Is the development of cue weighting strategies in children's speech perception context-dependent?

Catherine Mayo and **Alice Turk**

Theoretical and Applied Linguistics
University of Edinburgh
catherin@ling.ed.ac.uk, turk@ling.ed.ac.uk

ABSTRACT

It has been proposed that young children, in contrast to adults, may have a perceptual preference for transitional cues as compared to other more static cues. However, it appears that the extent to which children make use of transitional information as compared to adults may change with consonantal context. In perception of fricative contrasts (e.g. /su/–/ʃu/) young children are more influenced by vowel–onset formant transitions, relative to frequency of frication noise, than are older children and adults. But, in identifying the voice onset time contrast /da/–/ta/, children appear to pay less attention to transitional information, relative to the duration of silence, than adults. Possible explanations for these discrepant results are discussed.

1 INTRODUCTION

Numerous studies have shown that children are influenced differently from adults by the acoustic information available in the speech stream. Nittrouer and colleagues, for example, have found that for fricative contrasts (e.g. /su/–/ʃu/) young children give more attention or weight to vowel–onset formant transitions, relative to the frequency of frication noise, than do older children and adults [1, 2, 3]. Other researchers have found similar differences between children and adults in their relative weighting of acoustic cues [4, 5, 6, 7].

Nittrouer and colleagues have proposed an explanation for these differences in cue weighting between adults and children. This hypothesis, called the Developmental Weighting Shift theory (DWS), is based on the premise that children process speech more globally (less analytically) than adults, in terms of large units such as syllables or monosyllabic words [8, 9, 10]. Nittrouer suggests that this global processing manifests itself in speech perception in terms of the acoustic cues that children weight more heavily. Nittrouer goes on to propose that the acoustic correlates of more global speech perception could be syllable–internal vowel formant transitions, because these cues are "perceptually salient and delimit signal portions corresponding to syllables" [11, p. 268]. Children would therefore be perceptually biased towards making heavy use of these cues, thus explaining their behaviour in perception of the fricative contrasts noted above.

It is not clear, however, whether all transitional cues are equally salient to children in all acoustic–phonetic contexts. Nittrouer has found that the influence of transitions on children's perception fluctuates depending on the difference in magnitude between the transitions themselves: children weight transitional information relatively less for /si/–/ʃi/ and /sa/–/ʃa/ contrasts (which have relatively similar vowel onset transitions) than for /su/–/ʃu/ contrasts (which have fairly different onset transitions). Despite these differences, though, children were still found to weight transitions more heavily than adults for all three vowel contexts [2].

However it is possible that given an alternative *consonantal* context, children might give less weight than adults to transitional information. To test this possibility, therefore, this study examines children's and adults' acoustic cue weighting in a non–fricative contrast, specifically the voice onset time (VOT) contrast /da/–/ta/. This particular contrast was chosen because the magnitude of the difference between the transitions should be of similar salience to those in the fricative contrasts /su/–/ʃu/ and /sa/–/ʃa/. If children are indeed perceptually biased towards transitional cues, they should be so in all phonetic contexts, including this one.

2 METHOD

2.1 SUBJECTS

Data from ten three–year–olds, eleven five–year–olds, ten seven–year–olds and four adults were analysed for this study. All children were in full–time primary education in Edinburgh (Scotland) and all were monolingual native speakers of Scottish Standard English (SSE). Parental questionnaires determined that all of the children were free from speech/language disorders, hearing deficits and histories of chronic otitis media.

All adults were monolingual native speakers of English and all had lived in the Edinburgh area for at least 3 years at the time of testing (average duration 14 years). The adults ranged in age from 21 to 26 years (average age 23). All of the adults reported themselves as being free from

ISBN 1-876346-48-5

speech/language disorders, hearing deficits and histories of chronic otitis media.

2.2 STIMULI

The stimuli used in this study were copy–synthesised versions of /da/ and /ta/ based on detailed acoustic analysis of natural tokens spoken by a male native speaker of SSE.

The stimuli were designed following the modified trading relations paradigm used by Nittrouer in most of her studies of /s/–/ʃ/ contrasts [12, 1, 2]. In this paradigm, two cues to the contrast in question are varied: in this study these were (i) VOT, and (ii) frequency of the formants at vowel onset. VOT was varied along a continuum from a duration appropriate for /d/ to one appropriate for /t/. The vowel–onset formant frequencies, on the other hand, took only two forms, either frequencies appropriate for having followed /d/ or those appropriate for having followed /t/.

Combining the two vowel types with each point on the VOT continuum gives two continua which are identical in terms of their VOT, but differ in terms of the frequency of the vowel formants at onset. This allows for an investigation of the perceptual effect of the formant transitions. A listener who is not influenced by transitional information will perceive these two continua as the same, while one who weights transitional information more heavily will perceive the two continua differently.

Nine different VOT values were synthesised, varying in 5 msec steps from 0 msec (most /d/–like) to 40 msec (most /t/–like). Five different vowels were synthesised for each transition condition, based on five different natural utterances of each CV syllable [2]. The average /d/–transition formant onset frequencies were F1: 261 Hz, F2: 1642 Hz, F3: 2472 Hz; the average /t/–transition formant onset frequencies were F1: 537 Hz, F2: 1536 Hz, F3: 2551 Hz. The average vowel target values for all 10 synthetic vowels were F1: 711 Hz, F2: 1433 Hz, F3: 2665 Hz.

Each of the 10 vowels was combined with each of the 9 VOT values, resulting in 90 different stimuli. F0 for each stimulus began at 124 Hz at onset of voicing, rose to 130 Hz 90 msec after onset of voicing, and fell to 60 Hz at vowel offset. The stimuli were randomised for presentation.

2.3 PROCEDURES

All subjects were tested individually in a quiet room. The stimuli were presented over headphones (Sennheiser HD 490, frequency response 17–22000 Hz) via a CD player. Testing for the child subjects took place over two/three consecutive days. Testing for the adult subjects took place on one day, with a short break half way through testing.

All subjects were introduced the target words “da” and “ta.” The child subjects were also familiarised with pictures that corresponded to each word: “da” (short for “dada”) was a picture of a father, “ta” (British English slang for “thank you”) was a picture of someone receiving a present. During testing, the children indicated which word they had heard by placing a counter on the relevant picture. Before testing the children were given an opportunity to practice responding to natural productions of the target words. This ensured that the children were able to identify the targets in natural speech, and that they clearly associated each picture with the relevant target. The children received feedback throughout this practice.

A pre–test was administered to both child and adult subjects. This test consisted of the endpoints of the VOT continua with the appropriate vowel formant transitions for each duration, i.e. the 0 msec VOT plus vowels with /d/–transitions (the most /da/–like stimuli) and the 40 msec VOT plus vowels with /t/–transitions (the most /ta/–like stimuli. There were 10 stimuli in the pre–test (5 repetitions of each VOT condition), presented in random order. All listeners were required to correctly identify 9 of the 10 pre–test stimuli in order for their results to be included in analysis.

For the test proper, the five–year–old, seven–year–old and adult subjects heard the entire set of 90 stimuli twice, resulting in 180 responses per subject, and 10 responses per transition type for each point on the VOT continuum. The 3–year–old subjects heard the set of 90 stimuli only once, resulting in 5 responses per transition type for each point on the VOT continuum for this group. All listeners were required to respond correctly to 80% of the continuum endpoints presented within the test proper for their results to be included in analysis.

3 RESULTS

The results for all subjects were normalised using a probit transformation [2]. This extracts rate–of–change information from data on an S–shaped curve and gives estimates of the mean and the slope of the curve [13]. For this study, the mean is equivalent to the point on the VOT continuum at which the /t/ responses reach 50%, and the slope is equivalent to the degree of categoriality of the responses. From the mean it is also possible to calculate the degree of separation of the response curves by taking the difference of the two means (i.e. the mean of the continuum with /t/–transitions and mean of the continuum with /d/–transitions). This gives a measure of the extent to which response curves were shifted along the continuum as a result of the change in transitional information across continua.

The graphs in Figure 1 illustrate the average perceptual response curves for the three–, five–, and seven–year–old children and the adults.

Visual examination of the response curves indicates that the adults show a greater displacement of category boundaries due to the transition change than do any of the three groups of children. Statistical analysis shows that there is a significant effect of age on separation of response curves ($p = .04$). Further analysis shows that amongst the chil-

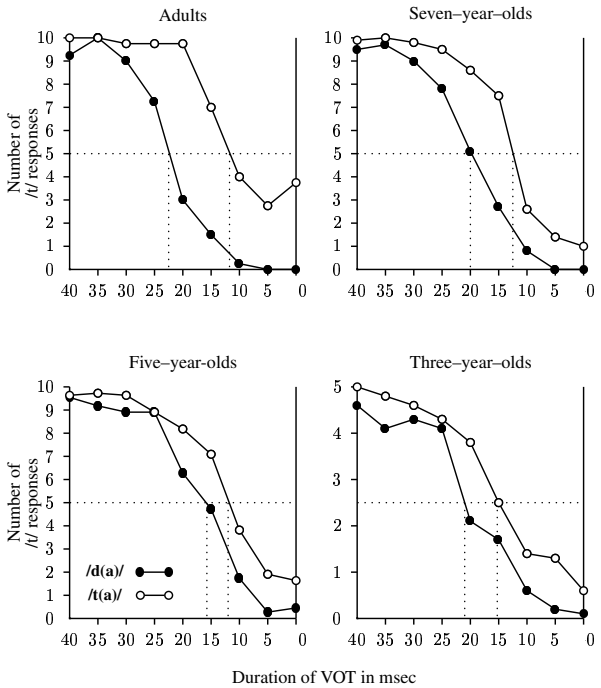

Figure 1: Adults' and children's responses to /da/–/ta/ stimuli. The dotted lines indicate the 50% /ta/ response point.

dren, there is no significant effect of age on separation of the response curves: the significant age–effect above is therefore due to a difference in separation of curves between the children as a group and the adults.

There is no significant effect of age on placement along the VOT continuum of either da–transition response curves or ta–transition response curves. There is also no significant effect of age on the slope of either da–transition response curves or ta–transition response curves. Non–significant age–related differences in placement of the /d/–transition curves are of similar or greater magnitude to non–significant age–related differences in placement of /t/–transition curves.

4 DISCUSSION

In identifying the VOT contrast between /da/ and /ta/, children appear to pay less attention to transitional information, relative to the duration of silence, than adults. This is indicated by the fact that children's perception of /da/ versus /ta/ is less influenced than adults by a change in transitions from those indicating /d/ to those indicating /t/. Comparing this to the previously examined /s/–vowel /ʃ/–vowel results, it would appear that the extent to which children make use of transitional information, as compared to adults, may change with the consonantal context.

There are two possible explanations for this result. The first is that the degree to which children make use of transitional information is indeed influenced by consonantal context. The second interpretation, put forward by Nittrouer [2] is that children are always more perceptually sensitive to transitions than adults. The reason that this second interpretation is possible is due to the nature of transitions following /d/ and /t/ compared to those following /s/ and /ʃ/. For vowels following /s/ and /ʃ/, formants differ in both direction and extent. For vowels following /d/ and /t/, the formant frequencies move in the same direction, but not to the same extent: vowels following /t/ have much less extensive formant transitions than those following /d/. This means that the onset of a vowel following /t/ looks like a truncated version of a vowel following /d/. One could therefore interpret the results of the /da/–/ta/ experiment as meaning that children are so much more sensitive to transitional information than adults that they are prepared to accept even "truncated" /d/–transitions (i.e. /t/–transitions) as indicating a /d/, and thus need only a short silence duration to persuade them that what they have heard is indeed /d/.

There is, however, evidence to suggest that this second interpretation is not the case. In order for this explanation to be correct, the category boundaries for all of the subjects should be the same for the unambiguous, "un–truncated" /d/–transition stimuli, but there should be a difference between children and adults for the ambiguous, "truncated" /t/–transition stimuli: if children require less silence to hear these as voiced, they should place their boundaries closer to the 0 msec end of the VOT continuum than do the adults [2]. However, as noted above, there was no significant effect of age on the placement of either the /da/–transition response curves or the /ta/–transition response curves.

Additionally, if the results for the /da/–/ta/ contrast are due to the fact that /t/–transitions are simply "truncated" versions of /d/–transitions then the above result (greater separation of response curves for adults as compared to children) should be seen for all contrasts in which the transitional information in one syllable is a "truncated" version of the other, no matter what the acoustic–phonetic context. However, this is not what is found. In the /su/–/ʃu/ contrasts that have been shown to engender cue weighting differences in children and adults, F2 following /s/ is a "truncated" version of F2 following /ʃ/. Therefore, if F3 is neutralised in both of the syllables, the resulting /su/–/ʃu/ contrast will be cued by the same type of relationship between transitions as found in the /da/–/ta/ contrast above. According to the above view, children should accept more of these neutral F3 /su/ stimuli as "truncated" versions of /ʃu/. They should therefore show a smaller separation of response curves than adults in perception of this contrast. Nittrouer and colleagues [3] have tested children's and adults' cue weighting of a /su/–/ʃu/ contrast in which F3 was neutralised. However, the results of that study showed that, unlike for the /da/–/ta/ contrast, young children (4 years) showed *greater* separation of response curves than older children (7 years) and both groups of children showed *greater* separation than adults. This is in keeping with studies of /s/–/ʃ/ contrasts

ISBN 1-876346-48-5 © 2003 UAB

in which F3 was not neutral. It would therefore appear that cue weighting differences between adults and children are dependent on the acoustic–phonetic context of the contrast being tested.

5 CONCLUSION

Differences between adults and children in acoustic cue weighting in speech perception is context–dependent. Children weight vowel–onset formant transitions more heavily than adults in the context of fricative contrasts such as /su/–/ʃu/, but less heavily than adults in the context of a VOT contrast such as /da/–/ta/.

This finding has implications for interpretation of the Developmental Weighting Shift theory. It remains clear that children and adults are very different in their speech perception strategies. It is therefore still possible that this is due to differences in the size of unit that children and adults make use of to process speech, with children making use of a larger, more global unit. If this is the case, it is also possible that such global processing should impact on speech perception in terms of the acoustic cues that children attend to most heavily. However, it cannot be the case that the acoustic correlates of global speech perception strategies are always vowel–onset formant transitions. Future studies will have to determine what such correlates might be.

ACKNOWLEDGMENTS

This study was funded by a grant from the Wellcome Trust. Many thanks go to all the children who participated in the study.

REFERENCES

[1] S. Nittrouer and M. Studdert-Kennedy, "The role of coarticulatory effects in the perception of fricatives by children and adults," *Journal of Speech and Hearing Research*, vol. 30, pp. 319–329, 1987.

[2] S. Nittrouer, "Age–related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," *Journal of Phonetics*, vol. 20, pp. 351–382, 1992.

[3] S. Nittrouer and M. E. Miller, "Predicting developmental shifts in perceptual weighting schemes," *Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2253–2266, 1997.

[4] F. Lacerda, "Young infants' discrimination of confusable speech signals," in *The Auditory Processing of Speech: From Sounds to Words*, M. E. H. Schouten, Ed. Mouton de Gruyter, 1992.

[5] B. A. Morrongiello, R. C. Robson, C. T. Best, and R. K. Clifton, "Trading relations in the perception of speech by five–year–old children," *Journal of Experimental Child Psychology*, vol. 37, pp. 231–250, 1984.

[6] R. N. Ohde and K. L. Haley, "Stop–consonant and vowel perception in 3–and 4–year–old children," *Journal of the Acoustical Society of America*, vol. 102, no. 6, pp. 3711–3722, 1997.

[7] M. M. Parnell and J. D. Amerman, "Maturational influences on perception of coarticulatory effects," *Journal of Speech and Hearing Research*, vol. 21, pp. 682–701, 1978.

[8] P. W. Jusczyk and C. Derrah, "Representation of speech sounds by young infants," *Developmental Psychology*, vol. 23, pp. 648–654, 1987.

[9] L. Menn, "Phonotactic rules in beginning speech," *Lingua*, vol. 26, pp. 225–241, 1971.

[10] M. Studdert-Kennedy, "The phoneme as a perceptuomotor structure," in *Language Perception and Production: Relationships Between Listening, Speaking, Reading and Writing*, A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer, Eds., pp. 67–84. Academic Press, London, 1987.

[11] S. Nittrouer, M. E. Miller, C. S. Crowther, and M. J. Manhart, "The effect of segmental order on fricative labeling by children and adults," *Perception & Psychophysics*, vol. 62, no. 2, pp. 266–284, 2000.

[12] H. L. Fitch, T. Halwes, D. M. Erickson, and A. L. Liberman, "Perceptual equivalence of two acoustic cues for stop–consonant manner," *Perception & Psychophysics*, vol. 27, no. 4, pp. 343–350, 1980.

[13] J. Cohen and P. Cohen, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, LEA, Hillsdale, NJ, second edition, 1983.