Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Predictability and adult-child cue weighting differences in speech perception

Catherine Mayo[1,2], *Alice Turk*[2] *and Robert A. J. Clark*[1]

[1]Centre for Speech Technology Research, University of Edinburgh
[2]Linguistics and English Language, University of Edinburgh
catherin@inf.ed.ac.uk, turk@ling.ed.ac.uk, robert@cstr.ed.ac.uk

Abstract

In this experiment, we tested the hypothesis that adult-child differences in cue weighting are influenced by adult-child differences in knowledge of (a) the relative predictability of word-initial vs. word-final consonants, and (b) of the relationship between predictability and acoustic salience/distinctiveness. We tested our hypothesis using synthetic speech continua with formant transitions varying from /edi/ to /ebi/, which listeners were encouraged to hear as either "Abe E/Ade E" (VC#V context) or as "A bee/A dee" (V#CV context). We tested the extent to which changes in formant transitions influence /d/ vs. /b/ categorisation. Results show that adults were more influenced by transitions cueing word-initial consonants (less predictable in English) than by transitions cueing word-final consonants (more predictable in English), whereas children showed a more balanced pattern, with marginally more influence of transitions cueing word-final consonants. Results are consistent with the view that adults have learned more about the relative predictability of word-initial vs. word-final consonants and have learned that acoustic cues to the less-predictable initial consonants are more distinctive. They therefore weight these cues more heavily than less-distinctive, more contextually predictable, word-final cues.

Index Terms: speech perception, acoustic cue weighting, perceptual development, Smooth Signal Redundancy

1. Introduction

It is well established that adults and young children differ in the perceptual attention, or weight, each group gives to different aspects of the speech stream [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. For example, when identifying members of a /so/-/ʃo/ ("sew-show") contrast, adults' responses tend to be more influenced by the frequency distribution of the frication noise than children's, while children's responses tend to be more influenced by the frequency and direction of movement of the vowel-onset formant transitions, as compared to adults [3, 4, 7, 13].

Two classes of explanation have been proposed to account for these differences: i) differences in the perceptual acuity of children's vs. adults' auditory systems ([10], but see [5]), and ii) language learning/exposure/experience [14, 15, 16, 17, 18, 19, 20, 21, 22]. In this paper, we test the hypothesis that a particular type of learning affects the development of cue weighting behaviour, namely learning about the relative predictability of speech sounds in different structural positions (here word-initial vs. word-final positions), as well as learning about the relationship between relative predictability and acoustic clarity/distinctiveness.

Our hypothesis derives from proposals in the production literature that acoustic distinctiveness/salience/clarity (acoustic redundancy) relates inversely with predictability from context (language redundancy), so that all elements in the speech stream have equal recognition likelihood (Smooth Signal Redundancy [23, 24]). Aylett and Turk [24, 25] propose that speakers control the inverse relationship between language and acoustic redundancy through prosodic prominence and constituent structure: Elements that are highly predictable from context are less likely to be prosodically marked, and therefore are less acoustically salient than less predictable elements. What this suggests is that less predictable, prosodically strong elements (elements preceded by a boundary, or prosodically prominent elements) are more distinct and salient than prosodically weak elements that are highly predictable from context. For example, it is well established that word-initial consonants are more physically distinct in both amplitude and formant transitions than are more predictable word-final consonants ([26] and others). The Smooth Signal Redundancy view explains this difference in terms of the greater contextual predictability of word-final consonants.

As Norris and McQueen [27] point out, Bayes' theorem leads us to expect that listeners asked to identify a linguistic element will rely less on acoustic information when this information is ambiguous or less distinct than when it is clear and salient, and consequently will rely more on prior beliefs about the likelihood of the element. Luce and Pisoni [28] presented supporting evidence that shows that listeners rely more on word frequency when identifying words presented in noise than when identifying words presented in the clear.

In this experiment, we test the hypothesis that adult-child differences in cue weighting are influenced by adult-child differences in knowledge of the relative predictability of word-initial vs. word-final consonants, and of the relationship between predictability and acoustic salience/distinctiveness. We hypothesise that adults will weight cues to final consonants less heavily than cues to initial consonants because they expect cues in this highly predictable position to be more ambiguous than cues to consonants in initial position. In contrast, children who have not yet learned about the relative predictability of word-initial vs. word-final consonants should show less of a difference in cue weighting between word-initial vs. word-final consonants.

We test our hypothesis using synthetic speech continua with formant transitions varying from /edi/ to /ebi/, which listeners are encouraged to hear as either the VC#V contrast "Abe E/Ade E" (ABE-ADE context) or as the V#CV contrast "A bee/A dee" (BEE-DEE context). We test listeners' reliance on /e/-C and C-/i/ transitional cues in the two word boundary contexts. Cue weighting is assessed by the extent to which changes in formant transitions influence /d/ vs. /b/ categorisation. Because there is a larger acoustic difference in /ed/ vs. /eb/ transitions compared to /di/ vs. /bi/ transitions, we expect both age groups to weight consonantal transitions more heavily in /e/-C contexts compared to

10.21437/SpeechProsody.2016-113

C-/i/ contexts.

2. Methods

2.1. Participants

Sixteen adults (age range 19-29 years, average age 21 years), and 34 five-year-olds (age range 4;8-6;0, average age 5;3) were tested. All were monolingual native speakers of Scottish Standard English (SSE), and all reported themselves (or were reported by parents) as being free from speech/language/hearing disorders and (at the time of testing) upper respiratory infection. Child subjects were in their first year of full-time education and performed appropriately for their age on standardised tests of reading (Schonell Graded Word Reading Test, [29]) and receptive vocabulary (BPVS, [30]).

2.2. Stimuli

Two sets of trading relations stimuli were designed to show whether children and adults weight formant transitions differently in word-initial vs. word-final contexts. These two sets of stimuli had identical segmental contrasts and contexts for those contrasts, namely /ebi/-/edi/. Both sets of stimuli also had physically identical vowel formant transitional cues to the contrasts: (i) the /e/-offset vowel formant transition into the /b/ or /d/ closure, which was varied across each pair of continua, and (ii) the /i/-onset vowel formant transition out of the /b/ or /d/ closure, which was varied along each pair of continua.

The two sets differed in terms of the placement of a word/syllable boundary. In the ABE-ADE set of stimuli, the syllable boundary appeared after the consonant: /eb#i/-/ed#i/ ("Abe E" vs. "Ade E") and in the BEE-DEE set of stimuli, the word/syllable boundary appeared before the consonant: /e#bi/-/e#di/ ("A bee" vs "A Dee").

The difference in word boundary location was signalled in two ways: 1) by manipulating two durational cues, the duration of the steady-state portion of the pre-stop /e/-vowel, which was longer in the ABE-ADE condition and shorter in the BEE-DEE condition, and duration of the stop closure, which was shorter in the ABE-ADE condition and longer in the BEE-DEE condition, and 2) by telling listeners what contrast they would be hearing (e.g., "You will now hear examples of two phrases, 'Abe E.' and 'Ade E.' "), which explicitly indicated how to segment the stimuli. Crucially, the formant transition manipulations were identical in both boundary conditions.

2.3. Synthesis

Synthetic speech continua were created for the contrasts used in the study using the Sensyn [31] version of the Klatt cascade/parallel synthesizer [32]. The congruent endpoints of the continua were copy synthesised based on acoustic analysis of natural tokens of /ebi/ and /edi/ produced by an adult male speaker of SSE.

Two /e/-vowels were created, with offset transitions appropriate for preceding either a /d/ closure or a /b/ closure. The /ed/-transition vowels had offset values of: F1=310 Hz, F2=2213 Hz, F3=2733 Hz; the /eb/-transition vowels had offset values of F1=310 Hz, F2=1500 Hz, F3=2640 Hz. The duration of the offset transitions of the /e/-vowels was 20 msec for the /ed/-transition stimuli and 30 msec for the /eb/-transition stimuli. The onset values for the formants of both /e/-vowels were: F1=389 Hz, F2=2210 Hz, F3=2735 Hz, while the target values for both vowels were: F1=389 Hz, F2=2256 Hz, F3=2766 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. A 9-point continuum of /i/-vowels was created by changing the time-varying values of F2 and F3 in the onset transition of the vowel, from values appropriate for having followed /d/ to values appropriate for having followed /b/. The onset frequencies of the formants at the /di/-appropriate end of the continuum were: F1=214 Hz, F2=2209 Hz, F3=2772 Hz. The onset frequencies of the formants at the /bi/-appropriate end of the continuum were F1= 214Hz, F2=1977 Hz, F3=2508 Hz. The /i/ vowel target values for all stimuli were: F1=263 Hz, F2=2282 Hz, F3=2816 Hz. 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. This resulted in 18 different stimuli.

The duration of the steady-state portion of the /e/-vowel and the duration of the intervocalic stop closure were manipulated to create the two different syllable boundary conditions. In the ABE-ADE condition, the vowel target portion of the /e/ was 235 msec and the intervocalic stop closure was 35 msec. In the BEE-DEE condition, the vowel target portion of the /e/ was 175 msec and the intervocalic stop closure was 75 msec. This doubled the number of individual stimuli from 18 to 36: 18 for each syllable boundary condition.

All of the synthesized stimuli had an F0 contour designed to signal a high pitch accent on the /e/ vowel, appropriate to contrast "Abe" with "Ade", and "A" with an implied "B", and a downstepped high accent on /i/. The total duration of each ABE-ADE stimulus was 530 msec, with 235 msec of /e/-vowel (including transition), 35 msec of stop closure, 5 msec burst and 255 msec of /i/-vowel. In these stimuli, F0 for each /e/-vowel began at 130 Hz at voicing onset, rose to 1400 Hz 60 msec after onset, rose again to 1500 Hz 45 msec following this, and fell to 1200 Hz at voicing offset. The total duration of each BEE-DEE stimulus was 510 msec, with 175 msec of /e/-vowel (including transition), 75 msec of stop closure, 5 msec burst and 255 msec of /i/-vowel. In these stimuli, F0 for each /e/-vowel began at 140 Hz at voicing onset, rose to 1500 Hz 45 msec after this, and fell to 1200 Hz at voicing offset. F0 for each /i/ vowel began at 1200 Hz at voicing onset and fell to 900 Hz at voicing offset. In the ABE-ADE contrast, the stop burst occurred 270 msec after the onset of the /e/ vowel (235 msec of vowel, plus 35 msec of closure), while in the BEE-DEE contrast, the stop burst occurred 250 msec after the onset of the /e/ vowel (175 msec of vowel, plus 75 msec of closure).

2.4. Procedure

All participants 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 at a comfortable listening level. Testing for the children took place over two or three days. Testing for the adults took place on one day, with a short break half way through testing. All listeners heard both ABE-ADE and BEE-DEE contrasts, but were only ever presented with stimuli from one contrast at a time. The listeners' task was to identify individual stimuli as either one or the other half of the relevant contrast (e.g., as either "Abe E." or "Ade E." *or* as either "A bee" or "A Dee"). The adult participants performed the task alone, by entering their responses on a form. The child participants provided their responses to the experimenter by saying the word aloud, and by placing a counter

on a picture corresponding to the relevant word (for "Abe E."-"Ade E.": a boy called "Abe" and a girl called"Ade", both with surnames beginning with the letter "E"; for "A bee"-"A Dee": a bee with an "A" on it and a girl called "Dee" with an "A" on her shirt).

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 the provided pictures with the relevant targets. The children received corrective feedback throughout this practice, and did not proceed until they had, unprompted, correctly identified a complete set of 10 randomly presented natural stimuli (5 of each VCV syllable). A pre-test was administered to both child and adult participants to ensure that they understood the task. This test consisted of the congruent endpoints of the synthetic continua presented in random order. There were 10 stimuli in the pre-test (5 per congruent endpoint). No corrective feedback was given during this pre-test.

During the main test, each individual stimulus was presented 10 times each (in random order) to the five-year-old and adult participants. This resulted in 180 stimulus presentations for each of the contrast pairs (ABE-ADE or BEE-DEE). The stimuli within a contrast pair were split into blocks of 10 for presentation. The inter-stimulus interval for the adults was 3 seconds, with an inter-block interval of 10-seconds. Following [33], there was no fixed inter-stimulus interval for the child participants. Instead, the presentation was paused briefly after every stimulus, allowing the children sufficient time to respond. At the end of each block, the children were allowed to choose a small prize (a sticker). Noncontingent encouragement was given to the child subjects throughout the pre-test and the main test.

3. Results

Multilevel logistic regression was used to model listener behaviour [34]. Logistic regression allows us to model listeners' responses ([d] or [g]) from a number of predictors (/e/-C context vs. C-/i/ context; VC#V context vs. V#CV context), and the multilevel structure allows us to fully accommodate data from individual listeners, and to also incorporate a group-level predictor of listener type (child vs. adult). Parameters are estimated as distributions rather than points; the equivalent to a null hypothesis is that a parameter's mean is zero. If the actual estimated value is greater than 1.96 standard deviations from zero, it is equivalent to being statistically significant in the traditional sense, because 95% of the distribution lies below 1.96 standard deviations from the mean. Parameter values that are not significant in that sense are not shown in the models below (they are assumed to be zero and the corresponding term in the model drops out).

For each listener, each contrast pair (ABE-ADE or BEE-DEE) engendered two S-shaped response curves. Figure 1 shows models of these curves for all adults and children (note that modelled curves may represent a stretched S-shape, or only a portion of an S-shape). For purposes of visual analysis, the *slope* of each curve can be seen as resulting from the perceptual weight assigned to the acoustic cue that changes *along* the corresponding continuum—here the /i/-onset transitions—while the degree of *separation* of two curves in a pair can be seen as resulting from the weight given to the acoustic cue that changes *across* the two continua—here the /e/-offset transitions.

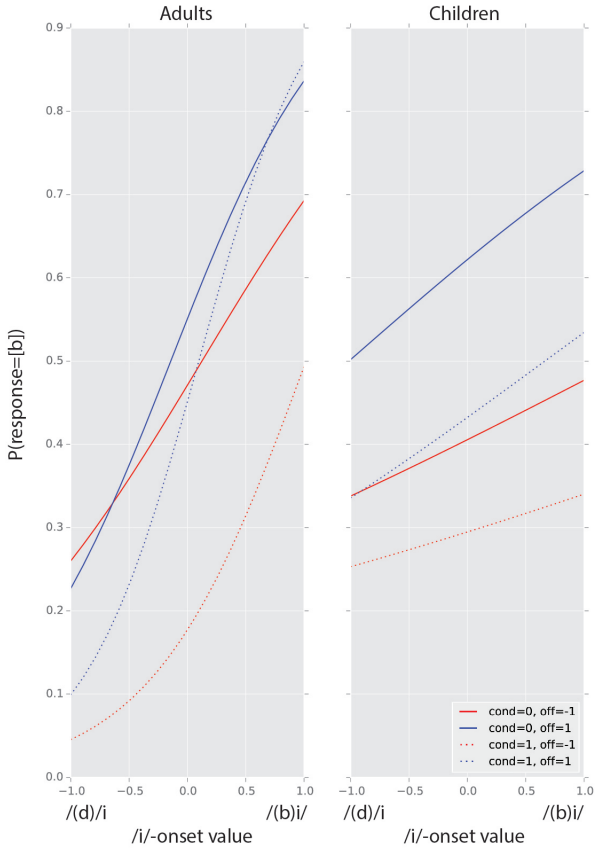

Figure 1: *Regression curves at the model level for adults and children. The x-axis corresponds to point along the /i/-onset transition continuum. The y-axis corresponds to number of [b] responses: solid curves (cond=0) correspond to stimuli presented in the ABE-ADE syllable boundary condition, and dotted curves (cond=1) correspond to stimuli presented in the BEE-DEE condition. Red curves (off=-1) correspond to stimuli with /e(d)/-appropriate offset transitions, and blue curves (off=1) correspond to stimuli with /e(b)/-appropriate offset transitions.*

3.1. Influence of /e/-transitions vs. /i/-transitions

The separation and slopes of listener response curves shown in Fig. 1 indicate that overall, adult listeners were more influenced by manipulations of /i/ transitions than by manipulations of /e/ transitions, whereas children were more influenced by manipulations of /e/ transitions.

The multilevel logistic regression analysis supports this interpretation. In the model, each individual listener has its own set of coefficients, but all parameters must be constrained by the modelled group behaviour (i.e. adult or child): this allows for the combination of individual and group effects. Each coefficient is modelled as a default or baseline for the non-adult (i.e., child) condition, plus a second coefficient that accounts for the difference when the response is for an adult. This model gives the following result: $P(y_i = 1) = [0.49 + 1.03 \times a] \times \mathbf{on} + [0.58 + -0.4 \times a] \times \mathbf{off} + [-1.01] \times \mathbf{cond} + [0.74 \times a] \times \mathbf{off} \times \mathbf{cond}$. Note that for this and all following models: y_i is the response (either [b] or [d]) to one presentation of a stimulus, **on** is the co-

efficient for the /i/-onset transitions, **off** is the coefficient for the /e/-offset transitions and **cond** is the coefficient for the syllable boundary condition (either ABE-ADE or BEE-DEE), a is adult (thus $1 =$ adult, $0 =$ child), and if $a = 0$, some terms drop out of the model.

We see from this that the /e/-offset transition coefficients are 0.58 for a child listener, and $0.58 - 0.4 = 0.18$ for an adult listener. The /i/-onset transition coefficients are 0.49 for a child listener, and $0.49 + 1.03 = 1.52$ for an adult listener. Therefore, as suggested by the response curves, adults are *not* more affected by /e/-offset transitions than by /i/-onset transitions, as might be expected given the relative acoustic informativeness of these two cues. Instead adults are more affected by /i/-onset transitions. Children are weakly more affected by /e/-offset transitions than by /i/-onset transitions. On the assumption that the /i/ transitions were less acoustically salient than the /e/ transitions (/i/ transitions: 30 ms, 232 Hz difference in F2 offset values; 264 Hz difference in F3 offset values; /e/ transitions: 20 ms, 713 Hz difference in F2 offset values, 93 Hz difference in F3 offset values), these results are consistent with the view that children respond more to acoustic salience, whereas adults respond more to CV transitions.

3.2. Adult listeners

Slopes and separation of response curves (see Fig. 1) show that adults are more responsive to /e/ and /i/ transitions cueing word-initial stop place of articulation in the ABE-ADE stimuli than to the same transitions when they cued word-final stop place of articulation in the BEE-DEE stimuli.

In the model built to account for adult listener behaviour the results are presented as a default condition (COND=0), which in this case is the ABE-ADE condition and a non-default condition (COND=1, or BEE-DEE), therefore when examining the non-default BEE-DEE condition it is necessary to look at the interaction terms ($\times$ COND). This model produces the following result: $P(y_i = 1) = 0.21 + 1.18 \times$ **on** $+ 0.16 \times$ **off** $+ -0.9 \times$ **cond** $+ 0.25 \times$ **on** $\times$ **off** $+ 0.58 \times$ **on** $\times$ **cond** $+ 0.51 \times$ **off** $\times$ **cond** (see above for definition of coefficients).

This model tells us that when **cond** $= 0$ (i.e., when the syllable boundary condition was ABE-ADE) , the effect on adults' responses of the /e/-offset transitions is $0.16/4 = 0.04$, and the effect of the /i/-onset transitions is $1.18/4 = 0.29$ (note that the above coefficient values are divided by 4 to reflect the approximate influence on probability: at the maximum of the derivative of the invlogit function it has a value of $\beta/4$). When **cond** $= 1$ (BEE-DEE), the effect of the /e/-offset transitions is increased by 0.51 compared to its effect in the ABE-ADE condition, which means that the /e/-offset effect is now $(0.16 + 0.51)/4 = 0.17$. This term is significant. The effect of the /i/-onset transitions in the BEE-DEE condition is increased by 0.58 compared to its effect in the ABE-ADE condition, which means that the /i/-onset effect is now $(1.18 + 0.58)/4 = 0.44$. This term is significant.

In summary, adult listeners show a much greater effect of /e/-offset transitions in the (word-initial, unpredictable) BEE-DEE condition than in the (word-final, predictable) ABE-ADE condition. Similarly, adults show a much greater effect of /i/-onset transitions in the BEE-DEE context than in the ABE-ADE condition.

3.3. Child listeners

Response curve separation and slope as shown in Fig. 1 show that, in contrast to adults, children are marginally less responsive to /e/ and /i/ transitions cueing word-initial stop place of articulation in the ABE-ADE stimuli than to the same transitions cueing word-final stop place of articulation in the BEE-DEE stimuli.

In the model of child listener behaviour the results are, as above, presented as a default condition (COND=0: ABE-ADE) and a non-default condition (COND=1: BEE-DEE). This model produces the following result: $P(y_i = 1) = 0.24 + 0.39 \times$ **on** $+ 0.44 \times$ **off** $+ -0.63 \times$ **cond** $+ 0.1 \times$ **on** $\times$ **off** $+ -0.08 \times$ **on** $\times$ **cond** $+ -0.14 \times$ **off** $\times$ **cond**.

From this we observe that when **cond** $= 0$ (ABE-ADE) the effect of the /e/-offset transitions on the children's responses is $0.44/4 = 0.11$, and the effect of the /i/-onset transitions was $0.39/4 = 0.09$. When **cond** $= 1$ (BEE-DEE) the effect of the /e/-offset transitions is decreased by 0.14 compared to its effect in the ABE-ADE condition, which means that the /e/-offset effect is now $(0.44 - 0.14)/4 = 0.07$. This term is significant, but tiny. The effect of the /i/-onset transitions in the BEE-DEE condition is decreased by 0.08 compared to its effect in the ABE-ADE condition, meaning that the /i/-onset effect is now $(0.39 - 0.08)/4 = 0.02$. This term is not significant.

In summary, there is a small difference across syllable boundary placement in the effect of /e/-offset transitions: children are—weakly—more affected by /e/-offset transitions in the (word-final, predictable) ABE-ADE context than in the (word-initial, unpredictable) BEE-DEE context. There is no difference in the effect of /i/-onset transitions on children's responses across the two syllable boundary conditions.

4. Summary and discussion

Our results show that adults and children responded differently to /e/ and /i/ transitions in "Abe E.Ade E." and "A Bee/A Dee" stimuli. Adults were more influenced by transitions cueing word-initial consonants than by the exact same transitions cueing word-final consonants, whereas children showed a more balanced pattern, with marginally more influence of transitions cueing word-final consonants. Our proposal is that 5-year-old children have not yet learned about the relative predictability of word-initial vs. word-final consonants in English, and therefore are "equal opportunity" listeners who give approximately equal weight to cues to word-initial and word-final consonants. On this view, children have not yet learned that they can usually identify word-final consonants on the basis of stored, top-down, lexical information, and can afford to pay less attention to cues to these consonants. In contrast, adults who have learned that word-initial consonants are less predictable than word-final consonants in English, pay more attention to cues for consonants that are less predictable in their language. In addition, we found that overall, adults were more influenced by /i/ transitions (CV transitions in our stimuli), as compared to /e/ transitions, whereas children were more influenced by /e/ transitions. These results are consistent with existing findings that children give less weight than do adults to some less distinctive or acoustically salient cues [10].

5. References

[1] S. E. Krause, "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults," *Journal of the Acoustical Society of America*, vol. 71, no. 4, pp. 990–995, 1982.

[2] 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, pp. 229–238.

[3] C. Mayo, J. M. Scobbie, N. Hewlett, and D. Waters, "The influence of phonemic awareness development on acoustic cue weighting in children's speech perception," *JSLHR*, vol. 46, pp. 1184–1196, 2003.

[4] 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," *JASA*, vol. 115, pp. 3184–3194, 2004.

[5] ——, "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception," *JASA*, vol. 118, pp. 1730–1741, 2005.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] C. Wardrip-Fruin and S. Peach, "Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants," *Language and Speech*, vol. 27, no. 4, pp. 367–379, 1984.

[12] J. Watson, "Sibilant-vowel coarticulation in the perception of speech by children with phonological disorder," Ph.D. dissertation, Queen Margaret College, Edinburgh, 1997.

[13] S. Nittrouer, "The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM," *Journal of Speech and Hearing Research*, vol. 39, no. 5, pp. 1059–1070, 1996.

[14] R. N. Aslin, D. B. Pisoni, B. L. Hennessy, and A. J. Perey, "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," *Child Development*, vol. 52, pp. 1135–1145, 1981.

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

[16] J. M. McQueen, M. D. Tyler, and A. Cutler, "Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds," *Language Learning and Development*, vol. 8, pp. 317–339, 2012.

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

[18] J. L. Metsala and A. C. Walley, "Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability," in *Word Recognition in Beginning Literacy*, J. L. Metsala and L. C. Ehri, Eds. Hillsdale, NJ: Erlbaum, 1998, pp. 89–120.

[19] 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. London: Academic Press, 1987, pp. 67–84.

[20] S. E. Trehub, "The discrimination of foreign speech contrasts by infants and adults," *Child Development*, vol. 47, pp. 466–472, 1976.

[21] J. F. Werker, J. H. V. Gilbert, K. Humphrey, and R. C. Tees, "Developmental aspects of cross-language speech perception," *Child Development*, vol. 52, pp. 349–355, 1981.

[22] J. F. Werker and R. C. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behaviour and Development*, vol. 7, pp. 49–63, 1984.

[23] M. Aylett, "Stochastic suprasegmentals - relationships between redundancy, prosodic structure and care of articulation in spontaneous speech," Ph.D. dissertation, University of Edinburgh, 2000.

[24] M. Aylett and A. Turk, "The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and Speech*, vol. 47, pp. 31–56, 2004.

[25] A. Turk, "Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis," *Laboratory Phonology*, vol. 1, no. 2, pp. 227–262, 2010.

[26] M. A. Redford and R. L. Diehl, "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," *Journal of the Acoustical Society of America*, vol. 106, pp. 1555–1565, 1999.

[27] D. Norris and J. M. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological Review*, vol. 115, pp. 357–395, 2008.

[28] P. A. Luce and D. B. Pisoni, "Recognising spoken words: The neighbourhood activation model," *Ear and Hearing*, vol. 19, no. 1, pp. 1–36, 1998.

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

[30] L. M. Dunn, L. M. Dunn, C. Whetton, and J. Burley, *British Picture Vocabulary Test*, 2nd ed. Berkshire, UK: NFER-NELSON, 1997.

[31] Sensimetrics Corp., *SenSyn: Speech Synthesizer Package*, Cambridge, MA.

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

[33] A. C. Walley and T. D. Carrell, "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," *Journal of the Acoustical Society of America*, vol. 73, pp. 1011–1022, 1983.

[34] C. Mayo, F. Gibbon, and R. A. J. Clark, "Phonetically trained and untrained adults' transcription of place of articulation for intervocalic lingual stops with intermediate acoustic cues," *JSLHR*, vol. 56, pp. 779–791, 2013.