



City Research Online

City, University of London Institutional Repository

Citation: De Clerk, I., Pettinato, M., Verhoeven, J. & Gillis, S. (2017). Is prosodic production driven by lexical development ? Longitudinal evidence from babble and words. *Journal of Child Language*, 44(5), pp. 1248-1273. doi: 10.1017/s0305000916000532

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/15249/>

Link to published version: <https://doi.org/10.1017/s0305000916000532>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk



**Is prosodic production driven by lexical development?
Longitudinal evidence from babble and words.**

Journal:	<i>Journal of Child Language</i>
Manuscript ID	JCL-10-15-112.R3
Manuscript Type:	General Article
Keywords:	Prosodic development, Babbling, Early words

SCHOLARONE™
Manuscripts

For Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Running headline: PROSODIC PRODUCTION IN BABBLE AND WORDS

Is prosodic production driven by lexical development?

Longitudinal evidence from babble and words.

For Peer Review

Abstract

1
2
3
4
5
6 This study investigated the relation between lexical development and the production of prosodic
7 prominence in disyllabic babble and words. Monthly recordings from nine typically developing
8 Belgian-Dutch-speaking infants were analysed from the onset of babbling until a cumulative
9 vocabulary of 200 words was reached. The differentiation between the two syllables of isolated
10 disyllabic utterances was computed for f0, intensity and duration measurements. Results showed
11 that the ambient trochaic pattern emerged in babble, but became enhanced in words. Words
12 showed more prosodic differentiation in terms of f0 and intensity and a more even duration ratio.
13 Age or vocabulary size did not predict the expansion of f0 or intensity in words, whereas
14 vocabulary size was related to the production of more even-timed syllables. The findings are
15 discussed in terms of lexicalist accounts of phonetic development and a potential phonetic
16 highlighting function of first words.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Speech is not a uniform, constant signal. Instead, it evokes the impression of melodic and rhythmic fluctuations, with certain parts (i.e. words, syllables) standing out more than others. The patterning of alternations between stronger and weaker syllables belongs to the study of prosody. Prosody seems to be a universal feature of human languages, as it has also been described in sign languages, which do not use the acoustic modality (Brentari, 1998). This paper aims to study the development of prosodic prominence in early utterances, babble and words, of typically developing infants acquiring Belgian Dutch. Moreover, the impact of the emergence of vocabulary on prosodic abilities is of particular interest in this study?.

Prosodic phenomena have been described at several levels. At word level, the most prominent syllable is said to bear WORD STRESS, e.g. the syllable *straw* in *strawberry*. SENTENCE STRESS or SENTENCE ACCENT denotes the most prominent word at the sentence level (for in-depth discussions, see Nespor and Vogel, 1986 and Selkirk, 1984). Acoustically, prominence at different linguistic levels is usually the result of the single or combined use of an increase in fundamental frequency (f_0), intensity and duration in syllables (Lieberman, 1960; Kochanski, Grabe, Coleman, & Rosner, 2005).

The acoustic properties of word stress seem to be highly salient to infants and they already play a role from birth onwards (Sansavini, Bertoncini, & Giovanelli, 1997). Studies focusing on the perception of word stress in the first year of life confirm this early sensitivity by showing a preference for the predominant stress pattern of the ambient language (Friederici, Friedrich, & Christophe, 2007; Jusczyk, Cutler, & Redanz, 1993). These results with very young infants suggest that it is possible to acquire the acoustic correlates of stress through a purely bottom-up, acoustic strategy. However, the acoustics of prosodic phenomena are not always transparent. Pierrehumbert (2003) offers a thorough discussion of the problem of phonetic

1
2
3 variability in acquisition, pointing out that frequently a one-to-one relation between a prosodic
4
5 phenomenon and an acoustic cue may not be evident: different acoustic correlates of stress are
6
7 used across [languages and talkers](#) (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, &
8
9 Donati, 2008; Sluijter & van Heuven, 1995).
10

11
12 Experimental evidence further suggests that a bottom-up acquisition strategy cannot be
13
14 the sole mechanism, since the manner in which word stress is encoded and processed also
15
16 depends on its function within a language (Cutler, 2005; Dupoux, Sebastian-Galles, Navarrete, &
17
18 Peperkamp, 2008). For instance, Skoruppa et al. (2009) found that the acoustic sensitivity of
19
20 nine-month-old infants was determined by the lexical function of word stress in the ambient
21
22 language. Infants were more sensitive to changes in stress patterns if they acquired a language
23
24 which uses contrastive lexical stress to differentiate between homonyms, like Spanish, than when
25
26 acquiring a language with fixed word stress, like French. Hence, the phonology of stress and its
27
28 role in the vocabulary influences the ability of infants to abstract and store patterns of the
29
30 acoustic input.
31
32
33
34

35 A similar situation has been described in the literature on early speech production, where
36
37 a role for the infant's vocabulary in learning to produce word stress has been posited. This
38
39 suggestion emerged because although some research found evidence that infants potentially had
40
41 the necessary control of the acoustic cues for the production of prominences in speech, they did
42
43 not always apply this consistently, or in accord with the stress pattern of their native language
44
45 ([Davis, MacNeilage, Matyear, & Powel, 2000](#)). Indeed, findings of the early acquisition of word
46
47 stress are not always conclusive, and there is conflicting evidence regarding the use of prosodic
48
49 cues in babble and words. For example, whilst Hallé, De Boysson-Bardies, and Vihman (1991)
50
51 reported language-appropriate control of f0 contours for eighteen-month-old French and Japanese
52
53 infants, with French infants mainly producing rising and Japanese infants falling f0 contours,
54
55
56
57
58
59
60

1
2
3 Vihman, DePaolis, and Davis (1998) found no language-specific strategies in the use of f0 and
4
5 intensity on first disyllabic words of English and French infants. Hallé et al. (1991) found that
6
7 vowel duration in the babble and words of their participants showed features of the ambient
8
9 language: French infants displayed final syllable lengthening whereas Japanese infants did not, as
10
11 this is also reduced in adult Japanese. In contrast, Vihman et al. (1998) and Davis et al. (2000)
12
13 found little overall evidence for final syllable lengthening in disyllabic babble of seven- to
14
15 fourteen-month-old English infants in comparison to adults.
16
17

18
19 Davis et al. (2000) approached these inconsistent findings by comparing disyllabic
20
21 babbled utterances that were judged by adult raters to be clearly stressed with those that were not.
22
23 Fewer than half of the babbled utterances were judged to contain a clearly prominent syllable, but
24
25 in these clearly stressed utterances, f0, intensity and duration were produced in a similar manner
26
27 to a control group of adult speakers. Furthermore, there was no obvious influence of the
28
29 predominant strong-weak (trochaic) pattern of English word stress in the clearly stressed
30
31 syllables. Instead, infants seemed to follow individual preferences for stress assignment, with
32
33 some producing mainly trochaic forms whilst others favoured iambs (weak-strong). Davis et al.
34
35 some producing mainly trochaic forms whilst others favoured iambs (weak-strong). Davis et al.
36
37 (2000) concluded that infants had the capacity to render the necessary cues for prominence, but
38
39 did not do this consistently in babbled utterances. Moreover, even when they were producing
40
41 adult-like cues, the preferred adult stress pattern was not yet evident.
42
43

44
45 These findings led DePaolis, Vihman and Kunnari (2008) to conjecture that prosodic
46
47 control may emerge with the advent of first words. They therefore compared the disyllables at the
48
49 four-word-point (i.e. the first recording in which the infant spontaneously produced four different
50
51 identifiable words, with ages ranging from 10- to 18-months) in French, Welsh, English and
52
53 Finnish. These languages were chosen because the first two have predominantly iambic word
54
55 stress patterns and the latter have a majority of trochaic word forms. These stress patterns were
56
57
58
59
60

1
2
3 however not clearly discernible in babble. They argue that language-specific differences in the
4 use of acoustic cues might only emerge at around eighteen months, when relatively more words
5 than babble are produced. All four groups realised durational differences in line with their target
6 language. It was concluded that ‘a more finely tuned use of prosody may require a level of
7 attention to linguistic detail that begins to be possible only as word production becomes well
8 established’ (DePaolis et al., 2008: 417). It should however be noted that this hypothesis could
9 not be verified with inferential statistics, as the study only included the very beginnings of
10 vocabulary acquisition and therefore yielded an insufficient number of words for statistical
11 comparison to babble.
12
13
14
15
16
17
18
19
20
21
22

23 [The hypothesis of DePaolis et al. \(2008\)](#) is especially relevant, as theories emphasising
24 the role of vocabulary in phonetic and phonological development have gained [increasing](#) interest
25 (see e.g. Edwards, Beckman, & Munson, 2004; Edwards, Munson, & Beckman, 2011; Stoel-
26 Gammon, 2011; Vihman, 2009). Pierrehumbert (2003) conceptualised phonetic development as
27 the process of abstracting phonetic detail over stored lexical items. Werker and Tees (2005) have
28 argued for a cascading model of perceptual tuning for speech sounds, where complex and
29 integrated tasks such as word learning rest on earlier, more basic sensory acquisition of phonetic
30 categories. These levels are interdependent and linked by feed-forward as well as feed-backward
31 mechanisms, so that lexical representations can in turn guide the uptake of new phonetic
32 information to establish and solidify categories.
33
34
35
36
37
38
39
40
41
42
43
44
45

46 These theories were examined in children and adults in a number of studies. Munson
47 (2001) used a nonsense word repetition task to investigate the influence of lexical factors on
48 speech production in three-year-old children and adults. Children repeated less frequent phoneme
49 sequences with reduced accuracy, longer durations and more phonetic variability. Adults'
50 productions were also affected, with less fluent and more variable repetitions of infrequent
51
52
53
54
55
56
57
58
59
60

1
2
3 sequences. This showed that phoneme-sequence frequency effects in the vocabulary had a direct
4
5 impact on children's phonological abilities to repeat novel words, whilst for adults it had an
6
7 effect on phonetic skill. The relationship between vocabulary and speech production in children
8
9 was further explored by Edwards et al. (2004), who found that children with smaller vocabularies
10
11 showed a larger influence of lexical frequency on the repetition accuracy of phoneme sequences
12
13 in nonsense words. Again, low-frequency sequences had longer durations and more phonetic
14
15 variability than sequences with high lexical frequency. It was argued that children with larger
16
17 vocabularies had extracted more robust and flexible independent representations for each
18
19 phoneme and were therefore able to extend them to new contexts more easily. Finally, Storkel
20
21 (2001; 2003) showed that novel word learning was affected by the phonotactic probability
22
23 contained within the words: words with sound sequences that were more common in children's
24
25 lexicons were learned more easily. In summary, these studies suggest that different aspects of the
26
27 lexicon influence phonological and phonetic abilities in children and adults. These factors are the
28
29 vocabulary size and the frequency and phonotactic probabilities of phoneme sequences contained
30
31 in the lexicon.
32
33
34
35
36

37 Given the link between lexical and phonetic/phonological abilities in later developmental
38
39 stages and adulthood, the next question is how early the influence of higher-order language
40
41 processing can be detected on lower-level phonetic acquisition. The studies reviewed in the
42
43 previous paragraph showed lexical effects on segmental development, but the impact of
44
45 vocabulary acquisition on prosodic development is an under-researched area. Moreover, recall
46
47 that DePaolis et al (2008) were not able to confirm the link between the emergence of stable,
48
49 language-specific stress patterns and the appearance of words through inferential statistics.
50
51
52

53 The present study therefore seeks to begin to address this gap in the literature by means of
54
55 a longitudinal phonetic study of prominence patterns in spontaneous speech and by looking at
56
57
58
59
60

1
2
3 disyllabic babble and first words. This study **not only compares** the prelexical and lexical stage,
4
5 but also includes babble which is produced during the lexical stage. By comparing prelexical and
6
7 lexical utterances which occur at the same time in development, it is possible to create a more
8
9 complete picture of the acquisition of prominence than in previous studies. Fundamental
10
11 frequency, intensity and duration were measured in the utterances of infants acquiring Belgian-
12
13 Dutch, from the onset of babbling until a vocabulary of 200 words was reached. By including a
14
15 longer developmental time span than DePaolis et al. (2008), we aim to assess the effect of
16
17 vocabulary development through a direct comparison of disyllabic babble and words over time.
18
19 Just like DePaolis et al. (2008), the present study does not make representational claims for the
20
21 prosodic phenomena measured, instead the focus is on the infants' ability to render the acoustic
22
23 surface form of prominence in Dutch, and whether the emergence of vocabulary affects this
24
25 ability.
26
27
28
29

30 To summarise, two sets of research questions are the focus of the present study:

- 31
32
33 1. Is the acoustic realisation of prosodic differentiation more prominent in words
34
35 than in babble? Is there a bias toward the trochaic stress pattern, and if so, is this
36
37 more prominent in words than in babble?
38
39
- 40
41 2. Is the development of the acoustic correlates of stress predicted by infants'
42
43 vocabulary size, or is stress mastered early in the production of words?
44

45 For the first research question, if DePaolis et al's (2008) prediction is correct, the ambient
46
47 trochaic pattern will be clearly present in early words and individual acoustic cues should show
48
49 **greater differentiaton across** stressed and unstressed syllables in words than in babble. It is also
50
51 hypothesised that infants will be aiming at a strong-weak (trochaic) pattern in words, since in
52
53 Dutch, the adult lexicon is predominantly trochaic (Cutler, 2005; Daelemans, Gillis, & Durieux,
54
55
56
57
58
59
60

1
2
3 1994; Schiller, Meyer, Baayen, & Levelt, 1996)¹. For babble, it is doubtful whether a clear effect
4
5 of the strong-weak pattern will be evident, given the ambiguous results of the literature on
6
7 babbling.
8
9

10 The hypothesis for the second research question is less clear-cut, as there are two equally
11
12 likely outcomes. As some accounts suggest that phonetic development unfolds as an abstraction
13
14 over an increasing mass of lexical items (Pierrehumbert, 2003; Vihman, 2009), it might be the
15
16 case that an infant's vocabulary size determines how detailed phonetic representations are.
17
18 Alternatively it can be hypothesised that the nature of lexical utterances themselves strengthens
19
20 phonetics, i.e. the fact that they are spoken with communicative intent and have linguistic
21
22 representations may already be enough to increase the infant's ability to produce more
23
24 differentiated prominence patterns. Therefore, although the lexical status of disyllables should
25
26 influence cues, an effect of vocabulary size per se will not be seen.
27
28
29

30 **Materials and Method**

31 **Participants**

32
33 The data for this study were taken from the CLiPS Child Language Corpus (CCLC),
34
35 which is a collection of longitudinal audio-video data and transcriptions of 40 typically
36
37 developing Belgian-Dutch acquiring infants.
38
39
40
41

42 For the purpose of this study, 9 children were randomly selected from the corpus. The
43
44 infants were recruited from day-care centres, families known by the researchers and by
45
46 announcements. The infants had been raised in monolingual homes acquiring the standard variety
47
48 of Belgian Dutch (Verhoeven, 2005). All children were tested for congenital hearing loss in the
49
50 first month of life by means of an otoacoustic emissions test. All the infants had normal hearing
51
52
53

54
55 ¹ Although note that for duration, utterance-final lengthening may also influence the realisation of
56
57 prominences in stand-alone utterances and obscure word stress effects (White & Turk, 2010;
58
59 Sluijter & van Heuven, 1995).
60

1
2
3 and no cases of otitis media were reported. The typical development of these children had been
4 established on the basis of parent report and a checklist of the attainment of communicative and
5 motor milestones, which is largely based on the checklist developed by KIND EN GEZIN, the
6 Belgian infant welfare centre (Molemans, van den Berg, Van Severen, & Gillis, 2011). Normal
7 language development was verified by means of the Dutch version of the CDI (N-CDI)
8 administered at ages (years; months) 1;0, 1;6 and 2;0 (Zink & Lejaegere, 2001).
9
10
11
12
13
14
15

16
17 The mean age of these children at the beginning of the recordings was 0;06 ($SD =$
18 0;00.22). The mean age at the cut-off point of 200 words was 1;10 ($SD = 0;03$) (see *Data*
19 *selection* for more information on the cut-off point). The ages of the individual children at the
20 time of recording are specified in Table I.
21
22
23
24

25
26 **INSERT TABLE I. ABOUT HERE**
27

28 29 **Corpus**

30
31 The CCLC consisted of monthly recordings of spontaneous interactions between the
32 children and their caretakers in their home environment. For more details on the corpus and
33 transcription, see Van Severen, Gillis, van den Berg, De Maeyer, and Gillis (2013). A JVC digital
34 video camera with built-in unidirectional microphone was used to record the data. Recordings
35 lasted 60-90 minutes and the fragments during which the child was most vocally active and in
36 uninterrupted interaction with a caretaker were selected. The final selection was approximately
37 20 minutes long.
38
39
40
41
42
43
44
45

46
47 The recordings were phonemically transcribed following the CHAT conventions
48 (MacWhinney, 2000). The criteria for distinguishing words (lexical items) from babble were
49 based on Vihman and McCune (1994). In order to qualify as a lexical item, utterances had to
50 meet at least two out of three criteria: The first was based on the recurring interactional context in
51 which the utterance occurred: i.e. a context where no other word could fit (e.g. the child utterance
52
53
54
55
56
57
58
59
60

1
2
3 *baba* is interpreted as *bal* ‘ball’ as it was produced regularly whilst the child showed a ball), or
4
5 the mother's identification clarified the meaning (e.g. the child utterance *baba* was interpreted as
6
7 *bal* ‘ball’ by the mother). Secondly, the shape of the vocalisation was used as a criterion: it was
8
9 either an exact phonological match or it matched two or more segments of the target (i.e. *pal* for
10
11 *bal* ‘ball’), or the utterance matched the target in its intonation contour or prosody. The third
12
13 criterion took into account the relation to other vocalisations: the utterance was imitated or all
14
15 occurrences of the utterance were identical (i.e. consistent use of *popo* for *opa* ‘grandpa’).

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CHAT transcriptions were converted to a Praat (Boersma & Weenink, 2014) TextGrids using the CHAT2PRAAT function in the CLAN program (MacWhinney, 2013). The video files were converted to audio files using Free-Video-Converter ("Free-Video-Converter," 2012). The resulting TextGrids were time-aligned to the audio files at the utterance level, as illustrated in Figure 1.

INSERT FIGURE 1 ABOUT HERE

Data selection

For the present study, speech data were included from the ONSET OF BABBLING until children had reached a CUMULATIVE VOCABULARY of 200 words. This cut-off point was chosen because it provided enough relevant data to include all children in statistical analyses. Moreover, vocabulary level as a developmental point is used in other studies (Vihman et al., 1998). Cumulative vocabulary is used as a standard measure to track lexical development in psycholinguistic studies (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Vanormelingen, De Maeyer, & Gillis, 2015), and it is based on a list of the word forms a child produces per individual recording session in a longitudinal study. The first cumulative number of vocabulary items comes from the number of words in the first recording session. For every recording that follows, the list of word types in the new session is compared to the previous one, and the

1
2
3 number of new word forms produced are added to the cumulative vocabulary count. The period
4
5 over which data were collected thus spans the onset of babbling, the emergence of first words
6
7 and, as vocabulary sizes increased, the appearance of two and multiword utterances (Bates &
8
9 Goodman, 1997).
10

11
12 Onset of canonical babbling was determined by a True Canonical Babbling Ratio
13
14 (tCBR) of 0.15 or higher (Chapman, Hardin-Jones, Schulte, & Halter, 2001; Molemans et al.,
15
16 2011). The tCBR is the proportion of syllables with true consonants (i.e. all consonants except
17
18 glottals (/h/, glottal stop) and glides (/w/, /j/)) over all syllables produced. Since a longitudinal
19
20 approach was taken in the current study, no artificial boundaries were placed between a babbling
21
22 phase and a lexical phase as there was a transitional period where babble and words co-occurred.
23
24

25
26 **Inclusion criteria for disyllables.** The waveforms and spectrograms associated with the
27
28 speech files were examined in order to identify the disyllabic babble and words to be tagged in
29
30 the PRAAT TextGrids (see Figure 1). Sound sequences were considered to be disyllables when
31
32 they consisted of two vocalic phases minimally separated by a clear consonantal phase (see
33
34 Segmentation criteria for consonants and vowels for specification). Additional consonants
35
36 flanking the vocalic sections were allowed. The inclusion criteria for the disyllables were based
37
38 on DePaolis et al. (2008). In order to be considered a single unit, the disyllables had to occur
39
40 within the same intonation contour or adjacent to a prosodic break such as a pause or an in-breath
41
42 at the beginning of a new breath group (Lieberman, 1984). Furthermore, disyllables had to be
43
44 clearly separated from the surrounding speech by a silence of at least 400 ms with an intersyllabic
45
46 pause of less than 400 ms. For a small number of items produced at a low speech rate, the pause
47
48 criterion was relaxed up to 500 ms as long as the two syllables were part of a single intonation
49
50 contour, indicating cohesion (Cruttenden, 1986). However, lexical items from a multiword
51
52 context were not included in this study. Disyllables were excluded if there was concurrent speech
53
54
55
56
57
58
59
60

1
2
3 or noise or if they were produced with a creaky, breathy, excessively loud or whispery voice. An
4
5 example of a selected utterance is provided in Figure 1.
6
7

8 The recordings in the range studied contained a total of 40 928 child utterances, of
9
10 which 27 807 were prelexical (i.e. babble and vocal play) and 13 121 lexical. The inclusion
11
12 criteria significantly reduced the number of utterances analysed. The number of babbled and
13
14 lexical disyllables per subject is summarised in Table II.
15
16

17 **INSERT TABLE II. ABOUT HERE**
18
19

20 **Segmentation criteria for consonants and vowels.** The disyllables identified by the
21
22 procedure described above were further segmented into consonants and vowels, since the
23
24 acoustic measurements in this study were conducted in the vocalic portion of each syllable.
25
26 Figure 1 illustrates the annotation of an utterance. The waveform, spectrogram, f0 and intensity
27
28 curve were used (a) to determine the word boundaries and (b) to segment the disyllables into
29
30 consonants and vowels. The segment boundaries were identified on the basis of the consistent
31
32 application of the segmentation criteria which are described in detail in DePaolis et al. (2008) and
33
34 to which the interested reader is referred to for more information. Since DePaolis et al. (2008) did
35
36 not specify any criteria for the segmentation of /l/, the onset and offset of the lateral approximant
37
38 were determined on the basis of the discontinuity on the spectrogram in the intensity and/or
39
40 frequency of the formants of /l/ and those associated with the preceding or following vowel.
41
42
43
44

45 **Reliability of the annotations**
46

47 The data were annotated by the author IDC. Approximately 20% of the data ($n = 250$)
48
49 was re-annotated by the author MP. The reliability of the placement of the segment boundaries
50
51 was analysed by means of the Pearson's product-moment correlation between the time points of
52
53 both annotators. The correlation between both sets of annotations was 0.99 ($p < .001$), which
54
55 indicates excellent inter-annotator agreement; annotators located the boundaries of segments at
56
57
58
59
60

1
2
3 virtually the same time points.
4

5 6 **Acoustic analysis**

7
8 The disyllables that were identified by the procedure described above were analysed acoustically
9
10 for the prosodic cues which are relevant to the perception of syllable prominence, duration,
11
12 intensity and f0. The acoustic analyses were carried out by means of a PRAAT script (Boersma &
13
14 Weenink, 2014). Each of the three acoustic parameters was measured in the vowels of the
15
16 disyllables only, not the entire syllable. This was done to reduce potential effects of syllable
17
18 composition on the measurements. Duration (in ms) was measured from the start to the end of
19
20 each vowel. Intensity was measured in dB as the mean energy averaged over the total number of
21
22 analysis frames in the vowel. F0 was determined by means of the PRAAT autocorrelation method
23
24 and expressed in Hz as the mean f0 averaged over the total number of analysis frames in the
25
26 vowel. Intensity and f0 were analysed by the PRAAT analysis settings adjusted to child speech,
27
28 i.e. the f0 range was set at 150-800 Hz. [As there is no gold standard procedure for f0](#)
29
30 [measurements in child speech we performed multiple analyses to test the robustness of the f0](#)
31
32 [range setting and reliability of automatic f0 measurements used in this study. These are provided](#)
33
34 [under Supplementary material 1.](#)
35
36
37
38

39 The intensity range was set at 0-100 dB. It should be mentioned that intensity
40
41 measurements in general need to be treated with caution. Intensity is very sensitive to background
42
43 noise and recording quality. Since participants in this study were highly mobile and clip-on
44
45 microphones were not used, we controlled for possible problematic intensity values by applying
46
47 rigid selection criteria, cleaning the collected dataset for outliers and, most importantly, by
48
49 computing the ratio between the intensity measurements of the two syllables. Ratio measures
50
51 have been used in other studies relying on acoustic (Shriberg, Campbell, Karlsson, Brown,
52
53 McSweeny, & Nadler, 2003) and kinematic (Goffman, 1999) analyses of prosodic modulation.
54
55 The purpose of this study was to investigate the acoustic differentiation between syllables of
56
57
58
59
60

1
2
3 utterances. Therefore, a ratio between the measurements in each syllable was computed to
4
5 quantify this differentiation (i.e. $\text{durationV1}/\text{durationV2}$ and $\text{intensityV1}/\text{intensityV2}$). This also
6
7 had the effect of normalising the intensity in louder utterances. The direction of the prosodic
8
9 differentiation is contained in the ratio: a trochaic tendency is a ratio above one and an iambic
10
11 tendency is a ratio below one. The perceptual f_0 distance between the first and the second vowel
12
13 in each disyllable was calculated by means of the formula $|39,86 \log_{10}(f_0V2/f_0V1)|$. The [absolute](#)
14
15 [pitch distance](#) specifies the perceptual distance in semitones between the first and the second
16
17 vowel in each disyllable. For the pitch distance, the polarity of the signed value indicates the
18
19 direction of the differentiation: a negative value represents a trochaic pattern whereas a positive
20
21 number is an iambic pattern.
22
23
24

25
26 The final dataset consisted of 1151 disyllables of which 525 were babble and 626 were
27
28 words (for numbers per participant, see Table II). The data were cleaned by means of the
29
30 interquartile rule (IQR). [This was done separately for every acoustic measurement, i.e. pitch](#)
31
32 [distance, intensity ratio and duration ratio, and for babble and words separately. Per](#)
33
34 [measurement, every value above or below the IQR threshold was identified as an outlier and](#)
35
36 [excluded from the statistical analysis. For pitch distance, 4.1% of the disyllables were considered](#)
37
38 [outliers, and 1103 utterances were included in the statistical analysis. For intensity ratios, 5.7% of](#)
39
40 [the disyllables were outliers, and 1085 utterances were statistically analysed. For duration ratios,](#)
41
42 [0.01% of the utterances were outliers, and 1139 utterances were analysed. Utterances excluded](#)
43
44 [for one of the three cues were not necessarily outliers for the other cues.](#)
45
46
47

48 **Statistical approach**

49
50 Linear mixed models (LMM) were used for the data analysis (Baayen, 2008). LMM are
51
52 particularly suited for analysing longitudinal corpus data because of their hierarchical structure:
53
54 the observations ($n = 1151$) are measured at different time points embedded in different
55
56
57
58
59
60

1
2
3 participants ($n = 9$). Moreover, linear mixed models are robust to missing and unequal numbers
4
5 of observations for participants and time points. Importantly, when examining the effects of
6
7 independent variables LMM take into account variation at the participant level as well as
8
9 variation over time.
10

11
12 The analyses were carried out in R (R Core Team, 2013) using the lme4 package (Bates,
13
14 Maechler, Bolker, & Walker, 2014) to generate models for the ratio of each prosodic cue. Each
15
16 model consisted of random and fixed effects. In all analyses, the random effects provided
17
18 random intercepts and slopes for each individual and the age points within individuals. The fixed
19
20 effects that were added to each model depended on the research question addressed by the
21
22 analysis and are clarified in the next paragraphs. The statistical procedure consisted of two
23
24 phases. A hierarchical approach was taken to build the models; the random and fixed effects were
25
26 added one by one. At each step, a likelihood ratio test was carried out to arrive at the best-fitting
27
28 model for the data, i.e. the model explaining the largest amount of variance with the fewest
29
30 predictors. In the second phase, we took the best-fitting model and checked which effects were
31
32 significant predictors. The estimates (henceforth E), standard errors (S.E.), t- and p-values of the
33
34 fixed effects of the best-fitting models are reported in the results section.
35
36
37
38

39
40 The first research question concerned the effect of age and utterance type (i.e. babble or
41
42 word) on the [three](#) acoustic cues. Therefore these predictors and their interactions were entered as
43
44 fixed effects in the first set of analyses. For all models, age squared was added to test for non-
45
46 linear development, but since this effect never significantly improved the fit of the models it will
47
48 not be reported. The second research question concerned the effect of vocabulary size on the
49
50 realisation of the three cues. Therefore a second set of analyses with cumulative vocabulary as a
51
52 fixed effect was conducted on the lexical utterances only ($n = 626$) for each of the acoustic cues.
53
54
55 Again, the random effects provided random intercepts and slopes for each individual and the
56
57
58
59
60

1
2
3 developmental points within individuals.
4

5
6 However, since age and cumulative vocabulary were highly correlated in our dataset
7
8 (the Pearson's correlation coefficient $r = 0.81$, $p < 0.001$), the possible vocabulary effect will be
9
10 linked to increasing age. The best-fitting model with vocabulary as fixed effect and the best-
11
12 fitting model with age as fixed effect were compared by inspecting the residual variances. All
13
14 models with cumulative vocabulary as a longitudinal predictor had the lowest residual variances
15
16 and were therefore deemed the most explanatory. Therefore only these models will be reported in
17
18 the results section (The output of all models is provided in the [Supplementary materials 5 to 10](#)).
19
20 In the build-up of the models, cumulative vocabulary squared was added, but as this never
21
22 significantly improved the fit it will not be reported further.
23
24

25 26 **Results**

27
28 Table III summarises all three measurements per subject and utterance type.
29

30
31 **INSERT TABLE III. and FIGURE 2 ABOUT HERE**
32

33 34 **Analysis 1: prosodic differentiation in babble and words**

35
36 **Fundamental frequency.** For pitch distance, the absolute pitch distances (i.e. the unsigned value
37
38 of the distance score between syllable 1 and syllable 2) were entered into the model. The best
39
40 fitting model consisted of the fixed effects of age and utterance type (babble or word). The results
41
42 are displayed in Figure 2 and the output of the statistical model is provided in [Supplementary](#)
43
44 [material 2](#). The estimate of the intercept was 3.261 ($S.E. = 0.236$, $t = 13.769$, $p < 0.001$), with
45
46 'words' as reference level for the factor of utterance type. The fixed effect of age showed no
47
48 significant development over time ($E = 0.045$, $S.E. = 0.028$, $t = 1.593$, $p = 0.127$), indicating that
49
50 pitch distance did not increase in older infants. However, there was a significant effect of
51
52 utterance type on pitch distance ($E = -0.694$, $S.E. = 0.197$, $t = -3.529$, $p < 0.001$); this fixed effect
53
54 indicates that the pitch distance was substantially larger in words than in babble. If we were
55
56
57
58
59
60

1
2
3 furthermore interested in the type of stress pattern, the polarity of the signed pitch distance
4 numbers need to be taken into account. Negative pitch distances indicate higher pitch on the first
5 syllable (i.e. a trochaic tendency); positive pitch distances imply higher pitch on the second
6 syllable (i.e. a iambic tendency). [Table III shows the descriptive statistics that ascertain the type](#)
7 [of stress pattern. Since no inferential statistics were performed on the stress pattern, these](#)
8 [findings should be interpreted cautiously. The values](#) in Table III show that a trochaic tendency
9 is apparent in the babble of 6 out of 9 infants and is boosted towards an even stronger trochaic
10 tendency in the words of 8 infants, as signed pitch distance values became more negative. Only
11 participant 6 did not expand the differentiation towards a trochaic pattern. This might be due to
12 the limited lexical utterances for this participant ($n = 14$, see Table II).
13
14
15
16
17
18
19
20
21
22
23
24

25 In answer to research question one, pitch distance on words was thus greater than on
26 babble and negative polarity suggested a trochaic pattern on words which was already partly
27 evident in babble.
28
29
30
31

32 **Intensity.** As for fundamental frequency, the best-fitting model for intensity had age and
33 utterance type as fixed effects in the statistical model (see [Supplementary material 3](#)). The
34 estimate of the intercept was 1.030 ($S.E. = 0.009$, $t = 115.786$, $p < 0.001$), again with 'words' as
35 the reference level for the factor of utterance type. Although the fixed effect of age significantly
36 improved the fit of the model, the effect itself was not significant, indicating hardly any
37 development over time ($E = 0.001$, $S.E. = 0.001$, $t = 0.809$, $p = .427$). However, the fixed effect of
38 utterance type ($E = -0.016$, $S.E. = 0.007$, $t = -2.183$, $p = .029$) suggests that intensity was not used
39 to the same extent for babble and lexical material, with Figure 2 confirming that the intensity
40 ratio was smaller in babble. In addition, the mean intensity ratios of babble and words were above
41 or around 1, indicating a tendency towards a trochaic pattern (see Table III). This was confirmed
42 in the words of all 9 participants and in the babble of 6 subjects, as can be seen in the individual
43 data. [Thus, in relation to the first research question, words showed increased amplitude](#)
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 modulation in comparison to babble, with the direction of modulation consistent with a trochaic
4
5 prosodic structure.
6
7

8 **Duration.** The best statistical model for duration included utterance type and the
9
10 interaction between age and utterance type as fixed effects in the statistical model (see
11
12 [Supplementary material 4](#)). The estimate of the intercept was 0.856 ($S.E. = 0.055$, $t = 15.678$, $p <$
13
14 0.001), with ‘words’ as reference for the factor of utterance type, as above. There was a
15
16 significant interaction effect between utterance type and age ($E = -0.037$, $S.E. = 0.010$, $t = -3.790$,
17
18 $p < 0.001$), [indicating that the ratios increased in the words with age. Since the estimate for the](#)
19
20 [duration ratio was smaller than one, an increased ratio over time indicates a decrease in](#)
21
22 [differentiation between syllables, i.e. a trend towards more even-timed syllables in words.](#) The
23
24 significant effect of age ($E = 0.037$, $S.E. = 0.009$, $t = 3.920$, $p < 0.001$), irrespective of utterance
25
26 type, is likely to be driven by the ratio increase on the lexical utterances, since the slope for
27
28 babble is flat (see Figure 2). Indeed, utterance type was also a significant effect, since babble
29
30 overall had smaller duration ratios than did lexical utterances ($E = -0.126$, $S.E. = 0.48$, $t = -2.614$,
31
32 $p = 0.009$) [indicating a bigger durational contrast between syllables in babble than in words.](#)
33
34
35
36

37 The mean durational ratios were below one (see Table III), indicating that the second
38
39 syllable was longer than the first in both babble and words. On the surface, this pattern is more
40
41 suggestive of an iambic acoustic pattern, although final lengthening effects cannot be ruled out
42
43 (Sluijter & Van Heuven, 1995). The individual data confirm this tendency. Interestingly, the
44
45 ratios on words grew closer to one, which indicates that the second syllable became shorter in
46
47 words compared to babble, and that the syllables were less differentiated in terms of length.
48
49 Therefore, in respect to research question one, duration differences between syllables actually
50
51 lessened on words in comparison to babble, [though the trochaic pattern was not evident in either](#)
52
53 [words or babble.](#)
54
55
56
57
58
59
60

Analysis 2: the effect of cumulative vocabulary on lexical utterances

The first set of analyses showed a significant effect of utterance type on all three correlates of stress, with words having greater differentiation than babble in terms of f_0 and intensity, and less differentiation for duration. Even the babble produced during the lexical period differed significantly from the lexical utterances produced at the same time in development. Consequently, the question arose whether these changes were systemic, i.e. whether the expansion of the lexicon enabled infants to produce finer and more controlled phonetic structures, or whether instead the intrinsic lexical nature of these utterances caused this development. Therefore, we examined whether the increasing cumulative vocabulary had an effect on the acoustic correlates of stress in lexical utterances.

Fundamental frequency and intensity. Comparable results were obtained for pitch and intensity. The only fixed effect in both analyses was cumulative vocabulary. The estimate of the intercept for pitch was 2.943 ($S.E. = 0.323, t = 9.109, p < 0.001$). For intensity this was 1.029 ($S.E. = 0.011, t = 90.643, p < 0.001$). Both for pitch ($E = 0.002, S.E. = 0.003, t = 0.785, p = 0.454$) and intensity ($E = 0.000, S.E. = 0.000, t = 0.094, p = 0.928$) there was no significant effect of cumulative vocabulary on the prosodic differentiation (see [Supplementary material 5 and 7](#)). This indicates that there was no gradual increase in prosodic differentiation between syllables as infants acquired a larger vocabulary. In answer to the second research question, neither pitch distance nor intensity were predicted by infants' vocabulary sizes.

Duration. For duration, the construction of the best models was the same as for fundamental frequency and intensity. The estimate of the intercept was 0.648 ($S.E. = 0.049, t = 13.349, p < 0.001$). However, the results for this cue were different: the analysis showed a significant effect of increasing cumulative vocabulary ($E = 0.001, S.E. = 0.000, t = 3.159, p = 0.016$) on duration ratios (see [Supplementary material 9](#)). In response to the second research

1
2
3 question, it thus seems that vocabulary level has an effect on prosodic differentiation between
4
5 syllables in terms of duration: the more words infants acquire, the more the ratios tend towards
6
7 one (see Table III), indicating a tendency towards a more equally stressed acoustic pattern.
8
9

10 Discussion

11
12 This study examined the development of prominence production in the spontaneous
13
14 babble and words of 9 typically developing infants acquiring Dutch. Two research questions were
15
16 addressed: firstly, we explored whether there was a difference between babble and words
17
18 regarding the production of fundamental frequency, intensity and duration in differentiating
19
20 between syllables. Furthermore, it was investigated whether an influence of the ambient trochaic
21
22 stress pattern would be evident in either the babble or words of the infants. The second point of
23
24 interest concerned the impact of vocabulary acquisition, i.e. whether there was an effect of
25
26 vocabulary size on the realisation of the cues to prosodic prominence in words. In discussing the
27
28 present findings, it should be borne in mind that 9 infants make up a limited population sample,
29
30 and that further studies would strengthen the present claims. Nevertheless, the substantial size of
31
32 the dataset allowed us to report findings that were robust to variation from individual participants
33
34 (see Table II).
35
36
37
38

39
40 Regarding the first research question, the analyses showed similar findings for f0 and
41
42 intensity in that age did not have a significant impact, whereas utterance type did: the difference
43
44 in f0 and intensity between syllables was significantly larger in words than in babble. Age was
45
46 not a determining factor in prosodic development, whereas prosodic prominence was more
47
48 contrastive in lexical disyllables than in babble. If the advent of vocabulary exerted an influence
49
50 on babble, we should see a concomitant increase in the regression lines for babble. However, this
51
52 was not the case as regression lines stayed remarkably flat even once vocabulary disyllables
53
54 appeared. On words, first syllables typically received higher f0 and intensity than second
55
56
57
58
59
60

1
2
3 syllables and this trochaic pattern is in keeping with the predominant stress pattern in Dutch
4
5 (Cutler, 2005; Daelemans et al., 1994; Schiller et al., 1996). Although the acoustic signature of a
6
7 trochaic pattern is present, it is not known whether adults also perceive this pattern as trochaic,
8
9 and this will be investigated in future research. Moreover, as pointed out in the results section, we
10
11 only consider descriptive statistics to ascertain the type of stress pattern. These findings should
12
13 therefore be interpreted cautiously.
14
15

16
17 Intriguingly, the results for pitch difference and intensity ratio seem to suggest subtle
18
19 traces of the trochaic pattern in the babbling stage, although this tendency is smaller than in
20
21 words, both in terms of signed pitch distance and in the number of children who display it. Again,
22
23 the perceptual effect of these measurements is unclear, but the descriptive statistics indicate that a
24
25 trochaic pattern is potentially present at this early stage. The present results on babble as well as
26
27 the ambiguous results on language-specific stress patterns reported by Davis et al. (2000) and
28
29 DePaolis et al. (2008) are reminiscent of the concept of covert contrasts in child phonology. A
30
31 child is said to have a COVERT CONTRAST between two phonemes when small acoustic differences
32
33 can be detected, e.g. VOT differences for /b/ and /p/, which are not perceivable by adult listeners
34
35 (Macken & Barton, 1980; Gierut & Dinnsen, 1986). The pattern in the data of the present study
36
37 could similarly signal the beginning of language structure, which gradually emerges through fine
38
39 phonetics for the majority of the infants. Since this is such a subtle phenomenon, a fairly large
40
41 dataset might be required to detect it. Previous studies have used relatively small datasets: Davis
42
43 et al. (2000) had 162 babble and DePaolis et al. (2008) analysed 687 disyllables distributed over
44
45 four languages. The present study measured prosody in 525 babble and 626 words; by virtue of
46
47 this large number of observations it may have been possible to observe this broad tendency in the
48
49 phonetic system that had gone unnoticed so far. Alternatively, Goffman (1999) has suggested that
50
51 children initially produce trochees by falling back on a motoric default which emerges during
52
53 canonical babble. In order to disentangle motoric from linguistic influences, similar studies
54
55
56
57
58
59
60

1
2
3 should be carried out with languages which have an iambic default stress pattern, such as Hebrew
4
5 (Segal, Nir-Sagiv, Kishon-Rabin, & Ravid, 2009) to name but one.
6

7
8 For duration, the results were somewhat different. Whilst there was a clear effect of
9
10 lexical status on duration, [in that there was a decrease in durational contrast on words](#), the data
11
12 also show that the second syllable was longer in both babble and words. The overall longer
13
14 second syllable appears to contradict the findings for f_0 and intensity and seems to go against the
15
16 trochaic tendency in Dutch. However, it should be recalled that for duration, it is unclear exactly
17
18 which aspect of prosody is being tapped: since these are isolated disyllables, they are also
19
20 utterance-final, which means that they are subject to prosodic lengthening effects (White & Turk,
21
22 2010), and we are therefore unable to dissociate the effect of emerging word stress from utterance
23
24 final lengthening. A comparison between words which have an iambic target in the adult lexicon
25
26 and those with a trochaic target would have shed further light on the control of duration for word
27
28 stress marking, but infants produced too few iambic target words to allow meaningful statistical
29
30 analyses to be made. For the duration ratios, there was an effect of age, and a significant
31
32 interaction between age and lexical status, suggesting that age had a different effect in babble and
33
34 words. When considering the slopes for the development of duration in Figure 2 it appears that
35
36 the significance of the age effect is caused by words alone, as the slope for babble is again
37
38 virtually flat.
39
40
41
42

43
44 [Interestingly, the duration ratio closer to one on words points to the fact that there is less](#)
45
46 [modulation between the syllables in words than in babble, i.e. there was a reduction in the final](#)
47
48 [lengthening effect on words, leading to a more even-timed pattern. The question that follows is](#)
49
50 [whether this trend is a step towards the adult language or not. On the one hand, metrical theories](#)
51
52 [such as the Trochaic/Iambic law \(Hayes, 1985\) claim that trochaic feet are marked with no](#)
53
54 [durational contrast: the head and the tail of a trochee have even duration. Hence, final](#)
55
56
57
58
59
60

1
2
3 lengthening applied to a weak syllable might give the appearance of an unmodulated syllable. It
4
5 might thus be the case that the words of the infants in the present study show this adult trend.
6

7
8 In contrast, Snow (1994) states that isochrony in stress production signals a period of
9
10 reorganisation. In his study of durational and intonational abilities, he described less modulation
11
12 at the transition from single to multiword utterances than?. Infants displayed adult-like final
13
14 lengthening on their first words, before regressing to more equally-timed syllables between the
15
16 single and multiword stage and eventually reverted to final lengthening once they were fully
17
18 established in the multiword period. Snow (1994) suggested that two different mechanisms were
19
20 at the root of the early and later abilities: physiological constraints (Lieberman, 1984; Robb &
21
22 Saxman, 1990) are thought to underlie the first stage, and volitional, linguistic control over final
23
24 lengthening is only evident once the child has entered the stage of combining words. The
25
26 intervening period of relative isochrony signals a period of reorganisation, where the emerging
27
28 word-combinations prompt children to actively try to control duration in utterances. As this skill
29
30 is not yet acquired, and has to be integrated with nascent word combination, children temporarily
31
32 fall back on rhythmic simplification. In the data of the present study, the shift towards more even-
33
34 timed syllables in words may represent a similar process at the transition from babble to first
35
36 words: the effort of producing an utterance which has an adult phonetic target and is semantically
37
38 and pragmatically appropriate for the context may put enough pressure on the infant's language
39
40 processing to trigger rhythmic simplification. However, since our data also encompass
41
42 measurements at the transition to the multiword stage (Bates & Goodman, 1997; Devescovi,
43
44 Caselli, Marchione, Pasqualetti, Reilly, & Bates, 2005), the loss of final lengthening may partly
45
46 be driven by utterances measured at this stage. Future research may therefore attempt to
47
48 disentangle the effects of the transition from babble to words and single words to multiwords on
49
50 durational phenomena in child speech.
51
52
53
54

55
56 In summary, regarding the first research question, it can be concluded that the arrival of
57
58
59
60

1
2
3 first words does indeed represent a marked transition in the ability to render prosodic differences
4
5 between syllables: all three cues changed with the advent of first words, i.e. the f0 and intensity
6
7 distinction between syllables became significantly larger, and there was a significant reduction of
8
9 utterance final lengthening towards a more evenly-timed pattern in words. Moreover, the
10
11 influence of the ambient trochaic stress pattern was reflected in mean f0 and intensity, possibly
12
13 already at the babbling stage. Utterance-final lengthening appeared to be present in babble and
14
15 reduced significantly with age on first words. A discontinuity in phonetic learning due to the
16
17 emergence of vocabulary has been described for the segmental domain by Vihman and
18
19 colleagues (Vihman & Velleman, 2000; Vihman & Kunnari, 2006). In their surface form, Welsh
20
21 and Finnish share long medial consonants. In Welsh, these are due to accentual lengthening, in
22
23 Finnish, long and short consonants have a lexical function as they distinguish words. At the 4-
24
25 word point, Welsh and Finnish infants produced long medial consonants. By the 25-word point
26
27 Finnish infants approximated the adult bimodal distribution, whereas the productions of Welsh
28
29 infants did not show development. The authors concluded that lexical learning seemed to provide
30
31 a further basis for pattern induction (Vihman & Kunnari, 2006: 158), and we would like to extend
32
33 their proposal to the prosodic domain.
34
35
36
37
38

39 The second research question concerned the nature of the influence of the vocabulary.
40
41 To address this, lexical disyllables were analysed separately, and we examined whether
42
43 cumulative vocabulary or age were significant predictors for each cue. Analyses were also carried
44
45 out with age squared and cumulative vocabulary squared as predictors to test for non-linear
46
47 development. For f0 and intensity, none of these predictors were significant, suggesting that
48
49 vocabulary size does not influence use of these cues. Vocabulary size was however a significant
50
51 predictor for duration, and it provided a model with a significantly better fit for the data than age.
52
53 This indicated that an increase in vocabulary size brought about reduced final lengthening
54
55 independently of chronological age. This may be consistent with Snow (1994), as bigger
56
57
58
59
60

1
2
3 vocabulary sizes are also more likely to be associated with the appearance of word combinations
4
5 (Bates & Goodman, 1997; Devescovi et al., 2005). The fact that results on the role of vocabulary
6
7 size dissociate between f_0 and intensity on the one hand and duration on the other lends further
8
9 support to the idea that the three cues are tapping different levels of prosody (Sluijter & van
10
11 Heuven, 1995).
12
13

14
15 There was no impact of cumulative vocabulary size on f_0 and intensity, although a link
16
17 would be predicted by lexicalist theories of phonetic development which rely purely on frequency
18
19 effects and strength of associations. If it were a question of needing a critical mass of vocabulary
20
21 items before change can take place (Vihman & Velleman, 1989; Walley, 1993) we should see an
22
23 effect for age squared, which signals non-linear development. Instead it seems that by virtue of
24
25 being lexical items, disyllables become prosodically more distinct. Unlike babble, lexical
26
27 disyllables have a clear adult model and a communicative function, although babble can be
28
29 communicative too in a more global way (Stoel-Gammon & Cooper, 1984). In agreement with
30
31 Vihman and colleagues (Vihman & Velleman, 2000; Vihman & Kunnari, 2006), it is further
32
33 proposed that [the linguistic nature of words causes](#) infants to pay greater attention to phonetic
34
35 detail and to consolidate representations. This is backed by a number of findings in different
36
37 domains: in speech perception, a similar refinement of phonetic learning in words has been
38
39 described. Yeung and Werker (2009) showed that nine-month-old infants were able to
40
41 discriminate a non-native phonetic contrast if the training phase used the contrast in a meaningful
42
43 way: if syllables differing by the relevant contrast were consistently used with two different
44
45 objects, the infants, who had not been able to discriminate the contrast before training, were able
46
47 to do so after. If, however, the training phase randomly assigned different syllables to the two
48
49 objects, infants did not succeed at test. Therefore, a labeling function appears to aid phonetic
50
51 perception. Further support comes from Swingley (2009) who suggested that a small proto-
52
53 lexicon may help infants in the acquisition of phonetics by making boundaries between phonetic
54
55
56
57
58
59
60

1
2
3 categories more evident. Crucially, it was hypothesised that this mechanism may be more
4
5 efficient with a smaller than a larger vocabulary, as a small vocabulary results in clearer
6
7 boundaries between phonetic categories. Feldman, Griffiths, and Morgan (2009) explored this
8
9 idea by simulating lexically-constrained phonetic category learning. They reported better
10
11 category distinction for a [Bayesian](#) model which incorporated some lexical knowledge than a
12
13 more classical model (see e.g. Maye, Werker, & Gerken, 2002), which only computed categories
14
15 from the distributional properties of the acoustic space. Finally, in the speech-motor domain, the
16
17 impact of lexical status on articulatory stability has been examined experimentally in four- to six-
18
19 year-old children (Heisler, Goffman, & Younger, 2010). In a fast-mapping task, children's
20
21 repetitions had greater kinematic stability when a nonsense string was presented with a visual
22
23 referent than without. It should be noted that this task did not incorporate a communicative
24
25 function, suggesting that lexicalisation alone was sufficient to stabilise motor patterns in these
26
27 older children.
28
29
30
31

32 We agree with lexicalist accounts that the size of the vocabulary has a prominent role to
33
34 play in phonetic and phonological development, but it is suggested that this may become more
35
36 important at later stages of language development. Instead, we propose that for the early stages
37
38 investigated in the present study, the role of vocabulary development may be somewhat different,
39
40 in that the function of first words for prosody production is similar to the highlighting effect on
41
42 phonetic perception suggested by Werker, Yeung, and Yoshida (2012). Furthermore, this can be
43
44 seen as analogous to work showing that language, and in particular labels, have a facilitatory
45
46 effect on visual category formation (Waxman & Markow, 1995) and can even influence looking
47
48 behaviour towards objects (Althaus & Marechal, 2014). In the same manner, the emergence of
49
50 true linguistic behaviour brings about a re-organisation and refinement of prosodic phonetics.
51
52
53
54
55
56
57
58
59
60

1
2
3 In conclusion, it can be said that traces of the ambient trochaic stress pattern are already present
4
5 in the babble of Belgian-Dutch learning infants, which was only observable by virtue of a
6
7 substantial dataset. Vocabulary items bring about an abrupt transition: f_0 and intensity increase
8
9 significantly, and the control of durational aspects appears to change. The increase in f_0 and
10
11 intensity differentiation is not related to vocabulary size, and therefore accounts which only rely
12
13 on frequency or practice effects may not be appropriate for the emergence of early prominence
14
15 marking. Instead, the communicative and linguistic nature of first words engenders phonetic
16
17 enhancement.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- 1
2
3
4
5
6 Althaus, N., & Marechal, D. (2014). Labels direct infants' attention to commonalities during
7
8 novel category learning. *PLoS ONE*, *9*: e99670.
9
- 10 Baayen, H. (2008). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge:
11
12 Cambridge University Press.
13
- 14 Bates, D., Maechler, M., Bolker, B. M. , & Walker, S. (2014). lme4: linear mixed-effects models
15
16 using Eigen and S4. R package version 1.1-7. Retrieved from [http://CRAN.R-](http://CRAN.R-project.org/package=lme4)
17
18 [project.org/package=lme4](http://CRAN.R-project.org/package=lme4).
19
- 20 Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: evidence
21
22 from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*,
23
24 *12*, 507-584.
25
26
- 27 Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. 5.4. Retrieved 5
28
29 October, 2014, from <http://www.praat.org>.
30
31
- 32 Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.
33
34
- 35 Chapman, K. L., Hardin-Jones, M., Schulte, J., & Halter, K. A. (2001). Vocal development of 9-
36
37 month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*,
38
39 *44*, 1268-1283.
40
41
- 42 Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
43
- 44 Cutler, A. (2005). Lexical stress. In D. Pisoni & R. Remez (Eds.), *The handbook of speech*
45
46 *perception* (pp. 264-289). Oxford: Blackwell.
47
- 48 Daelemans, W., Gillis, S., & Durieux, G. (1994). The acquisition of stress: a data-oriented
49
50 approach. *Computational Linguistics*, *20*, 421-451.
51
- 52 Davis, B., MacNeilage, P., Matyear, C., & Powell, J. (2000). Prosodic correlates of stress in
53
54 babbling: an acoustical study. *Child Development*, *71*, 1258-1270.
55
56
57
58
59
60

- 1
2
3 DePaolis, R. A., Vihman, M. , & Kunnari, S. (2008). Prosody in production at the onset of word
4
5 use: a cross-linguistic study. *Journal of Phonetics*, **36**, 406-422.
6
7
8 Devescovi, A., Caselli, M., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A
9
10 crosslinguistic study of the relationship between grammar and lexical development.
11
12 *Journal of Child Language*, **32**, 759-786.
13
14 Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress
15
16 'deafness': the case of French learners of Spanish. *Cognition*, **106**, 682-706.
17
18
19 Edwards, J., Beckman, M., & Munson, B. (2004). The interaction between vocabulary size and
20
21 phonotactic probability effects on children's production accuracy and fluency in nonword
22
23 repetition. *Journal of Speech, Language, and Hearing Research*, **47**, 421-436.
24
25
26 Edwards, J., Munson, B., & Beckman, M. (2011). Lexicon-phonology relationships and
27
28 dynamics of early language development - a commentary on Stoel-Gammon's
29
30 'Relationships between lexical and phonological development in young children'. *Journal*
31
32 *of Child Language*, **38**, 35-40.
33
34
35 Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). Learning phonetic categories by
36
37 learning a lexicon. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual*
38
39 *Conference of the Cognitive Science Society* (pp. 2208-2213). Austin, TX: Cognitive
40
41 Science Society.
42
43
44 Free-Video-Converter (Version 2.1.0). (2012): QIXINGSHI TECHNOLOGY.
45
46
47 Friederici, A., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are
48
49 already language specific. *Current Biology*, **17**, 1208-1211.
50
51
52 Gierut, J. A., & Dinnsen, D. A. (1986). On word-initial voicing. Converging sources of evidence
53
54 in phonologically disordered speech. *Language and Speech*, **29**, 97-114.
55
56
57 Goffman, L. (1999). Prosodic influences on speech production in children with specific language
58
59 impairment and speech deficits. *Journal of Speech Language and Hearing Research*, **42**,
60

1
2
3 1499-1517.

4
5 Hallé, P., De Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organization:
6
7 intonation and duration patterns of disyllables produced by Japanese and French infants.
8
9
10 *Language and Speech*, **34**, 299-318.

11
12 [Hayes, B. \(1985\). Iambic and trochaic rhythm in stress rules. *Proceedings of the Annual Meeting,*](#)
13
14 [Berkeley Linguistics Society](#), **11**, 429-446.

15
16 Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's
17
18 language production. *Developmental Science*, **13**, 722-730.

19
20
21 Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth:
22
23 relation to language input and gender. *Developmental Psychology*, **27**, 236-248.

24
25 Jusczyk, P., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress
26
27 patterns of English words. *Child Development*, **64**, 675-687.

28
29
30 Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence:
31
32 fundamental frequency lends little. *Journal of the Acoustical Society of America*, **118**,
33
34 1038-1054.

35
36 Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of*
37
38 *the Acoustical Society of America*, **32**, 451-454.

39
40
41 Lieberman, P. (1984). *The biology and evolution of language*. Cambridge: Harvard University
42
43 Press.

44
45 Macken, M., & Barton, D. (1980). Acquisition of the voicing contrast in English. A study of
46
47 voice onset time in word-initial stop consonants. *Journal of Child Language*, **7**, 41-74.

48
49
50 MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition.
51
52 Mahwah, NJ: Lawrence Erlbaum Associates

53
54 MacWhinney, B. (2013). CLAN (Version 04-Nov-2013 11:00).

55
56
57
58
59
60 Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can

- 1
2
3 affect phonetic discrimination. *Cognition*, **82**, B101-B111.
4
5 Molemans, I., van den Berg, R., Van Severen, L., & Gillis, S. (2012). How to measure the onset
6
7 of babbling reliably? *Journal of Child Language*, **39**, 523-552.
8
9
10 Munson, B. (2001). Phonological pattern frequency and speech production in adults and children.
11
12 *Journal of Speech, Language, and Hearing Research*, **44**, 778-792.
13
14 Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008).
15
16 Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio*,
17
18 **7**, 1-28.
19
20
21 Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
22
23
24 Oller, D. (1973). The effect of position-in-utterance on speech segment duration in English.
25
26 *Journal of the Acoustical Society of America*, **54**, 1235-1247.
27
28 Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology.
29
30 *Language and Speech*, **46**, 115-154.
31
32 Robb, M., & Saxman, J. (1990). Syllable durations of preword and early word vocalizations.
33
34 *Journal of Speech and Hearing Research*, **33**(3), 583-593.
35
36
37 Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of
38
39 multisyllabic stressed words. *Developmental Psychology*, **33**, 3-11.
40
41 Schiller, N., Meyer, A., Baayen, H., & Levelt, W. (1996). A comparison of lexeme and speech
42
43 syllables in Dutch. *Journal of Quantitative Linguistics*, **3**, 8-28.
44
45
46 Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeney, J. L., & Nadler, C. J.
47
48 (2003). A diagnostic marker for childhood apraxia of speech: the lexical stress ratio.
49
50 *Clinical Linguistics and Phonetics*, **17**, 549-574.
51
52
53 Segal, O., Nir-Sagiv, B., Kishon-Rabin, L., & Ravid, D. (2009). Prosodic patterns in Hebrew
54
55 child-directed speech. *Journal of Child Language*, **36**, 629-656.
56
57
58 Selkirk, E. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge,
59
60

- 1
2
3 MA: MIT Press.
4
5 Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastian-Galles, N., Limissuri,
6
7 R.A., Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French
8
9 and Spanish infants. *Developmental Science*, **12**, 914-919.
10
11
12 Sluijter, A., & van Heuven, V. (1995). Effects of focus distribution, pitch accent and lexical
13
14 stress on the temporal organization of syllables in Dutch. *Phonetica*, **52**, 71-89.
15
16
17 Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal*
18
19 *of Speech and Hearing Research*, **37**, 831-840.
20
21
22 Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in
23
24 young children. *Journal of Child Language*, **38**, 1-34.
25
26
27 Stoel-Gammon, C., & Cooper, J. (1984). Patterns of early lexical and phonological development.
28
29 *Journal of Child Language*, **11**, 247-271.
30
31
32 Storkel, H. L. (2001). Learning new words: phonotactic probability in language development.
33
34 *Journal of Speech, Language, and Hearing Research*, **44**, 1321-1337.
35
36
37 Storkel, H. L. (2003). Learning new words II: phonotactic probability in verb learning. *Journal of*
38
39 *Speech, Language, and Hearing Research*, **46**, 1312-1323.
40
41
42 Swingle, D. (2009). Contributions of infant word learning to language development.
43
44 *Philosophical Transactions of The Royal Society B: Biological Sciences*, **364**, 3617-3632.
45
46
47 Van Severen, L., Gillis, J., Molemans, I., van den Berg, R., De Maeyer, S., & Gillis, S. (2013).
48
49 The relation between order of acquisition, segmental frequency and function: the case of
50
51 word-initial consonants in Dutch. *Journal of Child Language*, **40**, 703-740.
52
53
54 Vanormelingen, L., De Maeyer, S., & Gillis, S. (2015). Interaction patterns of mothers of
55
56 children with different degrees of hearing: normally hearing children and congenitally
57
58 hearing-impaired children with a cochlear implant. *International Journal of Pediatric*
59
60 *Otorhinolaryngology*, **79**, 520-526.

- 1
2
3 Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association*,
4
5 35, 243-247.
6
- 7 Vihman, M. (2009). Word learning and the origins of phonological systems. In S. Foster-Cohen
8
9 (Ed.), *Language acquisition* (pp. 15-39). Basingstroke, Hampshire: Palgrave Macmillan.
10
- 11 Vihman, M., DePaolis, R., & Davis, B. (1998). Is there a "trochaic bias" in early word learning?
12
13 Evidence from infant production in English and French. *Child Development*, 69, 935-949.
14
- 15 Vihman, M., & Kunnari, S. (2006). The sources of phonological knowledge: a cross-linguistic
16
17 perspective. *Recherches Linguistiques de Vincennes*, 35, 133-164.
18
- 19 Vihman, M., & McCune, L. (1994). When is a word a word. *Journal of Child Language*, 21, 517-
20
21 542.
22
- 23 Vihman, M., & Velleman, S. (1989). Phonological reorganization: a case study. *Language*,
24
25 *Speech, and Hearing Services in Schools*, 33, 9-33.
26
- 27 Vihman, M., & Velleman, S. (2000). The construction of a first phonology. *Phonetica*, 57, 255-
28
29 266.
30
- 31 Walley, A. (1993). The role of vocabulary development in children's spoken word recognition
32
33 and segmentation ability. *Developmental Review*, 13, 286-350.
34
- 35 Waxman, S., & Markow, D. (1995). Words as invitations to form categories: evidence from 12-
36
37 to 13-month-old infants. *Cognitive Psychology*, 29, 527-302.
38
- 39 Werker, J., & Tees, R. (2005). Speech perception as a window for understanding plasticity and
40
41 commitment, language systems of the brain. *Developmental Psychobiology*, 46, 233-251.
42
- 43 Werker, J., Yeung, H., & Yoshida, K. (2012). How do infants become experts at native-speech
44
45 perception? *Current Directions in Psychological Science*, 21, 221-226.
46
- 47 White, L., & Turk, A. (2010). English words on the procrustean bed: polysyllabic shortening
48
49 reconsidered. *Journal of Phonetics*, 38, 459-471.
50
- 51 Yeung, H., & Werker, J. (2009). Learning words' sounds before learning how words sound: 9-
52
53
54
55
56
57
58
59
60

1
2
3 month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*,
4
5 234-243.
6

7 Zink, I., & Lejaegere, M. (2001). *N-CDIs: lijsten voor communicatieve ontwikkeling*.

8
9 *Aanpassing en hernormering van de MacArthur CDI's van Fenson et al.* Leuven: ACCO.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Tables and figures

Table I.

Ages of the individual children at the time of recording.

Subject	Gender	Age start (y;mm.dd)	Age end (y;mm.dd)	# recordings	Onset word use (y;mm.dd)
1	M	0;06.03	2;00.01	19	1;02.09
2	M	0;06.29	2;00.02	18	1;00.01
3	M	0;06.00	1;11.05	19	1;01.31
4	M	0;05.27	1;11.30	19	1;00.30
5	F	0;06.03	1;11.29	19	0;11.00
6	M	0;06.05	2;00.04	19	1;03.30
7	F	0;07.02	1;11.27	18	1;04.02
8	F	0;06.04	2;00.04	19	1;00.05
9	F	0;08.02	1;09.28	15	1;03.00

Note. Age start = the onset of babbling

Table II.

Number of selected disyllables per subject and utterance type (babble and lexical)

Subject	Babble	Lexical	Total
1	122	161	283
2	40	59	99
3	47	96	143
4	87	88	175
5	49	45	94
6	80	14	94
7	36	50	86
8	33	43	76
9	31	70	101
Total	525	626	1151
Mean	58,33	69,56	127,89
SD	29,41	39,85	62,19

Table III.

Summary of the measurements per subject and utterance type (Babble and Lexical)

Subject	Utterance type	Signed pitch distance	Absolute pitch	Intensity ratio	Duration Ratio
		Median (MAD)	distance Mean (SD; range)	Mean (SD; range)	Mean (SD; range)
1	Babble	-0,82 (2,04)	2,56 (1,91; 0,08-7,51)	1,01 (0,07; 0,84-1,20)	0,74 (0,40; 0,08-1,94)
	Lexical	-3,03 (3,43)	4,17 (2,87; 0,07-10,51)	1,05 (0,09; 0,82-1,26)	0,98 (0,51; 0,16-2,34)
2	Babble	-0,84 (1,12)	1,61 (1,31; 0,08-5,32)	1,00 (0,08; 0,80-1,16)	0,83 (0,43; 0,15-1,88)
	Lexical	-1,05 (1,93)	2,47 (1,84; 0,01-8,64)	1,00 (0,07; 0,84-1,14)	0,80 (0,41; 0,19-1,84)
3	Babble	0,78 (2,03)	2,43 (1,89; 0,07-6,35)	1,00 (0,09; 0,83-1,22)	0,60 (0,34; 0,09-1,59)
	Lexical	-1,18 (1,93)	2,93 (2,49; 0,09-10,41)	1,02 (0,08; 0,84-1,16)	0,82 (0,42; 0,10-2,05)
4	Babble	-0,15 (1,44)	1,95 (1,61; 0,01-7,26)	1,01 (0,08; 0,81-1,15)	0,68 (0,36; 0,06-1,93)
	Lexical	-0,44 (3,19)	3,79 (2,54; 0,03-10,06)	1,01 (0,08; 0,82-1,22)	0,78 (0,39; 0,20-2,26)
5	Babble	-0,30 (2,25)	2,31 (1,92; 0,02-7,18)	0,99 (0,08; 0,81-1,19)	0,78 (0,41; 0,13-1,82)
	Lexical	-1,67 (2,31)	3,29 (2,47; 0,01-9,22)	1,02 (0,07; 0,86-1,18)	0,69 (0,37; 0,15-1,52)
6	Babble	0,49 (1,90)	2,24 (1,60; 0,24-7,00)	0,98 (0,08; 0,78-1,21)	0,57 (0,35; 0,11-1,89)
	Lexical	0,78 (1,66)	2,16 (1,89; 0,04-6,01)	1,03 (0,09; 0,86-1,21)	0,60 (0,36; 0,13-1,31)
7	Babble	0,07 (1,69)	1,94 (1,59; 0,07-6,59)	1,00 (0,08; 0,85-1,16)	0,74 (0,35; 0,07-1,19)
	Lexical	-3,10 (2,04)	3,98 (2,48; 0,06-9,28)	1,05 (0,06; 0,92-1,18)	0,75 (0,37; 0,19-1,79)
8	Babble	-1,10 (1,29)	1,69 (1,25; 0,01-5,05)	0,99 (0,09; 0,83-1,17)	0,54 (0,29; 0,15-1,38)
	Lexical	-1,98 (1,63)	2,74 (2,23; 0,15-10,38)	1,04 (0,09; 0,83-1,23)	0,78 (0,48; 0,08-1,94)
9	Babble	-0,98 (1,29)	1,91 (1,73; 0,26-6,13)	1,07 (0,11; 0,84-1,24)	0,99 (0,39; 0,47-1,90)
	Lexical	-2,13 (1,41)	2,81 (2,10; 0,00-10,33)	1,06 (0,07; 0,80-1,23)	1,05 (0,47; 0,32-2,28)
Total	Babble	-0,36 (1,60)	2,17 (1,73)	1,01 (0,08)	0,70 (0,39)
	Lexical	-1,69 (2,42)	3,38 (2,54)	1,03 (0,08)	0,85 (0,46)

Note. MAD = Median absolute deviation; SD = standard deviation; Pitch distance: negative value = trochaic pattern, positive value = iambic pattern; Intensity and duration ratio: Ratio > 1 = trochaic pattern, ratio < 1 = iambic pattern.

1
2
3
4
5 Figure 1.
6

7 *An annotated disyllabic word (auto, /ˈɑːtoʊ/, 'car') in PRAAT. The waveform (top) and*
8 *spectrogram (bottom) were used to segment the speech stream.*
9
10

11 *Note. a = first syllable; b = second syllable; v = vowel; c = consonant*
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3
4
5 Figure 2.
6

7 *Scatterplots of the measurements for both lexical and babbled utterances.*
8

9
10 *Note.* Shaded area = confidence interval
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Supplementary material

Supplementary material 1.

In order to determine the optimal procedure for f_0 measurements in child speech we tested the robustness and reliability of the f_0 range settings:

F0 range

A large f_0 range might increase the risk for anomalous measurements. Therefore the entire dataset was re-analysed with a smaller pitch range of 400Hz (150-650Hz). A series of correlations and kappa scores were carried out on both sets of measurements, to see if the modification of the range could have impacted the findings of the study. The results were as follows:

The Pearson's product-moment correlation between the values reported in the present study (i.e. pitch range of 650 Hz) and the new values (i.e. pitch range of 400 Hz) was $r = 0.87$ ($p < .001$).

To see whether pitch tracker settings could affect the descriptive statistics used to discuss the appearance of trochaic versus iambic pattern, the pitch distances were categorised into trochees (negative values) and iambs (positive values). The kappa coefficient between the old and new datasets was $k = 0.90$.

The high correlations between the reported measurements and the values obtained after narrowing the pitch range to 400 Hz, indicate that both sets of measurements yield highly similar results, which are unlikely to affect the main conclusions of the study.

Automatic measurements

Secondly we tested whether it is appropriate to use automatic measurements for child data without hand correction. This boils down to the question of how much noise this leaves in the data, and whether this can affect the statistical results. Therefore the pitch-tracking of all 1381 files of disyllabic utterances was visually re-inspected.

1
2
3
4
5 (1) 127 files of 1381 were regarded potentially problematic on the basis of visual inspection. This
6
7 amounts to 9,19 % of the entire dataset. Some of the problems had to do with previously
8
9 unnoticed creaky voice, but also with segmental effects of the prevocalic consonant. In these
10
11 potentially problematic files, pitch what re-measured manually.
12

13
14 (2) The correlation between the original dataset and the new dataset with hand measured pitch
15
16 values was $r = 0.86$ ($p < .001$).
17

18
19 (3) We also converted the semitone measurements into a nominal expression of whether pitch
20
21 was higher on the first or second syllable. When comparing the original and new dataset, kappa
22
23 amounted to $k = 0.96$.
24

25 The high correlations between the automatic and hand corrected measurements, indicate that both
26
27 sets of measurements yield highly similar results, which are unlikely to affect the main
28
29 conclusions of the study.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

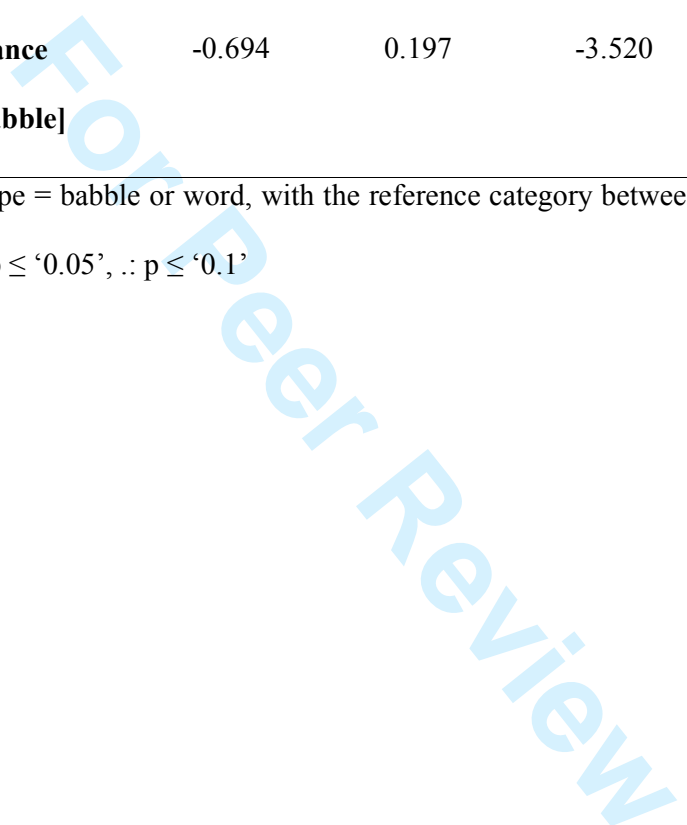
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary material 2.

Statistical output for the pitch distance.

	Estimate	Standard Error	t	p
Intercept	3.261	0.236	13.796	< 0.001***
Age	0.045	0.028	1.593	0.127
Utterance type[Babble]	-0.694	0.197	-3.520	< 0.001***

Note. Utterance type = babble or word, with the reference category between []; ***: $p \leq 0.001$, **: $p \leq 0.01$, *: $p \leq 0.05$, .: $p \leq 0.1$



Supplementary material 3.

Statistical output for the intensity ratios.

	Estimate	Standard Error	t	p
Intercept	1.030	0.009	115.786	< 0.001 ***
Age	0.001	0.001	0.809	0.427
Utterance type[Babble]	-0.016	0.007	-2.183	0.029 *

Note. Utterance type = babble or word, with the reference category between []; ***: $p \leq 0.001$,

** : $p \leq 0.01$, * : $p \leq 0.05$, . : $p \leq 0.1$

Supplementary material 4.

Statistical output for the duration ratios.

	Estimate	Standard Error	t	p
Intercept	0.856	0.055	15.678	< 0.001 ***
Age	0.037	0.009	3.920	< 0.001 ***
Utterance type[Babble]	-0.126	0.048	-2.614	< 0.009 **
Interaction: age*utterance type[age*Babble]	-0.037	0.010	-3.799	< 0.001 ***

Note. Utterance type = babble or word, with the reference category between []; ***: $p \leq '0.001'$,

** : $p \leq '0.01'$, * : $p \leq '0.05'$, . : $p \leq '0.1'$

[Supplementary material 5.](#)

Statistical output for the pitch distance: analysis on lexical subset with cumulative vocabulary as longitudinal predictor.

	Estimate	Standard	t	p
		Error		
Intercept	2.943	0.323	9.109	< 0.001***
Cumulative	0.002	0.003	0.785	0.454
vocabulary				

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, .: $p \leq '0.1'$

[Supplementary material 6.](#)

Statistical output for the pitch distance: analysis on lexical subset with chronological age as longitudinal predictor.

	Estimate	Standard Error	t	p
Intercept	3.254	0.264	12.315	< 0.001***
Age	0.042	0.076	0.552	0.591

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, .: $p \leq '0.1'$

Supplementary material 7.

Statistical output for the intensity ratios: analysis on lexical subset with cumulative vocabulary as longitudinal predictor.

	Estimate	Standard Error	t	p
Intercept	1.029	0.011	90.643	< 0.001***
Cumulative vocabulary	0.000	0.000	0.094	0.928

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, .: $p \leq '0.1'$

[Supplementary material 8.](#)

Statistical output for the intensity ratios: analysis on lexical subset with chronological age as longitudinal predictor.

	Estimate	Standard Error	t	p
Intercept	1.031	0.009	112.093	< 0.001***
Age	0.000	0.003	0.063	0.951

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, .: $p \leq '0.1'$

Supplementary material 9.

Statistical output for the duration ratios: analysis on lexical subset with cumulative vocabulary as longitudinal predictor.

	Estimate	Standard Error	t	p
Intercept	0.648	0.049	13.349	< 0.001***
Cumulative vocabulary	0.001	0.000	3.159	0.016 *

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, .: $p \leq '0.1'$

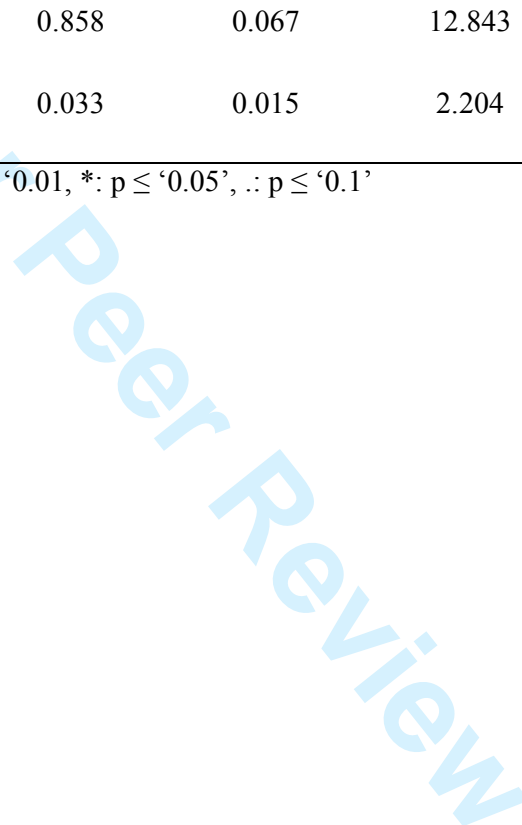
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

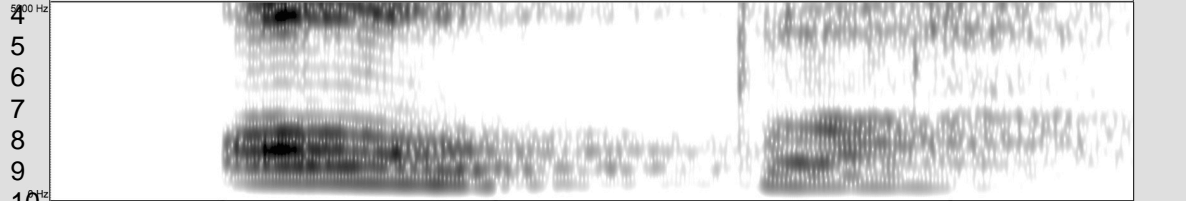
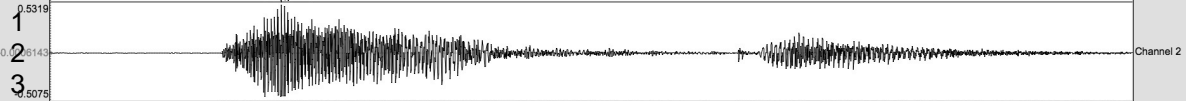
Supplementary material 10.

Statistical output for the duration ratios: analysis on lexical subset with chronological age as longitudinal predictor.

	Estimate	Standard Error	t	p
Intercept	0.858	0.067	12.843	< 0.001***
Age	0.033	0.015	2.204	0.047*

Note. ***: $p \leq '0.001'$, **: $p \leq '0.01'$, *: $p \leq '0.05'$, ..: $p \leq '0.1'$

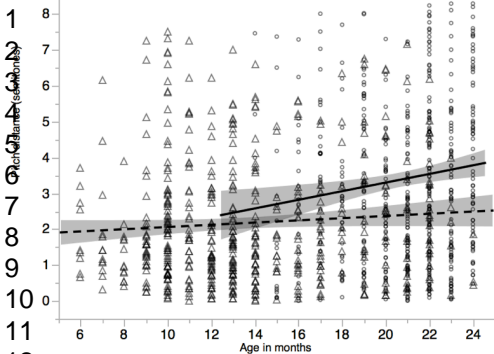




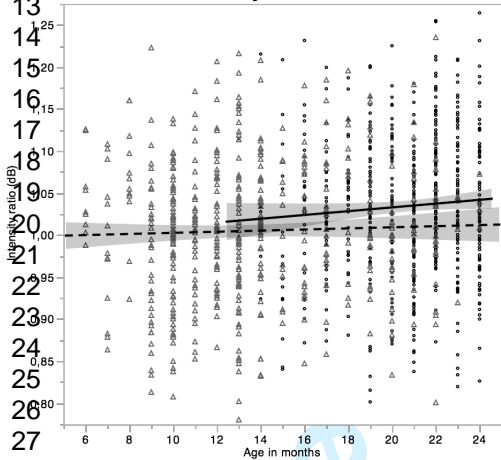
11		a	b		syll (1/4)
12					
13		v	c	v	word (5)
14					

0.747231 239.582231 0.761357 Visible part 0.761650 seconds 240.343882 0.069118

15 Total duration 1.572000 seconds 16



Intensity ratios



Duration ratios

