No available theories currently explain all adult-child cue weighting differences

Catherine Mayo & Alice Turk

Centre for Speech Technology Research & Theoretical and Applied Linguistics
University of Edinburgh, UK
catherin@ling.ed.ac.uk, turk@ling.ed.ac.uk

Abstract

Children and adults appear to weight some acoustic cues differently in perceiving certain speech contrasts. There are currently two main theories to explain this difference. One of these is the Developmental Weighting Shift theory, which proposes that children process speech in terms of more global, syllable-like units. Thus, in this view, children should always give more weight than should adults to cues like within-syllable vowel formant transitions, and less weight to across-syllable vowel formant transitions [1]. Other researchers have proposed that children have lower general auditory sensitivity than adults, which impacts on their speech perception. Thus, in this view, children should always give more weight than adults to cues that are longer, louder, or more spectrally informative than the alternative cues [2].

The current study tested these hypotheses in two ways. First, we examined adults' and three- to seven-year-old children's weighting of vowel-onset formant transitions in contrasts in which we systematically varied the consonantal context and the spectral distinctiveness of the transition. Second, we examined adults' and five-year-old children's weighting of vowel-formant offset transitions in spectrally identical within-monosyllabic-word and across-monosyllabic-word contexts. The results of the study showed that adult-child differences in cue weighting are affected by the segmental context of the cues, the salience of the cues, and the position of the cue in the word. However, neither of the above two theories, either on their own or in combination, can account for all of the observed cue weighting behaviour.

1. Introduction

A number of studies have found that, for some speech contrasts, children and adults give different amounts of perceptual weight to different acoustic cues. For example, when identifying /s-vowel/-/ʃ-vowel/ contrasts on the basis of frequency of frication noise and configuration of vowel-onset formant transitions, children have been found to weight transitions more heavily than do adults [3].

There are two prevailing theories to explain these differences in transitional cue weighting. Nittrouer and colleagues' Developmental Weighting Shift (DWS) theory proposes that children process speech more globally than adults, in terms of units such as syllables. This processing difference leads to children giving more weight than adults to cues that "delimit signal portions corresponding to syllables" [4, p. 268] such as spectrally dynamic within-syllable vowel formant transitions. Evidence for the DWS theory is equivocal: there are a number of studies which have indeed found that children weight transitional cues more than do adults for identifying /s-vowel/-/ʃ-vowel/ contrasts [3], for identifying voicing in final stops [5], and for identifying place of articulation in some initial stops [6]. However, other studies have found that children are less influenced than adults by transitional cues for identifying some voicing contrasts [7] and for identifying vowels [2].

An alternative explanation to the DWS theory is that adult-child differences in cue weighting are due to the maturation of some aspect(s) of the central auditory system [8, 2]. Under this explanation, children are less able to use incomplete or insufficient acoustic information than adults in perception of both speech and non-speech stimuli. Thus, in speech perception as in non-speech perception, children should give heavier weight than adults to "louder or longer duration speech cues...or [cues containing] greater amounts of spectral information" [2, p. 1173]. However, attempts to demonstrate a correlation between cue weighting patterns for speech and general auditory sensitivity have provided conflicting results [9, 2].

It is therefore unclear to what extent either the DWS theory, or a more general central auditory theory, can explain all adult-child cue weighting differences. The study described here was therefore designed to systematically test both theories, as described in more detail below.

2. Experiment 1

In this experiment, we examined adults' and three- to seven-year-old children's weighting of vowel-onset formant transitions in contrasts in which we systematically varied (i) the consonantal context of the transition: /s/-/sh/, /d/-/b/, /t/-/d/, and /n/-/m/, and (ii) the spectral distinctiveness of the transition (in terms of the onset frequency, extent, direction and duration of the vowel-onset formant transitions), varying from spectrally distinct (/do/-/bo/, /ta/-/da/ and /no/-/mo/) to spectrally similar (/de/-/be/, /ti/-/di/ and /ni/-/mi/).

The DWS theory would predict that children should give more weight than do adults to the within-syllable vowel-onset formant transition cues, regardless of consonantal context or spectral distinctiveness of the transitions. A general auditory theory predicts that children should give more weight than do adults to the spectrally distinct transitional cues, but less weight than do adults to the spectrally similar transitions, again regardless of consonantal context.

2.1. Method

2.1.1. Participants

Child participants were three-, five- and seven-year-old monolingual native speakers of Scottish Standard English (SSE). All of the children and their siblings were free from speech/language disorders, hearing deficits and histories of chronic otitis media, and all performed age-appropriately on tests of reading [10] and receptive vocabulary [11]. Adult par-

ticipants were monolingual native speakers of English living in Edinburgh (average duration of time in Scotland: 12 years). All of the adults were free from speech/language disorders, hearing deficits and histories of chronic otitis media. Exact numbers of participants who successfully identified each set of stimuli are given in the relevant figures in the Results section.

2.1.2. Stimuli

The contrasts used in Experiment 1 were /saɪ/-/ʃaɪ/, /de/-/be/, /do/-/bo/, /ta/-/da/, /ti/-/di/, /no/-/mo/ and /ni/-/mi/. Continua of synthetic speech sounds were created for each of these contrasts. The endpoints of the synthetic continua were copy-synthesised versions of the target syllables listed above, based on detailed acoustic analysis of natural tokens spoken by a male native speaker of SSE (aged 39 years, with normal speech, language and hearing). The stimuli were created using *SenSyn*, a cascade/parallel formant synthesiser based on [12].

A trading relations design was used to create the stimuli. In this paradigm, two continua of speech sounds are created for each contrast, in which two acoustic cues are manipulated. One of the two cues is varied *along* both continua. The two continua are therefore identical in terms of this cue. The other cue is varied *across* the two continua. The two continua therefore differ in terms of this cue. In this part of the study, the along-continua cue was one non-transitional cue for each contrast, as follows: Frequency of frication noise for /saɪ/-/ʃaɪ/, frequency of stop burst for /de/-/be/ and /do/-/bo/, duration of voice onset time for /ta/-/da/ and /ti/-/di/ and frequency of nasal murmur for /no/-/mo/ and /ni/-/mi/. The across-continua cues were the frequency- and time-varying properties of vowel-onset formant transitions in two conditions: (i) appropriate for having followed the first consonant in the contrast, or (ii) appropriate for having followed the second consonant in the contrast.

Two 9-point continua were created for every contrast. Five different repetitions of the same vowel were synthesised for each transition condition, and each of these vowels was combined with the 9 continuum values, resulting in 90 stimuli per contrast. (For details of the *SenSyn* parameters manipulated, see [13, 14])

2.1.3. Procedure

With the exception of the subjects who heard the /saɪ/-/ʃaɪ/ contrasts, each subject heard both sets of stimuli for each consonant contrast (i.e., those subjects who heard the VOT contrasts listened to both the /ta/-/da/ and the /ti/-/di/ contrasts). All subjects were tested individually in a quiet room. The stimuli were presented over headphones via a CD player, at a comfortable listening level.

Before the main test, a pre-test was administered which consisted of the congruent endpoints of the continua (the endpoint values of the along-continuum cue followed by the congruent vowel-formant transitions for each along-continuum cue condition). During the main test, the five-year-old, seven-year-old and adult subjects heard a complete set of 90 stimuli twice, in two different random orders, resulting in 180 responses per subject and 10 responses per transition type for each point on the continuum. The three- to four-year-old subjects heard a complete set of 90 stimuli only once, resulting in 5 responses per transition type for each point on the continuum for this group. During testing, the children indicated which word they had heard by saying the word aloud, and by placing a counter on one of two pictures; adults filled in their responses on a form.

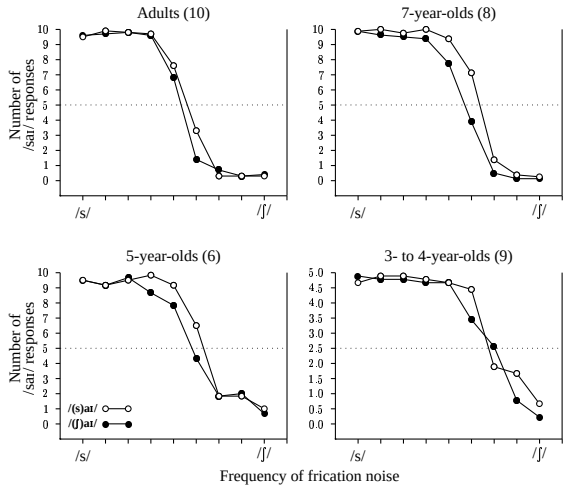

Figure 1: Adults' and children's responses to /saɪ/-transition stimuli (open circles) and /ʃaɪ/-transition stimuli (filled circles). Numbers in brackets indicate number of participants per age group.

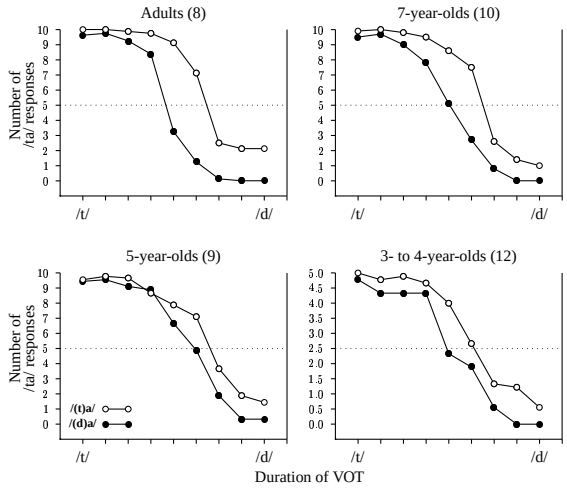

Figure 2: Adults' and children's responses to /ta/-transition stimuli (open circles) and /da/-transition stimuli (filled circles). Numbers in brackets indicate number of participants per age group.

2.2. Results

Because the results of this experiment are extensive, only those most relevant to the aims stated above will be discussed here. Figures 1, 2, and 3 show response curves for all 4 age groups for the contrasts /saɪ/-/ʃaɪ/, /ta/-/da/, and /ti/-/di/, respectively. For the /saɪ/-/ʃaɪ/ contrast, listeners responses were generally consistent with those found previously for other /s-vowel/-/ʃ-vowel/ contrasts, namely children as a group showed a significantly greater separation of response curves than did adults in response to a change in transitional information [$F(1, 31) = 5.50, p = .026$].

However, for the /ta/-/da/ and /ti/-/di/ contrasts, children did not appear to pay more attention to transitional information than did adults. For the /ti/-/di/ contrast, children gave the same weight as adults to transitional cues. This is indicated by the fact that there was no significant difference between children as a group and adults in separation of response curves for this contrast. For the /ta/-/da/ contrast, children gave *less* weight

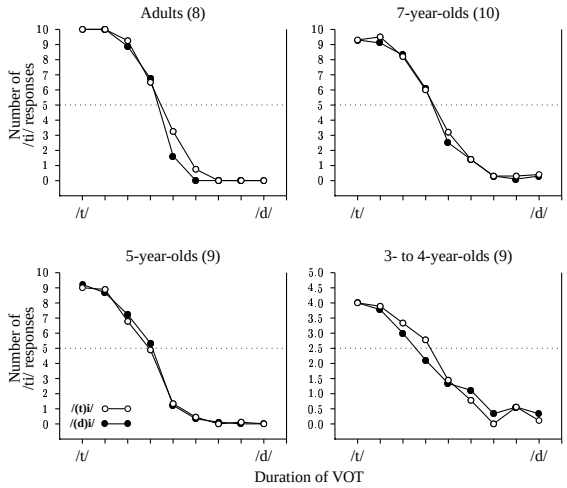

Figure 3: Adults' and children's responses to /ti/-transition stimuli (open circles) and /di/-transition stimuli (filled circles). Numbers in brackets indicate number of participants per age group.

than adults to transitional cues. This is reflected in the fact that children as a group showed a significantly smaller separation of response curves due to a change in transitional information than did adults [$F(1, 37) = 5.89, p = .02$].

3. Experiment 2

In this experiment we examined adults' and five-year-old children's weighting of vowel-formant offset transitions in spectrally identical within-monosyllabic-word and across-monosyllabic-word contexts. The transition in question was from the vowel /e/ into a following /d/ or /b/ closure, either within-word: "Ade E" versus "Abe E" (a woman and a man with a surname beginning with "E"), or across-word: "A Dee" versus "A bee" (a girl named "Dee" with the letter "A" on her shirt; a bee with the letter "A" on it).

The DWS theory would predict that children should give more weight to the within-word transitional cue presented in the "Ade E" versus "Abe E" contrast, while at the same time giving less weight to the across-word transitional cue presented in the "A Dee" versus "A bee" contrast; adults should display the opposite pattern of weighting.

3.1. Method

3.1.1. Participants

Child participants were 20 five-year-olds and 14 adults who met the same criteria as outlined for Experiment 1.

3.1.2. Stimuli

The contrast used in this cue weighting study was the VCV contrast /edi/-/ebi/, with the main acoustic cue of interest being the /e/-offset vowel formant transition. The contrast was presented so that this /e/-offset transition appeared in two contexts: (i) a within-monosyllabic word context, in the contrast "Ade E."–"Abe E.", and (ii) an across-monosyllabic word context, in the contrast "a Dee"–"a bee".

This part of the study also made use of a trading relations design using copy synthesised speech stimuli (see description above). Here, the along-continuum cue was the /i/-onset formant transition, which changed along a 9-point continuum from values appropriate for having followed /d/ (F1=214 Hz, F2=2209 Hz, F3=2772 Hz) to values appropriate for having followed /b/ (F1= 214Hz, F2=1977 Hz, F3=2508 Hz). The across-continua cue was the /e/-offset formant transition, which was appropriate either for preceding a /d/ (F1=310 Hz, F2=2213 Hz, F3=2733 Hz) or preceding a /b/ (F1=310 Hz, F2=1500 Hz, F3=2640 Hz). A complex stop burst was created with spectral peaks of 24 dB at 2500 Hz and 38 dB at 5200 Hz. This burst was neutral as to whether it cued a /d/ or a /b/ stop closure.

The two /e/ vowels were each combined with the neutral stop burst, and then with each of the points on the 9-point /(d)i/-/(b)i/ continuum, creating a pair of /edi/-/ebi/ continua, one with /ed/-offset transitions, and the other with /eb/-offset transitions. To create the within-word and across-word conditions, two durations were manipulated: the duration of the target (i.e., unchanging) portion of the /e/ vowel (235 msec in the within-word condition, and 175 msec in the across-word condition) and the duration of the intervocalic stop closure (35 msec in the within-word condition and 75 msec in the across word condition). Vowel durations were appropriate for signalling a contrast between "Ade" versus "Abe" and "Dee" versus "bee" with secondary phrasal stress on the /e/ in "A Dee" and "A bee". The stop closures were appropriate for word-final versus word-initial closures in these stress contexts. All other characteristics of the stimuli remained the same across the within- and across-word conditions.

3.1.3. Procedure

Each subject heard both sets of stimuli, that is, both the "Ade E" versus "Abe E" contrast and the "A Dee" versus "A bee" contrast. The general presentation procedure was the same as in Experiment 1, described above.

3.2. Results

The DWS theory would predict that adults should always weight across-word transitions more heavily than within-word transitions (referred to here as Pattern 1), while children should always weight within-word transitions more heavily than across-word transitions (referred to as Pattern 2). Paired t-tests carried out on listeners' separation of response curves due to the change in /e/-offset transition for the within-word "Ade E." versus "Abe E." contrast and the across-word "A Dee" versus "A bee" contrast support these predictions. Overall, children gave significantly more weight to within-word transitions than they gave to across-word transitions [$t(19) = 2.301, p = 0.03$]. In contrast, adults were found to give significantly more weight to across-word transitions than they gave to within-word transitions [$t(13) = -2.66, p = 0.02$].

However, an examination of the responses of individual subjects shows a great deal of variability in response patterns. Figure 4 shows examples of response curves from two different children and from two different adults: these graphs illustrate the fact that both children and adults showed *both* Pattern 1 and Pattern 2 type responses. Table 1 shows the distribution of Pattern 1 and Pattern 2 behaviour across the two age groups tested.

Table 1: *Number of subjects displaying each pattern of weighting, by age group.*

	within > across	within < across
5-year-olds	13	7
Adults	4	10

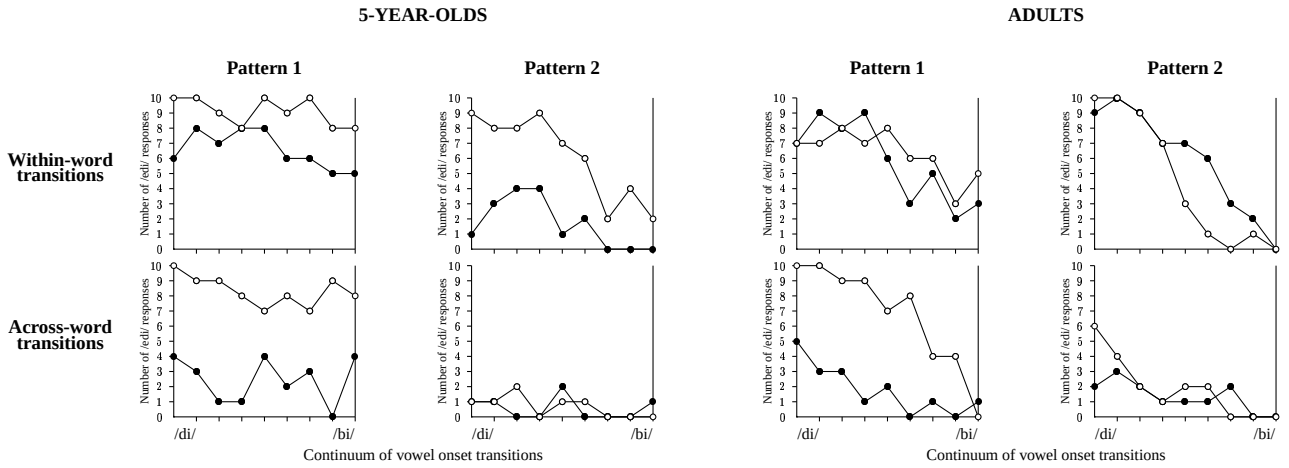

Figure 4: Adults' and children's responses to /ed/-offset transition stimuli (open circles) and /eb/-offset transition stimuli (filled circles). Each vertical pair of graphs represents the responses of a single child or adult participant.

χ^2 analysis shows that there is no significant relationship between age and pattern of weight given to within- versus across-word transitions: there are not more children who display the pattern predicted for children by the DWS (Pattern 2) than there are children who display the more "adult" pattern (Pattern 1). Similarly, there are not more adults who display the pattern predicted for them by the DWS (Pattern 1) than there are adults who display the "child" pattern (Pattern 2).

4. Discussion

The results from Experiment 1 for the /ta/-/da/ and /ti/-/di/ contrasts contradict both the DWS prediction and a more general auditory prediction for these stimuli. First, the findings that children and adults are equally influenced by transitions for /ti/-/di/, and that children show less influence of transitional cues than adults for /ta/-/da/ suggest that, rather than being consistently more biased than adults toward transitions, as predicted by the DWS, the extent to which children make use of transitional information as compared to adults changes with segmental context. Second, the fact that children were found to weight the spectrally distinctive transitions in /ta/-/da/ less heavily than adults suggests that not all adult-child cue weighting differences can be explained by a general auditory sensory hypothesis.

The results from Experiment 2 provide limited support for the DWS theory, in that children were shown to weight within-word transitions more than they weight across-word transitions, and adults were shown to weight across-word transitions more than they weight within-word transitions. However, there was a great deal of individual variability in responses, with both children and adults showing both possible patterns of weighting. This high level of individual variability suggests that it is inappropriate to say that children always weight within-monosyllabic word transitions more than do adults, and always weight across-monosyllabic word transitions less than do adults, as predicted by the DWS.

In conclusion, neither the DWS theory nor a more general auditory theory can account for all of the adult-child differences in cue weighting observed in this study and others. Further research is required to identify a possible cause of these differences.

5. References

[1] 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.

[2] J. E. Sussman, "Vowel perception by adults and children with normal language and specifi c language impairment: Based on steady states or transitions," *Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1173–1180, 2001.

[3] 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.

[4] 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.

[5] C. Wardrip-Fruin and S. Peach, "Developmental aspects of the perception of acoustic cues in determining the voicing feature of fi nal stop consonants," *Language and Speech*, vol. 27, no. 4, pp. 367–379, 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] P. Howell, S. Rosen, H. Lang, and S. Sackin, "The role of F1 transitions in the perception of voicing in initial plosives," in *Speech, Hearing and Language: Work in Progress*, University College London, 1992.

[8] L. S. Eisenberg, R. V. Shannon, A. Schaefer Martinez, J. Wygonski, and A. Boothroyd, "Speech recognition with reduced spectral cues as a function of age," *Journal of the Acoustical Society of America*, vol. 107, pp. 2704–2710, 2000.

[9] S. Nittrouer and C. S. Crowther, "Examining the role of auditory sensitivity in the developmental weighting shift," *Journal of Speech, Language, and Hearing Research*, vol. 41, pp. 809–818, 1998.

[10] F. Schonell and E. Goodacre, *The Psychology and Teaching of Reading*, Oliver and Boyd, London, 1971.

[11] L. M. Dunn, L. M. Dunn, C. Whetton, and J. Burley, *British Picture Vocabulary Test*, NFER-NELSON, Berkshire, UK, second edition, 1997.

[12] D. Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, pp. 971–995, 1980.

[13] C. Mayo and A. Turk, "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," *Journal of the Acoustical Society of America*, vol. 115, pp. 3184–3194, 2004.

[14] C. Mayo and A. Turk, "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception," submitted.