A comparative study of simple and choice reaction time in Parkinson's, Huntington's and cerebellar disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *56*(11), 1169–1177. https://doi.org/10.1136/jnnp.56.11.1169

Jarrold, C. (2003). A review of research into pretend play in autism. *Autism : The International Journal of Research and Practice*, *7*(4), 379–390. https://doi.org/10.1177/1362361303007004004

Jasnow, M., & Feldstein, S. (1986). Adult-Like Temporal Characteristics of Mother-Infant Vocal Interactions. *Child Development*, *57*(3), 754–761.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Johnson, C. P., & Myers, S. M. (2007). Identification and Evaluation of Children With Autism Spectrum Disorders. *Pediatrics*, *120*(5), 1183–1215. https://doi.org/10.1542/peds.2007-2361

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Anita, J. S., Tregay, J., … Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850–2858. https://doi.org/10.1016/j.neuropsychologia.2009.06.015

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., … Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850–2858. https://doi.org/10.1016/j.neuropsychologia.2009.06.015

Jones, C. R. G., Lambrechts, A., & Gaigg, S. B. (2017). Using Time Perception to Explore Implicit Sensitivity to Emotional Stimuli in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 2054–2066. https://doi.org/10.1007/s10803-017-3120-6

Jones, C. R. G., Rosenkranz, K., Rothwell, J. C., & Jahanshahi, M. (2004). The right dorsolateral prefrontal cortex is essential in time reproduction: an investigation with repetitive transcranial magnetic stimulation. *Experimental Brain Research*, *158*(3), 366–372.

Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., … Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, *11*(1), 95–109. https://doi.org/10.1002/aur.1873

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Juslin, P. N., Laukka, P., & Bänziger, T. (2018). The Mirror to Our Soul? Comparisons of Spontaneous and Posed Vocal Expression of Emotion. *Journal of Nonverbal Behavior*, *42*(1), 1–40. https://doi.org/10.1007/s10919-017-0268-x

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2004). Functional and Anatomical Cortical Underconnectivity in Autism: Evidence from an fMRI Study of an Ex, (2001).

Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K., & Varma, S. (2012). Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1292–1313. https://doi.org/10.1016/j.neubiorev.2012.02.007

Kakihara, Y., Takiguchi, T., Ariki, Y., & Nakai, Y. (2015). Classification of Children with Autism Spectrum and Typically Developing Children Using Pitch Features, *5*(1), 1–6. https://doi.org/10.5923/j.ajsp.20150501.01

Kaland, C., Swerts, M., & Krahmer, E. (2013). Accounting for the listener: comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *The Journal of the Acoustical Society of America*, *134*(3), 2182–2196. https://doi.org/10.1121/1.4816544

Kanner, L. (1943). Autistic disturbances of affective contact. *Acta Paedopsychiatrica*. https://doi.org/10.1105/tpc.11.5.949

Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., & Pellicano, E. (2016). Central tendency effects in time interval reproduction in autism. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep28570

Kargas, N., López, B., Reddy, V., & Morris, P. (2014). The Relationship Between Auditory Processing and Restricted, Repetitive Behaviors in Adults with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(3), 658–668. https://doi.org/10.1007/s10803-014-2219-2

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron*, *53*(3), 427–438. https://doi.org/10.1016/j.neuron.2007.01.006

Kendon, A. (1972). Some relationships between body motion and speech. *Studies in Dyadic Communication*, *7*(177), 90.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication*, *25*(1980), 207–227.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2015). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, 1–21. https://doi.org/10.1177/1362361315588200

Kerns, Connor M., & Kendall, P. C. (2012). The Presentation and Classification of Anxiety in Autism Spectrum Disorder. *Clinical Psychology: Science and Practice*, *19*(4), 323–347.

https://doi.org/10.1111/cpsp.12009

Kerns, Connor Morrow, Kendall, P. C., Berry, L., Souders, M. C., Franklin, M. E., Schultz, R. T., … Herrington, J. (2014). Traditional and atypical presentations of anxiety in youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(11), 2851–2861. https://doi.org/10.1007/s10803-014-2141-7

Killeen, P. R., & Weiss, N. a. (1987). Optimal timing and the Weber function. *Psychological Review*, *94*(4), 455–468. https://doi.org/10.1037/0033-295X.94.4.455

Kim, S. H., & Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research*, *3*(4), 162–173.

Kipp, M. (2001). Anvil-a generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*.

Kirsh, S. J., Crnic, K. A., & Greenberg, M. T. (1995). Relations between parent–child affect and synchrony and cognitive outcome at 5 years of age. *Personal Relationships*, *2*(3), 187–198. https://doi.org/10.1111/j.1475-6811.1995.tb00085.x

Kiss, G., van Santen, J., Prud'hommeaux, E., & Black, L. (2012). Quantitative Analysis of Pitch in Speech of Children with Neurodevelopmental Disorders. *13th Annual Conference of the International Speech Communication Association*. Retrieved from http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers/interspeech_2012/i12_1343.pdf

Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., … Devlin, B. (2012). Common genetic variants, acting additively, are a major source of risk for autism. *Molecular Autism*, *3*(1), 9. https://doi.org/10.1186/2040-2392-3-9

Klein, S. B., Chan, R. L., & Loftus, J. (1999). Independence of Episodic and Semantic Self-Knowledge: The Case from Autism. *Social Cognition*, *17*(4), 413–436. https://doi.org/10.1521/soco.1999.17.4.413

Klinger, L. G., & Dawson, G. (1995). A fresh look at categorization abilities in persons with autism. In *Learning and cognition in autism* (pp. 119–136). Springer.

Koch, G., Oliveri, M., Carlesimo, G. A., & Caltagirone, C. (2002). Selective deficit of time perception in a patient with right prefrontal cortex lesion. *Neurology*, *59*(10), 1658–1658. https://doi.org/10.1212/01.WNL.0000032504.45792.8F

Koch, Giacomo, Costa, A., Brusa, L., Peppe, A., Gatto, I., Torriero, S., … Caltagirone, C. (2008). Impaired reproduction of second but not millisecond time intervals in Parkinson's disease. *Neuropsychologia*, *46*(5), 1305–1313. https://doi.org/10.1016/j.neuropsychologia.2007.12.005

Koch, Giacomo, Oliveri, M., Torriero, S., Salerno, S., Gerfo, E. Lo, & Caltagirone, C. (2007). Repetitive TMS of cerebellum interferes with millisecond time processing. *Experimental Brain Research*, *179*(2), 291–299. https://doi.org/10.1007/s00221-006-0791-1

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kowalski, W. J. (1943). The effect of delay upon the duplication of short temporal intervals. *Journal of Experimental Psychology*, *33*(3), 239–246. https://doi.org/10.1037/h0054005

Krakowiak, P., Goodlin-Jones, B., Hertz-Picciotto, I., Croen, L. A., & Hansen, R. L. (2008). Sleep problems in children with autism spectrum disorders, developmental delays, and typical development: A population-based study. *Journal of Sleep Research*, *17*(2), 197–206. https://doi.org/10.1111/j.1365-2869.2008.00650.x

Kriete, T., & Noelle, D. C. (2015). Dopamine and the development of executive dysfunction in autism spectrum disorders. *PLoS ONE*, *10*(3), 1–19. https://doi.org/10.1371/journal.pone.0121605

Kubicek, L. F. (1980). Organization in two mother-infant interactions involving a normal infant and his fraternal twin brother who was later diagnosed as autistic. *High-Risk Infants and Children : Adult and Peer Interactions / Editor, Tiffany Martini Field, Co-Editors, Susan Goldberg, Daniel Stern, Anita Miller Sostek*. New York : Academic Press, 1980.

Kujala, T., Aho, E., Lepistö, T., Jansson-Verkasalo, E., Nieminen-von Wendt, T., von Wendt, L., & Näätänen, R. (2007). Atypical pattern of discriminating sound features in adults with Asperger syndrome as reflected by the mismatch negativity. *Biological Psychology*, *75*(1), 109–114. https://doi.org/10.1016/j.biopsycho.2006.12.007

Kunchulia, M., Tatishvili, T., Lomidze, N., Parkosadze, K., & Thomaschke, R. (2017). Time-based event expectancies in children with Autism spectrum disorder. *Experimental Brain Research*, *235*(9), 2877–2882. https://doi.org/10.1007/s00221-017-5024-2

Kupper, Z., Ramseyer, F., Hoffmann, H., Kalbermatten, S., & Tschacher, W. (2010). Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia. *Schizophrenia Research*, *121*(1–3), 90–100. https://doi.org/10.1016/j.schres.2010.03.032

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience*, *4*(January), 129. https://doi.org/10.3389/fnint.2010.00129

Lai, M. C., Lombardo, M. V., Chakrabarti, B., Sadek, S. A., Pasco, G., Wheelwright, S. J., … Suckling, J. (2010). A shift to randomness of brain oscillations in people with autism. *Biological Psychiatry*, *68*(12), 1092–1099. https://doi.org/10.1016/j.biopsych.2010.06.027

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review*, *18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

Lambrechts, A., Falter-Wagner, C. M., & van Wassenhove, V. (2017). Diminished neural resources allocation to time processing in Autism Spectrum Disorders. *NeuroImage: Clinical*, *17*(September 2017), 124–136. https://doi.org/10.1016/j.nicl.2017.09.023

Lambrechts, A., Mella, N., Pouthas, V., & Noulhiane, M. (2011). Subjectivity of Time

Perception: A Visual Emotional Orchestration. *Frontiers in Integrative Neuroscience*, *5*(November), 1–6. https://doi.org/10.3389/fnint.2011.00073

Lambrechts, A., Walsh, V., & Van Wassenhove, V. (2013). Evidence accumulation in the magnitude system. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0082122

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(6), 629–638. https://doi.org/10.1111/j.1469-7610.2006.01531.x

Large, E. W. (2008). *Resonating to musical rhythm: theory and experiment*. *Psychology of time*. https://doi.org/10.1016/B978-0-08046-977-5.00006-5

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., … Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, *32*(4), 195–214. https://doi.org/10.1007/s10919-008-0055-9

Lee, K.-H., Egleston, P. N., Brown, W. H., Gregory, A. N., Barker, A. T., & Woodruff, P. W. R. (2007). The role of the cerebellum in subsecond time perception: evidence from repetitive transcranial magnetic stimulation. *Journal of Cognitive Neuroscience*, *19*(1), 147–157.

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and Repetitive Behaviors in Autism Spectrum Disorders : A Review of Research in the Last Decade. *Psychological Bulletin*, *137*(4), 562–593. https://doi.org/10.1037/a0023341

Lejeune, H. (1998). Switching or gating? The attentional challenge in cognitive models of psychological time. *Behavioural Processes*, *44*(2), 127–145. https://doi.org/10.1016/S0376-6357(98)00045-X

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471. https://doi.org/10.1080/01690965.2010.500218

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., Näätänen, R., … N????t??nen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1–2), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology*, *117*(10), 2161–2171. https://doi.org/10.1016/j.clinph.2006.06.709

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Lever, A. G., Richard Ridderinkhof, K., Marsman, M., & Geurts, H. M. (2017). Reactive and proactive interference control in adults with autism spectrum disorder across the lifespan. *Developmental Psychology*, *53*(2), 379–395. https://doi.org/10.1037/dev0000219

Lewis, P. a., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: Evidence from neuroimaging. *Current Opinion in Neurobiology*, *13*(2), 250–255. https://doi.org/10.1016/S0959-4388(03)00036-9

Lewis, P. A., & Miall, R. C. (2009a). The precision of temporal judgement: milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1897–1905. https://doi.org/10.1098/rstb.2009.0020

Lewis, P. A., & Miall, R. C. (2009b). The precision of temporal judgement: milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1897–1905. https://doi.org/10.1098/rstb.2009.0020

Lind, S. E. (2010). Memory and the self in autism: A review and theoretical framework. *Autism*, *14*(5), 430–456. https://doi.org/10.1177/1362361309358700

Lind, S. E., & Bowler, D. M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology*, *119*(4), 896–905. https://doi.org/10.1037/a0020631

Lindell, A. K., & Hudry, K. (2013). Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychology Review*, *23*(3), 257–270. https://doi.org/10.1007/s11065-013-9234-5

Long, G., & Mo, S. (1970). Interval estimation as a function of constant and interrupted stimulation. *Psychonomic Science*, *21*(2), 87–88.

Lord, C, Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). *Autism Diagnostic Observation Schedule (ADOS)*. *Journal of Autism and Developmental Disorders* (Vol. 30). Western Psychological Services.

Lord, Catherine, Rutter, M. L., Dilavore, P. C., & Risi, S. (2002). Autism diagnostic observation schedule-generic (ADOS-G) manual. *Los Angeles, CA: Western Psychological Services*.

Lord, Catherine, Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/BF02172145

Loveland, K. a, Landry, S. H., Hughes, S. O., Hall, S. K., & McEvoy, R. E. (1988). Speech acts and the pragmatic deficits of autism. *Journal of Speech and Hearing Research*, *31*(4), 593–604.

Maister, L., & Plaisted-Grant, K. C. (2011). Time perception and its relationship to memory Autism Spectrum Conditions. *Developmental Science*, *14*(6), 1311–1322. https://doi.org/10.1111/j.1467-7687.2011.01077.x

Malapani, C., Levy, R., Meck, W. H., Deweer, B., Dubois, B., & Gibbon, J. (1998). Coupled Temporal Memories in Parkinson ' s Disease : A Dopamine-Related Dysfunction, 316–331.

Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. In T. Givon & B. F. Malle (Eds.), *The evolution of language out of prelanguage* (pp. 265–284). John Benjamins Publishing Company. Retrieved from http://cogprints.org/3317/

Malloch, S. N. (2000). Mothers and infants and communicative musicality. *Musicae Scientiae*, *3*, 29–57. https://doi.org/10.1177/10298649000030S104

Malow, B. A., Katz, T., Reynolds, A. M., Shui, A., Carno, M., Connolly, H. V., … Bennett, A. E.

(2016). Sleep Difficulties and Medications in Children With Autism Spectrum Disorders: A Registry Study. *Pediatrics*, *137*(Supplement), S98–S104. https://doi.org/10.1542/peds.2015-2851H

Maras, K. L., & Bowler, D. M. (2012). Eyewitness Testimony in Autism Spectrum Disorder: A Review. *Journal of Autism and Developmental Disorders*, *44*(11), 1–16. https://doi.org/10.1007/s10803-012-1502-3

Maras, K. L., Memon, A., Lambrechts, A., & Bowler, D. M. (2013). Recall of a live and personally experienced eyewitness event by adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(8), 1798–1810. https://doi.org/10.1007/s10803-012-1729-z

Marchi, E., Baron-cohen, S., Golan, O., Sven, B., Arora, P., & Reinhold, H. (2015). Typicality and Emotion in the Voice of Children with Autism Spectrum Condition : Evidence Across Three Languages Machine Intelligence & Signal Processing Group , MMK , Technische Universit ¨ Chair of Complex & Intelligent Systems , University of Passau , , 115–119.

Martin, J. S., Poirier, M., & Bowler, D. M. (2010). Brief report: Impaired temporal reproduction performance in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(5), 640–646. https://doi.org/10.1007/s10803-009-0904-3

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5–6), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Matthews, N., Welch, L., Achtman, R., Fenton, R., & FitzGerald, B. (2016). Simultaneity and temporal order judgments exhibit distinct reaction times and training effects. *PLoS ONE*, *11*(1), 1–17. https://doi.org/10.1371/journal.pone.0145926

Matthews, W. J., & Meck, W. H. (2014). Time perception: The bad news and the good. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(4), 429–446. https://doi.org/10.1002/wcs.1298

Maximo, J. O., Cadena, E. J., & Kana, R. K. (2014). The implications of brain connectivity in the neuropsychology of autism. *Neuropsychology Review*, *24*(1), 16–31. https://doi.org/10.1007/s11065-014-9250-0

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *42*(6), 682–702. https://doi.org/10.1080/13682820601170102

McCracken, J. T., McGough, J., Shah, B., Cronin, P., Hong, D., Aman, M. G., … others. (2002). Risperidone in children with autism and serious behavioral problems. *New England Journal of Medicine*, *347*(5), 314–321.

McDougle, C., Naylor, S. T., Cohen, D. J., Aghajanian, G. K., Heninger, G. R., & Price, L. H.

(1996). Effects of tryptophan depletion in drug-free adults with autistic disorder. *Archives of General Psychiatry*, *53*(11), 993–1000.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

McNeill, D. (2005). Gesture, gaze, and ground. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3869 LNCS*(1962), 1–14. https://doi.org/10.1007/11677482_1

McNeill, D., Bertenthal, B., Cole, J., & Gallagher, S. (2005). Gesture-first, but no gestures? *Behavioral and Brain Sciences*, *28*(2), 138–139. https://doi.org/10.1017/S0140525X05360031

Meck, W. H. (1983a). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(2), 171–201.

Meck, W. H. (1983b). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(2), 171–201.

Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research*, *3*(3–4), 227–242. https://doi.org/10.1016/0926-6410(96)00009-2

Meck, W. H. (2005). Neuropsychology of timing and time perception. *Brain and Cognition*, *58*(1), 1–8. https://doi.org/10.1016/j.bandc.2004.09.004

Meck, W. H. (2006). Temporal memory in mature and aged rats is sensitive to choline acetyltransferase inhibition. *Brain Research*, *1108*(1), 168–175. https://doi.org/10.1016/j.brainres.2006.06.047

Meck, W. H., & Hunter, W. S. (1984). Attentional bias between modalities: effect on the internal clock, memory, and decision stages used in animal time discrimination. *Annals of the New York Academy of Sciences*, *423*, 528–541. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=6588813&retmode=ref&cmd=prlinks

Meck, W. H., & N'Diaye, K. (2005). Un modèle neurobiologique de la perception et de l'estimation du temps. *Psychologie Francaise*, *50*(1), 47–63. https://doi.org/10.1016/j.psfr.2004.10.009

Meck, W. H., Penney, T. B., & Pouthas, V. (2008). Cortico-striatal representation of time in animals and humans. *Current Opinion in Neurobiology*, *18*(2), 145–152. https://doi.org/10.1016/j.conb.2008.08.002

Merchant, H., Luciana, M., Hooper, C., Majestic, S., & Tuite, P. (2008). Interval timing and Parkinson's disease: Heterogeneity in temporal performance. *Experimental Brain Research*, *184*(2), 233–248. https://doi.org/10.1007/s00221-007-1097-7

Meyer, B. J., Gardiner, J. M., & Bowler, D. M. (2014). Directed forgetting in high-functioning adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(10), 2514–2524. https://doi.org/10.1007/s10803-014-2121-y

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jefferies, H., & Plaisted, K. (2002). Visual motion coherence thresholds in people with autism. *Journal of Child Psychology and Psychiatry*, *2*(2), 255–263. https://doi.org/10.1111/1469-7610.00018

Ming, X., Brimacombe, M., & Wagner, G. C. (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain and Development*, *29*(9), 565–570. https://doi.org/10.1016/j.braindev.2007.03.002

Minter, M., Hobson, R., & Bishop, M. (1998). Congenital visual impairment and "theory of mind." *British Journal of Developmental Psychology*, *16*, 183–196. https://doi.org/10.1111/j.2044-835X.1998.tb00918.x

Mitchell, P., & Ropar, D. (2004). Visuo-spatial abilities in autism: A review. *Infant and Child Development*, *13*(3), 185–198. https://doi.org/10.1002/icd.348

Mitchell, W. L. (2015). *Communication profile in young adults with autism*. University of Alberta.

Mitchell, W. L., & Volden, J. (2015). Quality of communication in young adults with autism during a simulated employment interview. (Poster, IMFAR 2015).

Monk, C. S., Peltier, S. J., Wiggins, J. L., Weng, S. J., Carrasco, M., Risi, S., & Lord, C. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders,. *NeuroImage*, *47*(2), 764–772. https://doi.org/10.1016/j.neuroimage.2009.04.069

Monk, C. S., Weng, S.-J., Wiggins, J. L., Kurapati, N., Louro, H. M. C., Carrasco, M., … Lord, C. (2010). Neural circuitry of emotional face processing in autism spectrum disorders. *Journal of Psychiatry & Neuroscience : JPN*, *35*(2), 105–114. https://doi.org/10.1503/jpn.090085

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*. https://doi.org/10.1016/S0926-6410(03)00089-2

Morett, L. M., O'Hearn, K., Luna, B., & Ghuman, A. S. (2016). Altered Gesture and Speech Production in ASD Detract from In-Person Communicative Quality. *Journal of Autism and Developmental Disorders*, *46*(3), 998–1012. https://doi.org/10.1007/s10803-015-2645-9

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615–622. https://doi.org/10.1037/0278-7393.18.3.615

Mostofsky, S H, Goldberg, M. C., Landa, R. J., & Denckla, M. B. (2000). Evidence for a deficit in procedural learning in children and adolescents with autism: implications for cerebellar contribution. *Journal of the International Neuropsychological Society : JINS*, *6*(7), 752–759. https://doi.org/10.1017/S1355617700677020

Mostofsky, Stewart H., Powell, S. K., Simmonds, D. J., Goldberg, M. C., Caffo, B., & Pekar, J. J. (2009). Decreased connectivity and cerebellar activity in autism during motor task performance. *Brain*, *132*(9), 2413–2425. https://doi.org/10.1093/brain/awp088

Mostofsky, Stewart H, Dubey, P., Jerath, V. K., Jansiewicz, E. M., Goldberg, M. C., & Denckla, M. B. (2006). Developmental dyspraxia is not limited to imitation in children with autism spectrum disorders. *Journal of the International Neuropsychological Society : JINS*, *12*(3), 314–326. https://doi.org/10.1017/S1355617706060437

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128. https://doi.org/10.1007/bf02206861

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Murphy, E. R., Foss-Feig, J., Kenworthy, L., Gaillard, W. D., & Vaidya, C. J. (2012). Atypical Functional Connectivity of the Amygdala in Childhood Autism Spectrum Disorders during Spontaneous Attention to Eye-Gaze. *Autism Research and Treatment*, *2012*, 1–12. https://doi.org/10.1155/2012/652408

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–1108.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother-infant communications. *Journal of Child Language*, *13*(1), 15–29. https://doi.org/10.1017/S0305000900000271

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 378–383. https://doi.org/Doi 10.1001/Archpedi.161.4.378

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Nair, A., Treiber, J. M., Shukla, D. K., Shih, P., & Müller, R. A. (2013). Impaired thalamocortical connectivity in autism spectrum disorder: A study of functional and anatomical connectivity. *Brain*, *136*(6), 1942–1955. https://doi.org/10.1093/brain/awt079

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain & Development*, *36*(6), 516–522. https://doi.org/10.1016/j.braindev.2013.07.006

Nakano, T., Kato, N., & Kitazawa, S. (2011). Lack of eyeblink entrainments in autism spectrum disorders. *Neuropsychologia*, *49*(9), 2784–2790. https://doi.org/10.1016/j.neuropsychologia.2011.06.007

Nakano, T., & Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech.

Experimental Brain Research, *205*(4), 577–581. https://doi.org/10.1007/s00221-010-2387-z

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Ni Chuileann, S., & Quigley, J. (2013). Assessing recollection and familiarity in low functioning autism. *Journal of Autism and Developmental Disorders*, *43*(6), 1406–1422. https://doi.org/10.1007/s10803-012-1697-3

Nicholas, B. (2015). How Could Circadian Clock Genes Influence Short Duration Timing? *Time Distortions in Mind*, 356.

Nicholas, B., Rudrasingham, V., Nash, S., Kirov, G., Owen, M. J., & Wimpory, D. C. (2007). Association of Per1 and Npas2 with autistic disorder: support for the clock genes/social timing hypothesis. *Molecular Psychiatry*, *12*(6), 581–592. https://doi.org/10.1038/sj.mp.4001953

Noonan, S. K., Haist, F., & Müller, R. A. (2009). Aberrant functional connectivity in autism: Evidence from low-frequency BOLD signal fluctuations. *Brain Research*, *1262*, 48–63. https://doi.org/10.1016/j.brainres.2008.12.076

Noulhiane, M., Mella, N., Samson, S., Ragot, R., & Pouthas, V. (2007). How emotional auditory stimuli modulate time perception. *Emotion (Washington, D.C.)*, *7*(4), 697–704. https://doi.org/10.1037/1528-3542.7.4.697

O'Hearn, K., Lakusta, L., Schroer, E., Minshew, N., & Luna, B. (2011). Deficits in adults with autism spectrum disorders when processing multiple objects in dynamic scenes. *Autism Research*, *4*(2), 132–142. https://doi.org/10.1002/aur.179

O'Reilly, J. X., Mesulam, M. M., & Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *The Journal of Neuroscience*, *28*(9), 2252–2260. https://doi.org/10.1523/JNEUROSCI.2742-07.2008

Oller, D. K., Niyogi, P., Gray, S., Richards, J. a, Gilkerson, J., Xu, D., … Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13354–13359. https://doi.org/10.1073/pnas.1003882107

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621.Do

Ozonoff, S, Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, *32*(7), 1081–1105. https://doi.org/10.1111/j.1469-7610.1991.tb00351.x

Ozonoff, Sally, Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., … others. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(3), 256–266.

Ozonoff, Sally, Young, G. S., Goldring, S., Greiss-Hess, L., Herrera, A. M., Steele, J., … Rogers, S. J. (2008). Gross motor development, movement abnormalities, and early identification of autism. *Journal of Autism and Developmental Disorders*, *38*(4), 644–656. https://doi.org/10.1007/s10803-007-0430-0

Parish-Morris, J., Chevallier, C., Tonge, N., Letzen, J., Pandey, J., & Schultz, R. T. (2013).

Visual attention to dynamic faces and objects is linked to face processing skills: A combined study of children with autism and controls. *Frontiers in Psychology*, *4*(APR), 1–7. https://doi.org/10.3389/fpsyg.2013.00185

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Peeters, T., & Gillberg, C. (1999). *Autism: Medical and educational aspects*. John Wiley & Sons.

Pellicano, E. (2010). Individual Differences in Executive Function and Central Coherence Predict Developmental Changes in Theory of Mind in Autism, *46*(2), 530–544. https://doi.org/10.1037/a0018287

Pellicano, E., & Burr, D. (2012a). Response to Brock: Noise and autism. *Trends in Cognitive Sciences*, *16*(12), 574–575. https://doi.org/10.1016/j.tics.2012.10.004

Pellicano, E., & Burr, D. (2012b). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Peppé, S., Cleland, J., Gibbon, F. E., O'Hare, A., Castilla, P. M., Peppé, S., … Martinez-Castilla, P. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics*, *24*(1), 41–53. https://doi.org/10.1016/j.jneuroling.2010.07.005

Peppé, S. J. E. (2009). Why is prosody in speech-language pathology so difficult. *International Journal of Speech-Language Pathology*, *11*(4), 258–271. https://doi.org/10.1080/17549500902906339

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research : JSLHR*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Perbal-Hatif, S. (2012). Clinical research. *A Neuropsychological Approach to Time Estimation*, *14*, 425–432.

Perner, J. (2000). Memory and theory of mind. *The Oxford Handbook of Memory*, 297–312.

Perner, J., Kloo, D., & Gornik, E. (2007). Episodic memory development: Theory of mind is part of re-experiencing experienced events. *Infant and Child Development*, *16*(5), 471–490. https://doi.org/10.1002/icd.517

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215). https://doi.org/10.1126/science.1256846

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02). https://doi.org/10.1017/S0140525X04000056

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., … Betancur, C. (2010).

Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–372. https://doi.org/10.1038/nature09146

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(5), 765–775. https://doi.org/10.1111/1469-7610.00375

Poirier, M., Martin, J. S., Gaigg, S. B., & Bowler, D. M. (2011). Short-term memory in autism spectrum disorder. *Journal of Abnormal Psychology*, *120*(1), 247–252. https://doi.org/10.1037/a0022298

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*(3–4), 599–611. https://doi.org/10.1016/j.wocn.2003.08.001

Pouthas, V., & Perbal, S. (2004). Time perception depends on accurate clock mechanisms as well as unimpaired attention and memory processes. *Acta Neurobiologiae Experimentalis*, *64*(3), 367–385.

Praamstra, P., & Pope, P. (2007). Slow brain potential and oscillatory EEG manifestations of impaired temporal preparation in Parkinson's disease. *Journal of Neurophysiology*, *98*(August 2007), 2848–2857. https://doi.org/10.1152/jn.00224.2007

Presmanes, A. G., Walden, T. A., Stone, W. L., & Yoder, P. J. (2007). Effects of different attentional cues on responding to joint attention in younger siblings of children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(1), 133–144. https://doi.org/10.1007/s10803-006-0338-0

Quigley, J., McNally, S., & Lawson, S. (2016). Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months. *Language Learning and Development*, *12*(3), 295–310.

Ramenzoni, V. C., Riley, M. A., Shockley, K., & Baker, A. A. (2012). Interpersonal and intrapersonal coordinative modes for joint and single task performance. *Human Movement Science*, *31*(5), 1253–1267. https://doi.org/10.1016/j.humov.2011.12.004

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *52*(3), 273–286. https://doi.org/10.1080/713932708

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Redcay, E., & Courchesne, E. (2005). When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biological Psychiatry*, *58*(1), 1–9. https://doi.org/10.1016/j.biopsych.2005.03.026

Repp, B. H. (2005). Sensorimotor synchronization : A review of the tapping literature, *12*(6), 969–992.

Richards, W. (1964). Time Estimates Measured By Reproduction. *Perceptual and Motor Skills*, *18*, 929–943. https://doi.org/10.2466/pms.1964.18.3.929

Richardson, D. C., & Dale, R. (2005). Looking To Understand : The Coupling Between Speakers ' and Listeners ' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*, 1045–1060.

Richter, N., & Mehlhorn, G. (2006). Focus on contrast and emphasis: Evidence from prosody. *The Architecture of Focus. Berlin: Mouton de Gruyter*, 347–371.

Ring, M., Gaigg, S. B., & Bowler, D. M. (2016). Relational Memory Processes in Adults with Autism Spectrum Disorder. *Autism Research*, *9*(1), 97–106. https://doi.org/10.1002/aur.1493

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354. https://doi.org/10.1080/0163853X.2011.558002

Rodier, P. M., Ingram, J. L., Tisdale, B., Nelson, S., & Romano, J. (1996). Embryological origin for autism: Developmental anomalies of the cranial nerve motor nuclei. *Journal of Comparative Neurology*, *370*(2), 247–261. https://doi.org/10.1002/(SICI)1096-9861(19960624)370:2<247::AID-CNE8>3.0.CO;2-2

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575.

Roitman, J. D., Brannon, E. M., Andrews, J. R., & Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychologica*, *124*(3), 296–318. https://doi.org/10.1016/j.actpsy.2006.03.008

Rothwell, C. D. (2018). *Recurrence quantification models of human conversational grounding processes: Informing natural language human-computer interaction*. Wright State University.

Rubia, K., Halari, R., Christakou, A., & Taylor, E. (2009). Impulsiveness as a timing disturbance: neurocognitive abnormalities in attention-deficit hyperactivity disorder during temporal processes and normalization with methylphenidate. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1919–1931. https://doi.org/10.1098/rstb.2009.0014

Rumsey, J M, & Hamburger, S. D. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. *Journal of Clinical and Experimental Neuropsychology*, *10*(2), 201–221. https://doi.org/10.1080/01688638808408236

Rumsey, Judith M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders*, *15*(1), 23–36.

Russell, J., & Jarrold, C. (1999). Memory for actions in children with autism: self versus other. *Cognitive Neuropsychiatry*, *4*(4), 303–331. https://doi.org/10.1080/135468099395855

Russell, P. A., Hosie, J. A., Gray, C. D., Scott, C., & Hunter, N. (1998). The development of theory of mind in deaf people. *Journal of Child Psychology and Psychiatry*, *39*(6), 903–910. https://doi.org/10.1017/S0021963098002844

Rutherford, M. D., & Rogers, S. J. (2003). Cognitive underpinnings of pretend play in autism. *Journal of Autism and Developmental Disorders*, *33*(3), 289–302. https://doi.org/10.1023/A:1024406601334

Rutschmann, J., & Link, R. (1964). Perception of temproal order of stimuli differing in sense mode and simple reaction time. *Perceptual and Motor Skills*, *18*, 345–352.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Salazar, F., Baird, G., Chandler, S., Tseng, E., O'sullivan, T., Howlin, P., … Simonoff, E. (2015). Co-occurring Psychiatric Disorders in Preschool and Elementary School-Aged Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(8), 2283–2294. https://doi.org/10.1007/s10803-015-2361-5

Santos, F., Brosh, N., Falk, T. H., Zwaigenbaum, L., Bryson, S. E., Roberts, W., … Brian, J. a. (2013). Very early detection of Autism Spectrum Disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. *ICASSP*, 7567–7571.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Schain, R. J., & Freedman, D. X. (1961). Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *The Journal of Pediatrics*, *58*(3), 315–320. https://doi.org/10.1016/S0022-3476(61)80261-8

Scharfstein, L. a., Beidel, D. C., Sims, V. K., & Rendon Finnell, L. (2011). Social skills deficits and vocal characteristics of children with social phobia or asperger's disorder: A comparative study. *Journal of Abnormal Child Psychology*, *39*(6), 865–875. https://doi.org/10.1007/s10802-011-9498-2

Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. *Journal of Psycholinguistic Research*, *14*(4), 409–425. https://doi.org/10.1007/BF01067884

Schipul, S. E., Williams, D. L., Keller, T. a., Minshew, N. J., & Just, M. A. (2012). Distinctive neural processes during learning in autism. *Cerebral Cortex*, *22*(4), 937–950. https://doi.org/10.1093/cercor/bhr162

Schlegglof, E. A. (2000). Overlapping Talk and the Organization of Turn-Taking for Conversation. *Language in Society*, *29*(1), 1–63.

Schminder, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution. *European Research Journal of Methods for the Behavioral and Social Sciences*, *6*(4), 147–151.

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417. https://doi.org/10.1016/j.cognition.2013.08.004

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, *162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

Schöner, G. (2002). Timing, Clocks, and Dynamical Systems. *Brain and Cognition*, *48*(1), 31–51. https://doi.org/10.1006/brcg.2001.1302

Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, *23*(2-3 SPEC. ISS.), 125–141. https://doi.org/10.1016/j.ijdevneu.2004.12.012

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., & Skudlarski,

P. (2003). The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1430), 415–427. https://doi.org/10.1098/rstb.2002.1208

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology*, *87*(1), 146–151.

Selkirk, E. O. (1986). Phonology and syntax: The relation between sound and structure. Cambridge: The Massachusetts Institute of Technology Press, 1984. _. On derived domains in sentence phonology. *Phonology Yearbook*, (3), 371–405.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Shah, A, & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, *24*(4), 613–620. https://doi.org/10.1111/1469-7610.ep11422546.

Shah, Amitta, & Frith, U. (1993). Why do autistic individual show superior performance on the block design task? *J Child Psychol Psychiat*, *34*(8), 1351–1364.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(December), 2000. https://doi.org/10.1038/35048669

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., … Singh, N. C. (2010). Sounds of melody-Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/10.1016/j.neulet.2010.04.066

Shen, M. D., Nordahl, C. W., Young, G. S., Wootton-Gorges, S. L., Lee, A., Liston, S. E., … Amaral, D. G. (2013). Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain*, *136*(9), 2825–2835. https://doi.org/10.1093/brain/awt166

Sheridan, J., & McAuley, J. D. (1997). Rhythm as a cognitive skill: temporal processing deficits in autism. *Proceedings of the 4th {A}ustralasian Cognitive Science Conference*, 1–8.

Shih, P., Keehn, B., Oram, J. K., Leyden, K. M., Keown, C. L., & Müller, R. A. (2011). Functional differentiation of posterior superior temporal sulcus in autism: A functional connectivity magnetic resonance imaging study. *Biological Psychiatry*, *70*(3), 270–277. https://doi.org/10.1016/j.biopsych.2011.03.040

Shih, P., Shen, M., Öttl, B., Keehn, B., Gaffrey, M. S., & Müller, R. A. (2010). Atypical network connectivity for imitation in autism spectrum disorder. *Neuropsychologia*, *48*(10), 2931–2939. https://doi.org/10.1016/j.neuropsychologia.2010.05.035

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual Interpersonal Postural Constraints Are Involved in Cooperative Conversation. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(2), 326–332.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, a M., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research : JSLHR*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–393. https://doi.org/10.1016/j.cognition.2010.01.002

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The Voice Conveys Specific Emotions: Evidence From Vocal Burst Displays. *Emotion*, *9*(6), 838–846. https://doi.org/10.1037/a0017810

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(8), 921–929. https://doi.org/10.1097/CHI.0b013e318179964f

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(8), 813–821. https://doi.org/10.1111/j.1469-7610.2007.01766.x

Smith, J. G., Harper, D. N., Gittings, D., & Abernethy, D. (2007). The effect of Parkinson's disease on time estimation as a function of stimulus duration range and modality. *Brain and Cognition*, *64*(2), 130–143. https://doi.org/10.1016/j.bandc.2007.01.005

Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, *32*, 25–38.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665–681. https://doi.org/10.1080/01690965.2011.573220

Sodian, B., & Thoermer, C. (2008). Precursors to a theory of mind in infancy: Perspectives for research on autism. *Quarterly Journal of Experimental Psychology*, *61*(1), 27–39. https://doi.org/10.1080/17470210701508681

Solomon, M., Ozonoff, S. J., Ursu, S., Ravizza, S., Cummings, N., Ly, S., & Carter, C. S. (2009). The neural substrates of cognitive control deficits in autism spectrum disorders. *Neuropsychologia*, *47*(12), 2515–2526. https://doi.org/10.1016/j.neuropsychologia.2009.04.019

Souchay, C., Wojcik, D. Z., Williams, H. L., Crathern, S., & Clarke, P. (2013). Recollection in adolescents with autism spectrum disorder. *Cortex*, *49*(6), 1598–1609. https://doi.org/10.1016/j.cortex.2012.07.011

Souders, M. C., Mason, T. B. A., Valladares, O., Bucan, M., Levy, S. E., Mandell, D. S., … Pinto-Martin, J. (2009). Sleep behaviors and sleep quality in children with autism spectrum disorders. *Sleep (Rochester)*.

Spelke, E. S., & Cortelyou, A. (1981). Perceptual aspects of social knowing: Looking and listening in infancy. *Infant Social Cognition*, 61–84.

Spencer, J., Brien, J. O., Girges, C., Hill, H., & Johnston, A. (2007). Visual perception of facial biological motion in adults with ASD, *14*(2003), 2007.

Spencer, R. M. C., Zelaznik, H. N., Diedrichsen, J., & Ivry, R. B. (2003). Disrupted Timing of Discontinuous But Not Continuous Movements by Cerebellar Lesions. *Science*, *300*(5624), 1437–1439. https://doi.org/10.1126/science.1083661

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language*, *17*(1&2), 3–23. https://doi.org/10.1111/1468-0017.00186

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A {Twin} {Study} of {Autism} in {Denmark}, {Finland}, {Iceland}, {Norway} and {Sweden}. *Journal of Child Psychology and Psychiatry*, *30*(3), 405–416. https://doi.org/10.1111/j.1469-7610.1989.tb00254.x

Steiner, A. M., Goldsmith, T. R., Snow, A. V., & Chawarska, K. (2012). Disorders in Infants and Toddlers. *J Autism Dev Disord*, *42*(6), 1183–1196. https://doi.org/10.1007/s10803-011-1376-9.Practitioner

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Evidence for Diminished Multisensory Integration in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(12), 3161–3167. https://doi.org/10.1007/s10803-014-2179-6

Stewart, M. E., Griffiths, T. D., & Grube, M. (2015). Autistic Traits and Enhanced Perceptual Representation of Pitch and Time. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-015-2517-3

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Stone, J. V, Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., … Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings. Biological Sciences / The Royal Society*, *268*(1462), 31–38. https://doi.org/10.1098/rspb.2000.1326

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Stuss, D. T., Alexander, M. P., Shallice, T., Picton, T. W., Binns, M. A., Macdonald, R., … Katz, D. I. (2005). Multiple frontal systems controlling response speed. *Neuropsychologia*, *43*(3), 396–417. https://doi.org/10.1016/j.neuropsychologia.2004.06.010

Szelag, E., Kowalska, J., Galkowski, T., & Pöppel, E. (2004). Temporal processing deficits in high-functioning children with autism. *British Journal of Psychology (London, England : 1953)*, *95*(Pt 3), 269–282. https://doi.org/10.1348/0007126041528167

Tantam, D., Holmes, D., & Cordess, C. (1993). Nonverbal expression in autism of Asperger type. *Journal of Autism and Developmental Disorders*, *23*(1), 111–133. https://doi.org/10.1007/BF01066422

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(23), 13982–13987. https://doi.org/10.1073/pnas.95.23.13982

Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the naratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism & Developmental Disorders*, *23*(2), 309–322.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of Acoustical Society of America*, *134*(1), 628–639.

Tipples, J. (2008). Negative emotionality influences the effects of emotion on time perception. *Emotion (Washington, D.C.)*, *8*(1), 127–131. https://doi.org/10.1037/1528-3542.8.1.127

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15*(4), 434–444. https://doi.org/10.1111/j.1532-7078.2009.00025.x

Treffert, D. A. (1999). The Savant Syndrome and Autistic Disorder. *CNS Spectrums*, *4*(12), 57–60. https://doi.org/10.1017/S1092852900006830

Treffert, D. A. (2009). The savant syndrome: an extraordinary condition. A synopsis: past, present, future. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *364*(1522), 1351–1357. https://doi.org/10.1098/rstb.2008.0326

Treffner, P., Peter, M., & Kleidon, M. (2008). *Gestures and Phases: The Dynamics of Speech-Hand Communication*. *Ecological Psychology* (Vol. 20). https://doi.org/10.1080/10407410701766643

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychological Monographs*, *77*(13), 1–31. https://doi.org/10.1037/h0093864

Treisman, M. (1984). Temporal rhythms and cerebral rhythms. *Annals of the New York Academy of Sciences*, *423*(1), 542–565.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(1), 3–48. https://doi.org/10.1111/1469-7610.00701

Trevarthen, C., & Daniel, S. (2005). Disorganized rhythm and synchrony: Early signs of autism and Rett syndrome. *Brain and Development*, *27*(SUPPL. 1). https://doi.org/10.1016/j.braindev.2005.03.016

Trivino, M., Correa, A., Arnedo, M., & Lupianez, J. (2010). Temporal orienting deficit after prefrontal damage. *Brain*, *133*(4), 1173–1185. https://doi.org/10.1093/brain/awp346

Tulving, E. (1983). Elements of Episodic Memory. *Canadian Psychology*. https://doi.org/http://dx.doi.org/10.1017/S0140525X0004440X

Tunçgenç, B., & Cohen, E. (2018). Interpersonal movement synchrony facilitates pro-social behavior in children's peer-play. *Developmental Science*, *21*(1), 1–9. https://doi.org/10.1111/desc.12505

Uddin, L. Q., Supekar, K., Lynch, C. J., Khouzam, A., Phillips, J., Feinstein, C., … Menon, V. (2013). Salience Network–Based Classification and Prediction of Symptom Severity in Children With Autism. *JAMA Psychiatry*, *70*(8), 869. https://doi.org/10.1001/jamapsychiatry.2013.104

Vallesi, A., McIntosh, A. R., Shallice, T., & Stuss, D. T. (2009). When time shapes behavior: fMRI evidence of brain correlates of temporal monitoring. *Journal of Cognitive Neuroscience*, *21*(6), 1116–1126. https://doi.org/10.1162/jocn.2009.21098

Vallesi, A., Shallice, T., & Walsh, V. (2007). Role of the prefrontal cortex in the foreperiod effect: TMS evidence for dual mechanisms in temporal preparation. *Cerebral Cortex*, *17*(2), 466–474. https://doi.org/10.1093/cercor/bhj163

Van de Cruys, S., Evers, K., Hallen, R. Van Der, Eylen, L. Van, Boets, B., & Wagemans, J.

(2014). Precise minds in uncertain worlds : Predictive coding in autism. *Psychological Review*, *In press*, 1–36. https://doi.org/10.1037/a0037665

Van Rijn, H., Gu, B.-M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. In *Neurobiology of interval timing* (pp. 75–99). New York: Springer.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, *185*(3), 521–529. https://doi.org/10.1007/s00221-007-1168-9

Vesper, C., Van Der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, *211*(3–4), 517–530. https://doi.org/10.1007/s00221-011-2706-z

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*. (H. Laupp, Ed.), *Text*.

Villalobos, M. E., Mizuno, A., Dahl, B. C., Kemmotsu, N., & Müller, R. A. (2005). Reduced functional connectivity between V1 and inferior frontal cortex associated with visuomotor performance in autism. *NeuroImage*, *25*(3), 916–925. https://doi.org/10.1016/j.neuroimage.2004.12.022

Wachsmuth, I. (1999). Communicative rhythm in gesture and speech. *Gesture-Based Communication in Human-Computer Interaction*, 277–289.

Wallace, G., & Happe, F. (2008). Time perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(3), 447–455. https://doi.org/10.1016/j.rasd.2007.09.005

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Warlaumont, A. S., Oller, D. K., Buder, E. H., Dale, R., & Kozma, R. (2010). Data-driven automated acoustic analysis of human infant vocalizations using neural network tools. *The Journal of the Acoustical Society of America*, *127*(4), 2563–2577. https://doi.org/10.1121/1.3327460

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*, *22*(1), 25–39. https://doi.org/10.1044/1058-0360(2012/11-0145)

Wearden, J. H., & McShane, B. (1988). Interval Production as an Analogue of the Peak Procedure: Evidence for Similarity of Human and Animal Timing Processes. *The Quarterly Journal of Experimental Psychology Section B*, *40*(4), 363–375. https://doi.org/10.1080/14640748808402330

Wearden, J. H., Smith-Spark, J. H., Cousins, R., Edelstyn, N. M. J., Cody, F. W. J., & O'Boyle, D. J. (2008). Stimulus timing by people with Parkinson's disease. *Brain and Cognition*, *67*(3), 264–279. https://doi.org/10.1016/j.bandc.2008.01.010

Wearden, J H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*(2), 134.

Wearden, J H, Edwards, H., Fakhri, M., & Percival, A. (1998a). Why 'Sounds Are Judged

Longer Than Lights’: Application of a Model of the Internal Clock in Humans. *The Quarterly Journal of Experimental Psychology*, *51B*(2), 97–120.

Wearden, J H, Edwards, H., Fakhri, M., & Percival, A. (1998b). Why “Sounds Are Judged Longer Than Lights”’: Application of a Model of the Internal Clock in Humans. *The Quarterly Journal of Experimental Psychology*, *51B*(2), 97–120.

Wearden, John H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology*, *61*(4), 569–587. https://doi.org/10.1080/17470210701282576

Welchew, D. E., Ashwin, C., Berkouk, K., Salvador, R., Suckling, J., Baron-Cohen, S., & Bullmore, E. (2005). Functional disconnectivity of the medial temporal lobe in Asperger’s syndrome. *Biological Psychiatry*, *57*(9), 991–998. https://doi.org/10.1016/j.biopsych.2005.01.028

Wetherby, a M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders*, *34*(5), 473–493.

Wichmann, F. a, & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Wichmann, F. a, & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329. https://doi.org/10.3758/BF03194545

Wigham, S., Rodgers, J., South, M., McConachie, H., & Freeston, M. (2015). The Interplay Between Sensory Processing Abnormalities, Intolerance of Uncertainty, Anxiety and Restricted and Repetitive Behaviours in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(4), 943–952. https://doi.org/10.1007/s10803-014-2248-x

Wiltshire, A. (2007). Synchrony as the Underlying Structure of Gesture : The Relationship Between Speech Sound and Body Movement at the Micro Level. *Methodology*, 235–253.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraircing function of wrong bekfs in young children’s understanding of deception. *Cognition*, *13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wimpory, D. C., Hobson, R. P., Williams, J. M. G., & Nash, S. (2000). Are infants with autism socially engaged? A study of recent retrospective parental reports. *Journal of Autism and Developmental Disorders*, *30*(6), 525–536. https://doi.org/10.1023/A:1005683209438

Wimpory, D, Chadwick, P., & Nash, S. (1995). Brief report: musical interaction therapy for children with autism: an evaluative case study with two-year follow-up. *Journal of Autism and Developmental Disorders*, *25*(5), 541–552. https://doi.org/10.1007/BF02178299

Wimpory, Dawn. (2015). A social timing model of autism, informed by typical development. In A. Vatakis & M. J. Allman (Eds.), *Time Distortions in Mind: Temporal Processing in Clinical Populations* (pp. 57–92). Brill.

Wimpory, Dawn, Nicholas, B., & Nash, S. (2002). Social timing, clock genes and autism: A new hypothesis. *Journal of Intellectual Disability Research*, *46*(4), 352–358. https://doi.org/10.1046/j.1365-2788.2002.00423.x

Wing, A. M., & Kristofferson, A. B. (1973). The timing of interresponse intervals. *Perception & Psychophysics*, *13*(3), 455–460. https://doi.org/10.3758/BF03205802

Wing, L. (1996). Autistic spectrum disorders. *BMJ (Clinical Research Ed.)*, *312*(7027), 327–328.

Wing, Lorna. (1981). Asperger ’ s syndrome : a clinical account, 115–129.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). Embedded figures test, children’s embedded figures test, group embedded figures test: manual. Palo Alto, CA: Consulting Psychologists Press.

Wittmann, M., & Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends in Cognitive Sciences*, *12*(1), 7–12. https://doi.org/10.1016/j.tics.2007.10.004

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger Syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, *13*(6), 473–499. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14959500

Xingjian, S., Zhourong, C., Hao, W., & Yeung, D.-Y. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)* (pp. 1–9). https://doi.org/[]

Yang, B., Chan, R. C. K., Zou, X., Jing, J., Mai, J., & Li, J. (2007). Time perception deficit in children with ADHD. *Brain Research*, *1170*(SUPPL.: COMPLETE), 90–96. https://doi.org/10.1016/j.brainres.2007.07.021

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals - ProQuest. *Psychological Bulletin*, *124*(3), 283–307. Retrieved from http://search.proquest.com/docview/614329854?pq-origsite=360link&accountid=14749

Yoder, P., Stone, W. L., Walden, T., & Malesa, E. (2009). Predicting social impairment and ASD diagnosis in younger siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(10), 1381–1391. https://doi.org/10.1007/s10803-009-0753-0

Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using Extreme Learning Machine and financial expertise. *Neurocomputing*, *128*, 296–302. https://doi.org/10.1016/j.neucom.2013.01.063

Zakay, D., & Block, R. A. (1996). The role of attention in time estimation processes. *Advances in Psychology*, *115*, 143–164.

Zampini, M. (2005). Audio – visual simultaneity judgments. *Design*, *67*(3), 531–544.

Zimbardo, P. G., & Boyd, J. N. (1999). Putting Time in Perspective.pdf. *Journal of Personality and Social Psychology*, *77*(6), 1271–1288.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix 1: Violations of assumptions and data transformations applied (Chapter 2)

A.1.1. Point of Subjective Equality

Shapiro-Wilk tests of normality indicated the following violations of normality: In the auditory modality, PSE values for short durations in the ASD group showed a leptokurtic distribution with a small positive skew ($D(18)=.787$, $p<.005$). In the visual modality, PSE values for short durations in the TD group showed a small positive skew ($D(19)=.894$, $p<.05$), but visual inspection of P-P plots suggested that this was a very minor violation. In the audiovisual modality, PSE values in the ASD group showed a leptokuric distribution with a positive skew for short durations ($D(19)=.811$, $p<.005$), and a small positive skew for long durations ($D(19)=.885$, $p<.05$). In addition, Levene's test of homogeneity revealed that variance was not equal between groups in the auditory modality for long durations ($F(1, 36)=5.858$, $p<.05$) and in the audio-visual modality for short durations ($F(1,38)=4.341$, $p<.05$). To correct this, a log transformation was applied to PSE scores in the auditory and audiovisual modalies.

A.1.2. Weber Ratio

Shapiro-Wilk tests of normality indicated the following violations of normality: in the auditory modality, WR for short durations in the TD group showed a positive skew ($D(20)=.858$, $p<.01$).

In the visual modality, WR for short durations in the TD group showed a positive skew ($D(19)=.544$, $p<.001$) and WR for long durations in the ASD group showed a leptokurtic distribution ($D(13)=.817$, $p<.05$) although visual inspection of P-P plots indicated that the latter was a minor violation. In the audio-visual modality, WR values in the TD group showed a positive skew for both short durations ($D(21)=.842$, $p<.005$) and long durations ($D(21)=.885$, $p<.05$). WR values in the ASD group showed a positive skew for long durations only ($D(19) = .844$, $p<.01$). Levene's test indicated no violation of the homogeneity of variance in any of the conditions ($ps>.05$).

To correct for skewness, a log transformation was applied to WR scores in the auditory and visual modalities, and an inverse transformation was applied in the audiovisual modality.

A.1.3. Reaction Times

Shapiro–Wilk tests of normality indicated the following violations of normality: in the auditory modality, reaction times in the TD group showed a leptokurtic distribution and a positive skew for 0.9 ($D(20)=.822$, $p<.005$), 0.95 ($D(20)=.765$, $p<.001$), 1.05 ($D(20)=.760$, $p<.001$), 1.1 ($D(20)=.804$, $p<.005$) times the standard in the short durations range, and for 0.95 ($D(20)=.814$, $p<.005$), 1.05 ($D(20)=.821$, $p<.005$) and 1.1 ($D(20)=.719$, $p<.001$) times the standard in the long durations range. Reaction times in the ASD group showed a leptokurtic distribution and a positive skew for 1.25 times the standard in the short durations range ($D(18)=.863$, $p<.05$).

In the visual modality, reaction times in the TD group presented with a leptokurtic distribution and a positive skew for 0.75 ($D(19)=.802$, $p<.005$), 0.9 ($D(19)=.829$, $p<.005$), 0.95 ($D(19)=.583$, $p<.001$), 1.25 ($D(19)=.846$, $p<.01$) and 1.5 ($D(19)=.851$, $p<.01$) times the standard in the short durations range, and for 0.75 ($D(19)=.722$, $p<.001$), 0.9 ($D(19)=.697$, $p<.001$), 0.95 ($D(19)=.749$, $p<.001$), 1.05 ($D(19)=.581$, $p<.001$), 1.1 ($D(19)=.700$, $p<.001$), 1.25 ($D(19)=.696$, $p<.001$) and 1.5 ($D(19)=.842$, $p<.01$) times the standard in the long durations range. Reaction times in the ASD group showed a leptokurtic distribution and a positive skew for 1.25 times the standard in the short durations range ($D(13)=.868$, $p<.05$), and for 1.05 ($D(13)=.866$, $p<.05$) and 1.25 ($D(13)=.866$, $p<.05$) times the standard in the long durations range.

In the audio-visual modality, reaction times in the TD group showed a leptokurtic distribution and a positive skew for 0.9 ($D(21)=.688$, $p<.001$), 0.95 ($D(21)=.887$, $p<.05$) and 1.1 ($D(21)=.809$, $p<.005$) times the standard in the short durations range, and for 0.9 ($D(21)=.849$, $p<.005$), 1.05 ($D(21)=.746$, $p<.001$), 1.1 ($D(21)=.803$, $p<.005$) and 1.25 ($D(21)=.785$, $p<.001$) in the long durations range.

Levene's test for equality of variances indicated homogeneity of variances across all conditions and modalities for RTs.

A log transformation was applied to RT scores in all modalities.

Appendix 2: Syntax parsing (Chapter 4)

A.2.1. Constituency parse – sample sentences

ASD	TD
"She took me into the room and she'd said there had been a car crash." (ROOT (S (S (NP (PRP She)) (VP (VBD took) (NP (PRP me)) (PP (IN into) (NP (DT the) (NN room))))) (CC and) (S (NP (PRP she)) (VP (MD ’d) (VP (VBD said) (SBAR (S (NP (EX there)) (VP (VBD had) (VP (VBN been) (NP (DT a) (NN car) (NN crash)))))))) (. .)))	"Anna said that there had been an accident, showed me a picture on the wall." (ROOT (S (-LRB- [) (NP (UH yes)) (-RRB-]) (NP (NNP Anna)) (VP (VP (VBD said) (SBAR (IN that) (S (NP (EX there)) (VP (VBD had) (VP (VBN been) (NP (DT an) (NN accident))))))) (, ,) (VP (VBD showed) (NP (PRP me)) (NP (NP (DT a) (NN picture)) (PP (IN on) (NP (DT the) (NN wall)))))) (. .)))
"And then we put him in a foil blanket to keep him warm just in case he was going into shock" (ROOT (S (CC And) (ADVP (RB then)) (NP (PRP we)) (VP (VBD put) (NP (PRP him)) (PP (IN in) (NP (DT a) (NN foil) (NN blanket))) (S (VP (TO to) (VP (VB keep) (S (NP (PRP him)) (ADJP (JJ warm)) (SBAR (ADVP (RB just)) (IN in) (NN case) (S (NP (PRP he)) (VP (VBD was) (VP (VBG going) (PP (IN into) (NP (NN shock))))))))))))) (. .)))	"Then we put a piece of sponge or something between his arm and his chest and created an arm sling" (ROOT (S (ADVP (RB Then)) (NP (PRP we)) (VP (VP (VBD put) (NP (NP (DT a) (NN piece)) (PP (IN of) (NP (NN sponge) (CC or) (NN something))) (PP (IN between) (NP (NP (PRP$ his) (NN arm)) (CC and) (NP (PRP$ his) (NN chest)))))) (CC and) (VP (VBD created) (NP (DT an) (NN arm) (NN sling)))) (. .)))

"Um so she just said pass me um the cling film and she just wrapped that round so I helped her, passed her the cling film and then she wrapped it round and I just cut it off to make it easy for her, the cling film so we covered that, put the shirt back over."

```
(ROOT
 (S
  (INTJ (UH Um))
  (ADVP (RB so))
  (NP (PRP she))
  (ADVP (RB just))
  (VP (VBD said)
   (VP (VB pass)
    (NP (PRP me))
    (INTJ (UH um))
    (NP
     (NP (DT the))
      (NML
       (S
        (VP (VB cling)
         (S
          (UCP
           (NP (NN film))
           (CC and)
           (S
            (S
             (NP (PRP she))
             (ADVP (RB just))
             (VP (VBD wrapped)
              (SBAR (IN that)
               (S
                (NP
                 (NP (NN round) (RB so))
                 (SBAR
                  (S
                   (NP (PRP I))
                   (VP (VBD helped)
                    (NP (PRP her))))))
                (, ,)
                (VP (VBD passed)
                 (SBAR
                  (S
                   (NP (PRP her) (DT the))
                   (VP (VBP cling)
                    (NP (NN film)))))))))))
            (CC and)
            (S
             (ADVP (RB then))
             (NP (PRP she))
             (VP (VBD wrapped)
              (S
               (NP (PRP it))
               (ADJP (JJ round)))))
            (CC and)
            (S
             (NP (PRP I))
             (ADVP (RB just))
             (VP (VBD cut)
              (NP (PRP it))
              (PRT (RP off))
              (S
               (VP (TO to)
                (VP (VB make)
                 (S
                  (NP (PRP it))
```

"She got the bandage, and put the tea towel down, put the bandage on and I rolled up the edges of the bandages and then we wrapped it round and I tied the knot on the back of his hand"

```
(ROOT
 (S
  (S
   (NP (PRP She))
   (VP
    (VP (VBD got)
     (NP (DT the) (NN bandage)))
    (, ,)
    (CC and)
    (VP (VBD put)
     (NP (DT the) (NN tea) (NN towel))
     (ADVP (RB down)))
    (, ,)
    (VP (VBD put)
     (NP (DT the) (NN bandage))
     (PP (IN on)))))
  (CC and)
  (S
   (S
    (NP (PRP I))
    (VP (VBD rolled)
     (PRT (RP up))
     (NP
      (NP (DT the) (NNS edges))
      (PP (IN of)
       (NP (DT the) (NNS bandages))))))
   (CC and)
   (S
    (ADVP (RB then))
    (NP (PRP we))
    (VP (VBD wrapped)
     (S
      (NP (PRP it))
      (ADJP (JJ round)))))
   (CC and)
   (S
    (NP (PRP I))
    (VP (VBD tied)
     (NP (DT the) (NN knot))
     (PP (IN on)
      (NP
       (NP (DT the) (NN back))
       (PP (IN of)
        (NP (PRP$ his) (NN hand))))))))
  (. .)))
```

(ADJP (JJ easy)))))) (PP (IN for) (NP (PRP her))))))) (, ,) (NP (DT the)) (VP (VBP cling)))))) (NN film)) (SBAR (IN so) (S (NP (PRP we)) (VP (VBD covered) (SBAR (IN that) (S (, ,) (VP (VBD put) (NP (DT the) (NN shirt)) (ADVP (RB back) (RB over)))))))))))) (. .)))	
"I think she actually passed me the bandage and I unwrapped it up and asked me to put that on the cut so I put that square on the cut on the foot." (ROOT (S (S (NP (PRP I)) (VP (VBP think) (SBAR (S (NP (PRP she)) (ADVP (RB actually)) (VP (VBD passed) (S (NP (PRP me)) (NP (DT the) (NN bandage)))))))) (CC and) (S (NP (PRP I)) (VP (VP (VBD unwrapped) (NP (PRP it)) (PRT (RP up))) (CC and) (VP (VBD asked) (NP (PRP me)) (S (VP (TO to) (VP (VB put) (NP (DT that)) (PP (IN on) (NP (DT the) (NN cut))))))))) (ADVP (RB so)) (S (NP (PRP I)) (VP (VBD put) (NP (DT that) (NN square)) (PP (IN on) (NP (DT the) (NN cut))) (PP (IN on) (NP (DT the) (NN foot))))) (. .)))	"I think we went down to his legs after that and saw his left leg was broken um Anna asked me to pass her yeah I think I passed it to her, which we put under his um left leg um and then we covered him in a foil blanket to keep him warm." (ROOT (S (NP (PRP I)) (VP (VBP think) (SBAR (S (NP (PRP we)) (VP (VP (VBD went) (PRT (RP down)) (PP (IN to) (NP (PRP$ his) (NNS legs))) (PP (IN after) (NP (DT that)))) (CC and) (VP (VBD saw) (SBAR (S (NP (PRP$ his) (JJ left) (NN leg)) (VP (VBD was) (VP (VBN broken) (S (INTJ (UH um)) (NP (NNP Anna)) (VP (VBD asked) (NP (PRP me)) (S (VP (TO to) (VP (VB pass) (NP (PRP$ her) (NML (S (INTJ (UH yeah)) (NP (PRP I)) (VP (VBP think) (SBAR (S (NP (PRP I)) (VP (VBD passed) (NP (PRP it)) (PP (IN to) (NP (PRP her))) (, ,)

	(SBAR (WHNP (WDT which)) (S (S (NP (PRP we)) (VP (VBD put) (PP (IN under) (NP (PRP$ his) (NML (INTJ (UH um))) (JJ left) (NN leg))) (INTJ (UH um)))) (CC and) (S (ADVP (RB then)) (NP (PRP we)) (VP (VBD covered) (NP (PRP him)) (PP (IN in) (NP (DT a) (NN foil))))))))))))) (NN blanket)) (S (VP (TO to) (VP (VB keep) (S (NP (PRP him)) (ADJP (JJ warm))))))))))))))))))) (. .)))

A.2.2. Part-Of-Speech tags

The free recall transcripts were submitted to a Part-Of-Speech parser. There was no difference in the number of lemmas between groups (M_{ASD} = 602 ±319, M_{TD} = 608±284, $t(31)=-0.052$, $p>0.9$). The proportion of each type of tag across transcript was computed for each participant separately, and averaged by group (figure B.1). As a crude measure of syntactic structure, the proportion of coordinating conjunctions (CC) and the number of prepositions or subordinating conjunctions (IN) were analysed separately as t

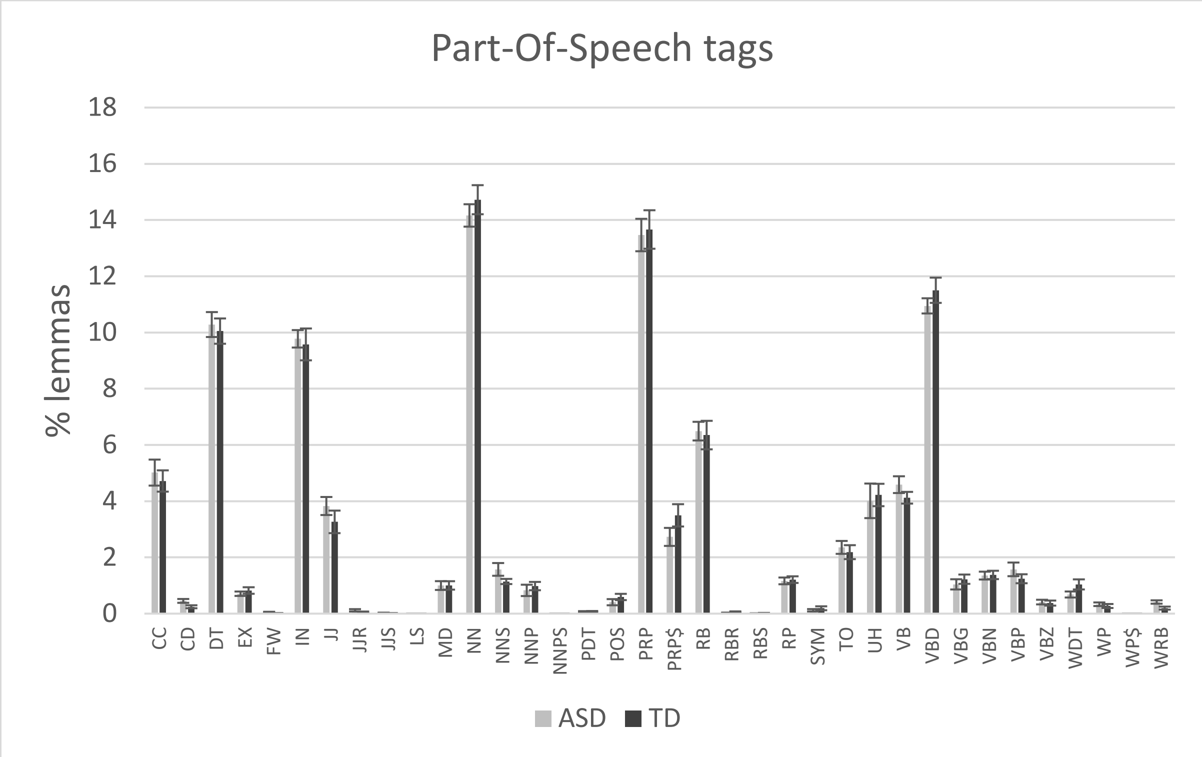

Figure B.1 Proportion of Part-Of-Speech tags across transcript for the ASD and TD groups separately. CC: Coordinating conjunction; CD: Cardinal number; DT: Determiner; EX: Existential there; FW: Foreign word; IN: Preposition or subordinating conjunction; JJ: Adjective; JJR: Adjective, comparative; JJS: Adjective, superlative; LS: List item marker; MD: Modal; NN: Noun, singular or mass; NNS: Noun, plural; NNP: Proper noun, singular; NNPS: Proper noun, plural; PDT: Predeterminer; POS: Possessive ending; PRP: Personal pronoun; PRP$: Possessive pronoun; RB: Adverb; RBR: Adverb, comparative; RBS: Adverb, superlative; RP: Particle; SYM: Symbol; TO: to; UH: Interjection; VB: Verb, base form; VBD: Verb, past tense; VBG: Verb, gerund or present participle; VBN: Verb, past participle; VBP: Verb, non-3rd person singular present; VBZ: Verb, 3rd person singular present; WDT: Wh-determiner; WP: Wh-pronoun; WP$: Possessive wh-pronoun; WRB: Wh-adverb. Errors bars indicate standard error of the mean.