Evaluation of a novel technique for assessing speech discrimination in children

Caroline Newton[1], Shula Chiat[2] and Lea Hald[3]

[1]Department of Human Communication Science, University College London, Chandler House
2 Wakefield Street, London WC1N 1PF

[2] Department of Language and Communication Science, City University, Northampton Square
London EC1V 0HB

[3]Department of Psychology, University of Sussex, Falmer, Brighton BN1 9QG

Address for correspondence:
Dr. Caroline Newton
Department of Human Communication Science
University College London
Chandler House
2, Wakefield Street
London WC1N 1PF

Email caroline.newton@ucl.ac.uk
Tel: 020-7679-4222
Fax: 020-7713-0861

Abstract

Methods used to assess children's speech perception and recognition in the clinical setting are out of step with current methods used to investigate these experimentally. Traditional methods of assessing speech discrimination, such as picture pointing, yield accuracy scores which may fail to detect subtle perceptual difficulties.

This paper will report a novel method of assessing speech input processing that uses measurement of children's eye movements to provide information on speed and confidence as well as accuracy in discriminating phonological contrasts. Participants were typically developing children aged two to seven years. Pairs of pictures representing auditory minimal pairs which varied in type and degree of phonological contrast were presented on a computer screen while the child heard a word matching one of these pictures. The child's eye movements in response to these stimuli were videorecorded for subsequent analysis of duration and direction of gaze. The effects of age and stimulus type on eye gaze were examined. The results were compared with those of a traditional picture pointing task using the same stimuli.

The informativeness of the novel technique is evaluated on the basis of the findings.

Introduction

When children have speech problems their speech output is a rich source of evidence about their production difficulties. These may be further investigated by comparing children's production on different output tasks, such as naming, repetition of real words, repetition of nonwords, with target phonology carefully selected. (Stackhouse and Wells, 2001, 1997; Dodd, 2005). In all cases speech produced reflects children's speech production abilities. The importance of investigating input processing is also recognised, but is more problematic. Such investigation cannot use spontaneous data and must therefore use techniques for eliciting responses to carefully selected speech input.

At the moment, speech discrimination is typically assessed using picture selection or speech judgement tasks. In picture selection tasks (for example, The Auditory Discrimination and Attention Test (Morgan-Barry, 1989)), the child is presented with two or more pictures representing words differentiated by one or more phonological features, and is asked to point to the picture matching one of these words. In speech judgement tasks (for example, Wepman's Auditory Discrimination Test (Wepman & Reynolds, 1987)), the child is typically asked to say whether a pair of words or nonwords are the same or different.

Each of these methods of assessment is informative up to a point. Reliably correct performance indicates that the child is able to perceive the targeted distinctions, at least in isolated words. However, children may have subtle or specific deficits in perception that are not exposed by standardised tests of perception (see, for example, Burnham, 2003; Locke, 1980).

Innovations in methods of presentation yielding more sensitive measures of response (for example, the gating task used by Edwards, Fox and Rogers (2002)) may also uncover differences in perceptual abilities that are masked by standardised tests. Research on perception in infants provides examples of such methods.

Since infants are not amenable to traditional methods of testing that require intentional pointing and conscious judgement, researchers have had to devise techniques that can detect unconscious responses to distinct stimuli. Preferential looking is one such technique. In this technique, gaze shift and/or duration of gaze are used as measures of infants' discrimination and recognition of linguistic stimuli. In a typical preferential looking scenario, infants are seated on a parent's lap in front of a monitor. Two pictures appear on the monitor, and the child then hears a word or sentence matching one of these pictures. A video camera is used to monitor the child's eye movements following the auditory input. This allows precise identification of changes in the location of the child's fixation, and the duration of the child's fixation at any one location. Using this technique, researchers have, for example, established word recognition skills in 14-15 month olds which are not evidenced until about 24 months when traditional picture or object selection tasks are used (e.g. Swingley and Aslin, 2002; Swingley, 2003; Ballem & Plunkett, 2005).

The preferential looking technique has now been used with a wide variety of linguistic stimuli (see, for example, Hirsh-Pasek & Golinkoff, 1996). However, it has not to our knowledge been used to assess word recognition beyond infancy, whether for research or clinical purposes. Recent research on input processing with clinical populations continues to use more conventional tasks such as the picture pointing and same-different tasks (Coady et al, 2005) and the ABX task (Hill et al., 2005), in which children are asked to indicate which of two auditory stimuli (A and B) matches a third (X).

Yet the preferential looking technique may have advantages even when children are amenable to the more conventional methods mentioned above:

- It taps speech discrimination more directly than either of these methods. Swingley and Aslin (2000:160) note that ‘eye movements are relatively automatic, and under appropriate conditions may reflect cognitive processes that are masked when children must make an overt choice.’
- Computerised presentation allows precise timing and precise measurement providing more information than a simple accuracy score provided by conventional assessments.
- Computerised presentation also permits the development of stimuli targeting a variety of contrasts and can therefore be tailored for individuals.

These advantages of the preferential looking technique motivated our development of Speech by Eye (SP-EYE), a procedure for recording gaze shift and gaze duration in response to speech input. The purpose of the research reported here was to evaluate the effectiveness and informativeness of SP-EYE with normally developing children and compare it with a traditional picture-pointing task.

Methodology

Participants

The participants in this investigation were 50 typically developing children, divided into five age groups of 2-2;11, 3-3;11, 4-4;11, 5-5;11, 6-6;11.

Stimuli

Stimuli were 30 pairs of monosyllabic object names (CVC or CV), distinguished by a single vowel or consonant. The object names distinguished by a vowel differed in the degree of phonological contrast: the distractor was either phonologically ‘close’ (five pairs, e.g. ship-sheep) or ‘distant’ (five pairs, e.g.

car-key). The names distinguished by a consonant differed in the degree and type of contrast of the initial consonant. ‘Close’ pairs differed by only one of the following features:

- Voicing (five pairs; e.g. pin-bin; fan-van)
- Place of articulation (five pairs; e.g. tap-cap; pea-key)
- Manner of articulation (five pairs; e.g. tail-sail; bat-mat)

‘Distant’ pairs differed by all three features, i.e. voice, place and manner (five pairs; e.g. cat-mat; ring-king).

Procedure

Each child participated in two tasks:

SP-EYE

The set of thirty items was presented in five blocks, each with six pairs of stimuli. Each pair of visual stimuli was displayed to the child on a laptop computer (see Figure 1), with the name of one of the objects played through speakers. In order to maintain the child's attention, the trials were embedded in a game of building a jigsaw puzzle, which also served as a means of bringing the child’s attention back to the centre of the screen between trials.

INSERT FIGURE 1 ABOUT HERE

On the naming of the target word, the child’s gaze was recorded by a webcam fixed to the top of the laptop (see Figure 2), for two seconds – until the visual stimuli were removed from the screen. This gave a two-second video file of looking behaviour stored on the computer for each stimulus pair.

INSERT FIGURE 2 ABOUT HERE

In the SP-EYE set-up, the same programme was used to control both the presentation of auditory and visual stimuli and the video recording of participants' looking behaviours, allowing fine temporal synchronisation between these. This ensured that video recording of the movements of the right eye and subsequent timing measurements began at the onset of the target word.

Though some pairs involved the same item (e.g. car-key and pea-key), no item appeared twice in the same block of trials. The positions of the target picture (i.e. left or right) and the order of items were randomised.

The two-second video files created for each child's looking behaviour in response to each trial item provided the basis for analysis. Looking behaviour was coded off-line using software which allowed frame-by-frame analysis of video files (Videolab © Department of Human Comunication Science, University College London). This enabled accurate recording of the direction and speed of first and subsequent shifts of gaze which was used to determine the following:

- Accuracy of first gaze shift

  Correct = gaze shifts to target

  Incorrect = gaze shifts to distractor or elsewhere.

- Response time for correct first gaze shift

  This was measured from the onset of the auditory stimulus (i.e. beginning of video file) to the point where the gaze shifted to the target picture.

- Speed of subsequent shift(s) of gaze

  This was calculated using timings of all shifts of gaze between target and distractor recorded during the two-second video file.

- Duration of gaze at target

  This was the time spent looking at the target picture, as a proportion of the total time spent looking at target and distractor.

Picture Pointing

In the picture-pointing task, the same set of thirty items was presented in hard copy, in the same order, with the same positioning of targets on the page, as in the SP-EYE presentation. The order of tasks (SP-EYE and picture-pointing) was counterbalanced within age and speech groups.

Results

Picture pointing task

The results indicated that children were able to perform the task very well overall (see Figure 3). However, a chi-square test verified that there were age differences in accuracy (chi-square (4) = 34.8, $p < 0.001$). A linear-by-linear association revealed that the differences were such that the older children were, the fewer errors they made (linear-by-linear chi-square (1) = 28.8, $p < 0.001$). Furthermore, a significant effect of contrast type was indicated (chi-square (5) = 46.6, $p < 0.001$).

Figure 3. Performance on picture pointing task according to age

INSERT FIGURE 3 ABOUT HERE

An examination of the errors by contrast type indicated that the majority of the consonant errors were found on the pairs that differed in voicing (see Table 1). A post-hoc analysis of the consonant contrast type voice compared to the combined contrast types place, manner and voice+place+manner (VPM) revealed that this difference was highly significant (chi-square (1) = 44.0, $p < 0.001$). However, in the case of vowels, almost all of the errors occurred with ‘close’ vowel pairs (see Table 1). This difference between the close and distant vowel types was also significant (chi-square (1) = 21.2, $p < 0.001$).

SP-EYE Task

The SP-EYE task yielded several measures of response. Specifically, these were accuracy (direction) of the first shift of gaze from the centre of the screen and the response time of the first shift of gaze from the centre of the screen, response time of subsequent gaze shift(s), and the duration of the gaze on the correct item.

INSERT FIGURE 4 ABOUT HERE

The direction of the first shift of gaze, which we had considered to be a measure of accuracy, indicated that contrary to the high accuracy found with the picture pointing task, the SP-EYE task did not reveal a significant overall effect of accuracy: there were 722 correct and 778 incorrect first shifts of gaze. This may be due to the lack of instruction to 'look at' the target, which we consider further in the Discussion section. Response time of first gaze shift also showed no effect of stimulus type (target vs distractor), and no effect of age. However, older children were slightly but significantly more accurate than younger children (for example two year olds were at 40% correct and six year olds at 54%, $F(4, 1495) = 2.9$, $MSe = 2.9$, $p = 0.02$).

A significant difference was indicated when looking at the speed of second shift by location of the first gaze. As shown in Figure 4, if the participants' first gaze was at the distractor, they were significantly faster to shift their gaze to the target than they were to do the reverse ($F(1, 1190) = 79.6$, $MSe = 94722.4$, $p < 0.001$).

A significant overall effect was seen for the duration of the gaze (see Figure 5), such that the duration of the gaze was longer for the target than the distractor ($F(1, 1190) = 79.6$, $MSe = 94722.4$, $p < 0.001$). However, this effect did not significantly vary by age ($F < 1$).

INSERT FIGURE 5 ABOUT HERE

Since the age differences were small, and not systematically reliable across the different measures, we did not include the age factor in the following analysis. Also, since there was no overall effect of stimulus type on direction of the first shift of gaze or response time of the first shift of gaze, no further analyses are conducted on the individual contrasts types for these measures.

Speed of second shift, on the other hand, had shown differences between target and distractor. From the results on this measure, it was clear that the place of articulation contrast led the participants to be less certain than the voice, manner and voice+place+manner contrasts. This was indicated by the fact that, when the participants first gaze was at the target, they were significantly faster to shift their gaze to the distractor when these contrasted in place (see Figure 6, t (228) = 3.0, p = 0.003). No other contrast types were significantly different from one another in speed of the second shift. Furthermore, no contrast types were significantly different from each other in the duration of gaze type.

INSERT FIGURE 6 ABOUT HERE

Discussion

The picture pointing and SP-EYE tasks were both informative, but may yield different information, since they differed in effects of age and there were subtle differences in contrast effects. Findings indicated that the picture pointing task was informative even for typically-developing children with no reported speech difficulties. Children's performance was not at ceiling, improved significantly with age and showed effects of consonant and vowel contrast type. These findings suggest that picture pointing may be sensitive to subtle differences in children's perceptual skills (provided that children are familiar with all vocabulary items used).

Although all children were close to ceiling in the picture pointing task, results showed age effects and effects of phonological contrast. Specifically, children were observed to make a greater number of errors on the minimal pairs differing by voice than the other stimulus types. It is possible that particular items were responsible for this contrast effect.

In the SP-EYE task, although our original measures of accuracy and reaction time of first gaze shift may have been expected to be informative, responses to target and distractor showed no differences on these measures. It seemed that the set-up of the task, with no instruction to look at the target, meant that the children were as likely to look first at the distractor as at the target. Previous research using the preferential looking methodology has included such an instruction, for example, 'Look! Look! __ !' (Ballem & Plunkett, 2005) or 'Where is the __?' (Swingley, 2003). In our study, the aim was to tap participants' natural response upon simply hearing the target word.

Despite not being instructed to look at the target, so that participants were no more likely to shift gaze first or faster to the target, they nevertheless looked at the target image more than the distractor, and were slower to shift their gaze from the target to the distractor than the reverse. Relative duration of gaze and speed of second shift of gaze appear to be sensitive measures.

There were no systematic changes across age groups in the SP-EYE task. However, there were differences relating to stimulus type. ‘Distant’ pairs showed longer duration of gaze and slower shift of second gaze from target to distractor than ‘close’ pairs. In addition, pairs differing by voice or manner of articulation showed longer duration of gaze and slower shift of second gaze from target to distractor than pairs differing by place of articulation. It might be speculated that this indicates that ‘close’ pairs of stimuli and those differing by place were discriminated less easily or confidently.

The results reported here suggest that SP-EYE provides an informative measure of speech discrimination abilities. As with the picture pointing task, some effects of contrast type (‘close’ vs ‘distant’) were found. Our tentative interpretation is that this task may provide evidence not only of discrimination ability, but of relative attention to differences that are discriminated. It may therefore supplement information about speech processing provided by traditional discrimination tasks.

As part of the evaluation of the SP-EYE technique further investigation is being carried out with 20 children, aged four to seven years, who have disordered phonological output. If the technique proves effective with this group of children, our goal will be to develop SP-EYE as a clinical tool. Given the increasing availability and use of technology in schools and clinics, such a clinical tool is a realistic and practical aim.

References

Ballem, K. & Plunkett, K. (2005) Phonological Specificity in children at 1;2. *Journal of Child Language* 32, 159-173.

Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing* 16, 573-609.

Coady, J.A., Kluender, K.R., and Evans, J.L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research* 48, 944-959.

Dodd, B. (2005). Differential Diagnosis and Treatment of Children with Speech Disorder. 2nd edition. London: Whurr.

Edwards, J., Fox, R. A. & Rogers, C. L. (2002) Final consonant discrimination in children: effects of phonological disorder, vocabulary size, and articulatory accuracy, *Journal of Speech, Language and Hearing Research* 45, 231-42.

Hill, P. R., Hogben, J., & Bishop, D. V. M. (2205). Auditory frequency discrimination in children with Specific Language Impairment: A longitudinal study. *Journal of Speech, Language, and Hearing Research* 48, 1136-1146.

Hirsh-Pasek K. & Golinkoff, R. M. (1996). *The origins of grammar: evidence from early language comprehension*. Camb, MA: MIT Press.

Locke, J. L. (1980b) The inference of speech perception in the phonologically disordered child. Part II: Some clinically novel procedures, their use, some findings, *Journal of Speech and Hearing Disorders* 45, 445-68.

Morgan-Barry, R. (1989) The Auditory Discrimination and Attention Test. Windsor: NFER-Nelson.

Stackhouse, J. and Wells, B. (2001) Children's Speech and Literacy Difficulties:Book Two: Identification and Intervention. London: Whurr.

Stackhouse, J. and Wells, B. (1997) Children's Speech and Literacy Difficulties: A psycholinguistic framework. London: Whurr.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech* 46, 265-294

Swingley, D. & Aslin, R. N. (2002) Lexical neighbourhoods and the word-form representations of 14-month-olds, *Psychological Science* 13, 480-484.

Swingley, D. & Aslin, R. N. (2000) Spoken word recognition and lexical representation in very young children, *Cognition* 76, 147-66.

Wepman, J. F. & Reynolds, W. M. (1987) Wepman's Auditory Discrimination Test. 2nd edition. Los Angeles: Western Psychological Services.

Contrast type	**Accuracy**	
	Incorrect	**Correct**
voice	37	213
place	12	238
manner	8	242
voice+place+manner	4	246
‘close’ vowel	23	227
‘distant’ vowel	1	249

Table1. Picture pointing accuracy by contrast type.

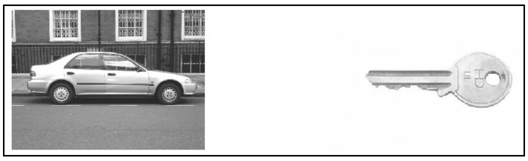

Figure 1. Example of visual stimuli

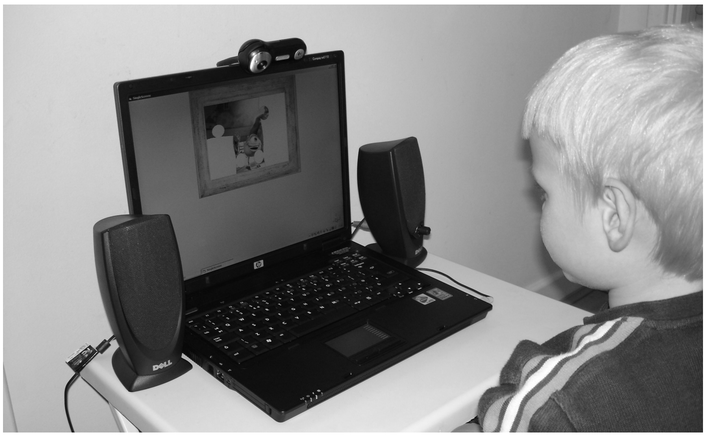

Figure 2. SP-EYE set-up

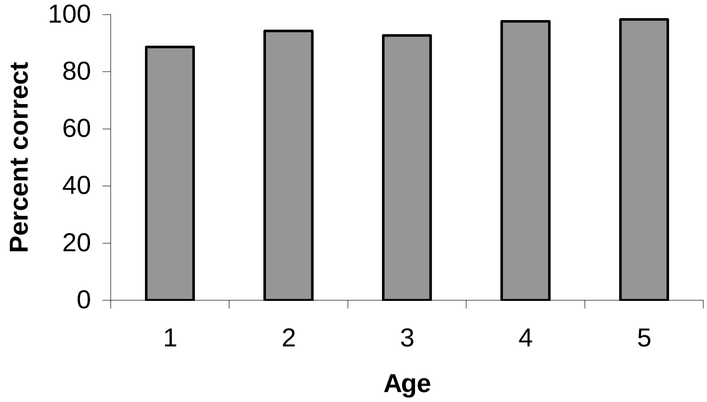

Figure 3. Performance on picture pointing task according to age.

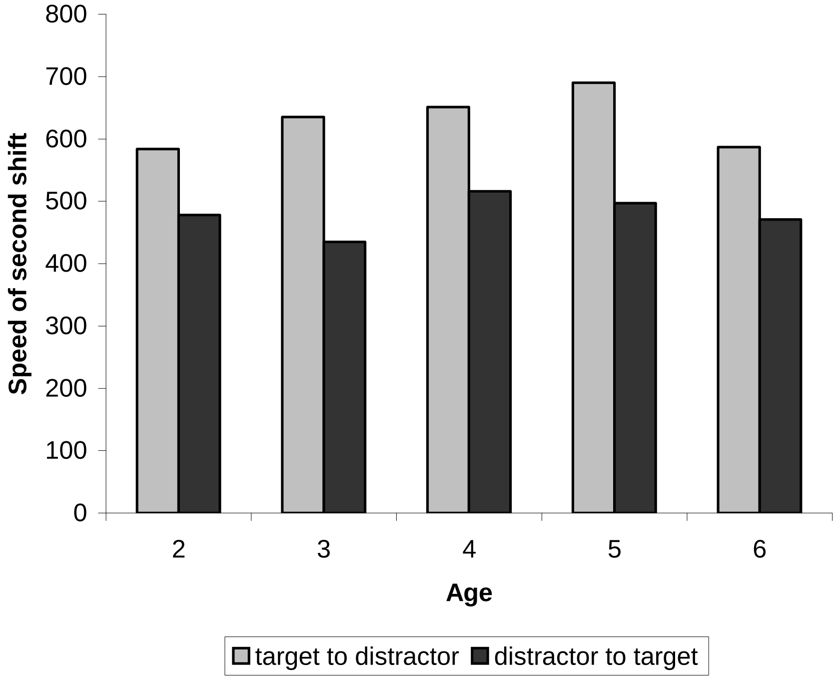

Figure 4. Speed of second gaze shift according to direction of shift and age

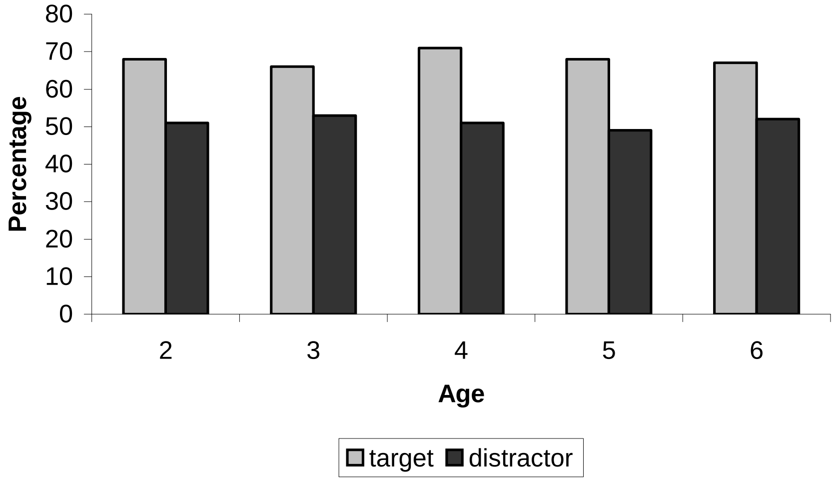

Figure 5. Duration of gaze on target vs distractor according to age

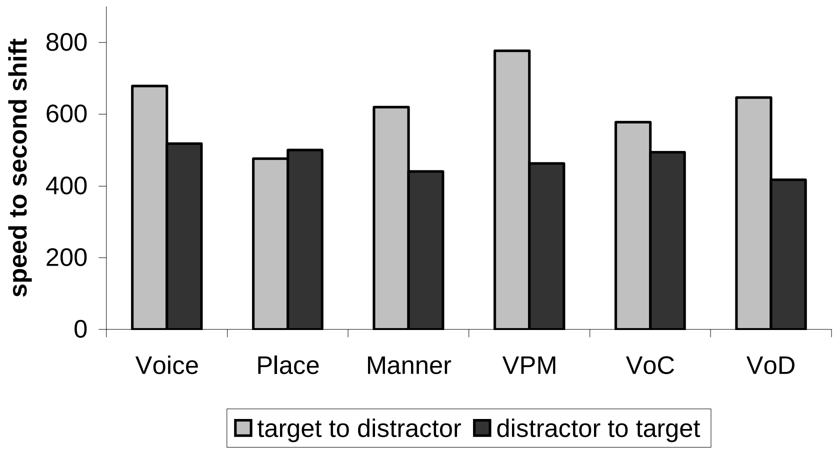

Figure 6. Speed of second gaze shift according to direction of shift and contrast type