Learning to perceptually organize speech signals in native fashion[a)]

Susan Nittrouer[b)] and Joanna H. Lowenstein
Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Road, Suite 4000, Columbus, Ohio 43212

(Received 12 June 2009; revised 30 October 2009; accepted 31 December 2009)

The ability to recognize speech involves sensory, perceptual, and cognitive processes. For much of the history of speech perception research, investigators have focused on the first and third of these, asking how much and what kinds of sensory information are used by normal and impaired listeners, as well as how effective amounts of that information are altered by "top-down" cognitive processes. This experiment focused on perceptual processes, asking what accounts for how the sensory information in the speech signal gets organized. Two types of speech signals processed to remove properties that could be considered traditional acoustic cues (amplitude envelopes and sine wave replicas) were presented to 100 listeners in five groups: native English-speaking (L1) adults, 7-, 5-, and 3-year-olds, and native Mandarin-speaking adults who were excellent second-language (L2) users of English. The L2 adults performed more poorly than L1 adults with both kinds of signals. Children performed more poorly than L1 adults but showed disproportionately better performance for the sine waves than for the amplitude envelopes compared to both groups of adults. Sentence context had similar effects across groups, so variability in recognition was attributed to differences in perceptual organization of the sensory information, presumed to arise from native language experience. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3298435]

PACS number(s): 43.71.Ft, 43.71.Hw, 43.71.An [MAH] Pages: 1624–1635

I. INTRODUCTION

The ability to recognize speech involves sensory, perceptual, and cognitive processes. Regarding the first of these, listeners must have some amount of access to the acoustic signal generated by the moving vocal tract of the speaker. Sensory impairments such as hearing loss can inhibit speech recognition if the impairment sufficiently precludes access to that sensory information. Research involving listeners with hearing loss has commonly focused on this aspect of perception, with questions explicitly addressing how much of the signal and what properties are needed for speech perception (e.g., Dorman *et al.*, 1985; Erber, 1971; Leek *et al.*, 1987; Revoile *et al.*, 1985; Stelmachowicz *et al.*, 1990). But while having access to sensory information in the speech signal is an obvious prerequisite to recognition, it is not sufficient to explain speech perception. This was demonstrated by Surprenant and Watson (2001), who tested the speech recognition of college students and also examined their discrimination for a number of related non-speech stimuli. Performance on the two kinds of tasks were not strongly correlated, leading the authors to conclude that "…factors higher in the sequence of processing than the auditory periphery account for significant variance in speech recognition—by both normal hearing and by hearing-impaired listeners." (p. 2094). The experiment reported here was designed to test the hypothesis that part of that variance is explained by how listeners perceptually organize the sensory information in the speech signal.

The idea that perceivers need to organize the sensory information reaching them is well accepted in the study of visual perception where displays such as the classic Ruben's vase illustrate the phenomenon: In that demonstration, viewers recognize either a vase or two symmetrical profiles depending on how they organize the sensory information that reaches their retinas. Scientists studying general auditory perception similarly examine the principles that underlie the recognition of the object responsible for generating the sensory information (Binder *et al.*, 2004; Bregman, 1990; Dau *et al.*, 2009; Griffiths and Warren, 2004). When it comes to speech perception, however, considerably less attention has been paid to how listeners perceptually organize the components of the signal in order to recover the sound-generating object, perhaps due to long-standing controversy over whether listeners actually ever do recover that object, which is the moving vocal tract (e.g., Liberman *et al.*, 1962; Lotto *et al.*, 2009). Be that as it may, the work that has been done on perceptual organization for speech demonstrates that listeners must integrate disparate spectral and temporal signal components appropriately in order to recover cohesive and linguistically relevant percepts (e.g., Best *et al.*, 1989; Remez *et al.*, 1994).

The work reported here follows years of investigation examining how the weighting of explicit acoustic cues to phonetic identity changes as children acquire experience with a native language. Acoustic cues are defined as spectrally discrete and temporally brief pieces of the speech sig-

[a)]Portions of this work were presented at the 154th Meeting of the Acoustical Society of America, New Orleans, LA, November 2007.
[b)]Author to whom correspondence should be addressed. Electronic mail:nittrouer.1@osu.edu

 0001-4966/2010/127(3)/1624/12/$25.00

TABLE I. Mean weighting coefficients for formant transitions and fricative-noise spectra for listeners in each age group. Group numbers are in italics, standard deviations are in parentheses (from Nittrouer and Lowenstein, 2009).

	3-year-olds (*23*)	5-year-olds (*22*)	7-year-olds (*20*)	Adults (*20*)
Formant transition	0.50 (0.15)	0.48 (0.19)	0.32 (0.15)	0.30 (0.17)
Fricative noise	0.66 (0.13)	0.71 (0.14)	0.81 (0.11)	0.82 (0.11)

nal that affect listeners' phonetic labeling when modified (Repp, 1982). Several cues can underlie a single phonetic distinction, and the way those cues are weighted varies depending on native language experience (e.g., Beddor and Strange, 1982; Crowther and Mann, 1994; Flege *et al.*, 1996; Gottfried and Beddor, 1988; Zampini *et al.*, 2000). Accordingly, children must discover the optimal weighting strategies in their native language, and their early weighting strategies differ from those of adults. In particular, children weight dynamic, spectral cues (i.e., formant transitions) more than adults, and weight stable spectral cues (e.g., noise spectra) less (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987; Watson, 1997). Some evidence of this developmental shift comes from a study in which adults and children between 3 and 7 years of age were asked to label syllable-initial fricatives based on the spectral shape of fricative noises and on formant transitions at voicing onset (Nittrouer and Lowenstein, 2009). Computed weighting coefficients are shown in Table I. They indicate that the weighting of formant transitions diminished with increasing age, while the weighting of fricative-noise spectra increased.

Of course, formant transitions are just brief sections of more broadly modulating resonant frequencies created by vocal-tract cavities, which are continually changing in shape and size. Results across studies demonstrating children's reliance on formant transitions for phonetic labeling led to the broader hypothesis that children would show a perceptual advantage in organizing this kind of signal structure over other kinds of structure. To test that hypothesis, Nittrouer *et al.* (2009) presented two kinds of processed signals to three groups of listeners. Sine wave replicas of speech (Remez *et al.*, 1981) were used to provide a signal condition that preserved dynamic spectral structure, just the sort of structure for which children are expected to show an advantage. Vocoded stimuli (Shannon *et al.*, 1995) were also used as a comparison condition. These stimuli preserve amplitude structure across the utterance in a preselected number of frequency bands. The sentences in that study were all four words long with English syntax, but no semantic constraints. Two groups of adult listeners participated: native speakers of American English and native speakers of another language who were excellent second-language (L2) speakers of American English. Although native speakers of any language could presumably have been used, Mandarin Chinese speakers were selected because Fu *et al.* (1998) showed that they recognize vocoded Chinese sentences as well as native English speakers recognize vocoded English sentences. This group of adults was included both to test the idea that the perceptual organization of sensory information is language specific, and to serve as controls for the third group of listeners: 7-year-old native speakers of American English. If native English-speaking 7-year-olds had poorer recognition than L1 adults for either kind of processed stimuli, the inclusion of L2 adults would mitigate against the explanation that age-related differences in processing at the auditory periphery accounted for the outcome, assuming that L2 adults also performed more poorly than L1 adults.

Results showed that the L2 adults performed more poorly with both kinds of stimuli than the L1 adults, suggesting that how listeners perceptually organize acoustic structure in the speech signal is at least somewhat dependent on native language experience. That interpretation was motivated by outcomes of other studies. For example, Boysson-Bardies *et al.* (1986) showed that the long-term spectral shape of speech signals varies across languages, with some having more prominent spectral peaks and others having flatter spectra. The fact that listeners are sensitive to the spectral shape of their native language was demonstrated by Van Engen and Bradlow (2007), who reported that speech-shaped noise degraded speech recognition more when the shape of the noise was derived from speakers of the language in which testing was occurring, rather than from a different language. Apparently the way in which speech is produced creates structure in signals at global levels (i.e., longer than the phonetic segment and spectrally broad), and listeners are sensitive to how signals are structured at those levels in their native language.

The 7-year-old children in Nittrouer *et al.* (2009) were poor at recognizing sentences that preserved only amplitude structure, scoring similarly to the L2 adults. At the same time, 7-year-olds performed similarly to the L1 adults with the sine wave replicas. In fact, the 7-year-olds were the only group of listeners to show better performance for the sine wave stimuli than for the four-channel amplitude envelope (AE) stimuli (25.65% vs 13.68% correct, respectively); L1 and L2 adults had statistically identical recognition scores for the four-channel amplitude envelope and the sine wave stimuli: 28.17% correct for L1 adults across the two conditions and 9.99% correct for L2 adults. Of course, 7-year-olds' performance could have been rendered equivalent across conditions by increasing the number of channels in the amplitude envelope stimuli, but the critical outcome was precisely that they performed differently on two sets of data that evoked similar responses from adults. However, a caveat in those results was that different listeners participated in testing with the two kinds of signals, so within-group comparisons could not be made. That problem was corrected in the current study.

The perceptual advantage demonstrated by 7-year-olds in that experiment for sine wave stimuli was viewed as complimentary to earlier findings showing that children weight formant transitions more than adults in phonetic decisions. Speech processed to preserve only the amplitude envelopes of the original signal do not preserve that kind of dynamic spectral structure particularly well, whereas sine wave replicas do. Consequently, how well listeners of different ages can perceptually organize various kinds of signals seems partly related to the signal properties themselves. If the processed signal preserves properties to which listeners are especially sensitive, those listeners are better able to recover the auditory (or in the case of speech, the linguistic) object.

The current study was conducted to extend Nittrouer *et al.* (2009) by examining how children younger than 7 years of age, the youngest group in that study, perceptually organize processed speech signals. Because children younger than 7 years of age have consistently demonstrated even stronger weighting of dynamic spectral components in the speech signal, they should show even a stronger advantage in their recognition of the sine wave replicas over the amplitude envelope signals. However, the 7-year-olds in that study recognized fewer than 30% of the words correctly in two out of three conditions, and so we worried that younger children might not recognize any words at all if the same or similar stimuli were used. Therefore, we elected to use sentences that provided both syntactic and semantic constraints. Children as young as 3 years of age participated in this experiment. The goal was to examine listeners' abilities to perceptually organize the sensory information provided by dynamic spectral structure (as preserved by sine waves) and by amplitude envelopes. Four-channel amplitude envelope signals were used because adults had shown identical performance with those signals to that obtained with sine wave speech in Nittrouer *et al.* (2009). Native Mandarin-speaking adults participated again because their inclusion could help bolster arguments that age-related differences in speech recognition, if found, are likely not due to differences in auditory processing of the sensory information.

Our focus in this experiment was on the perceptual organization of linguistically relevant acoustic structure in the speech signal. Because we used sentence materials, however, we needed a way to ensure that any group differences observed were not due to linguistic context effects. Miller *et al.* (1951) are generally credited with being the first investigators to examine differences in speech recognition scores due to the context, in which the stimulus is presented. They showed that when isolated words presented in background noise were part of small sets made known to research participants, recognition was more accurate than when set size was larger. As their metric they used the difference in signal-to-noise ratio (SNR) needed to achieve a preselected target of 50% correct recognition. It was not long before other investigators extended the notion of set size to frequency of word usage: Commonly occurring words can be recognized at poorer SNRs than uncommonly occurring words (e.g., Broadbent, 1967; Howes, 1957; Rosenzweig and Postman, 1957). In the rhetoric of information theory, word frequency influences the number of effective channels of sensory information available to the perceiver.

Boothroyd (1968) developed several metrics that provide a way of quantifying the numbers of effective channels of information used by the perceiver. The metric of most relevance to the current work is known as the *j* factor and derives from the fact that the probability of recognizing a whole word is dependent on the probabilities of recognizing the separate parts (or phonemes) forming that word. If lexicality played no role in recognition, then the probability of recognizing a whole word correctly would be directly related to the probability of recognizing each of the parts, such that

$$p_w = p_p^n, \tag{1}$$

where p_w is the probability of recognizing the whole word, p_p is the probability of recognizing each part, or phoneme, and n is the number of parts in the whole. However, this relation holds only when it is necessary to recognize each part in order to recognize the whole, as happens with nonwords, for instance. When listening to real words, the lexical status of those words generally reduces the need to recognize each separate part in order to recover the whole. Therefore, we can change n to a dimensionless exponent, such as j, to quantify the effect of hearing these phonemes in a lexical context. This exponent varies between 1 and n, and the extent to which it is smaller than n indicates the magnitude of the context effect. Now Eq. (1) can be changed to

$$p_w = p_p^j, \tag{2}$$

where j is the number of independent channels of information, and we can solve for the effective number of channels of information in the word with

$$j = \log(p_w)/\log(p_p). \tag{3}$$

Boothroyd and Nittrouer (1988) demonstrated the validity of the j factor with the finding that j equaled 3.07 when listeners were asked to recognize nonsense CVCs. That value was not statistically different from the value of 3, which would be expected if listeners need to recognize each segment independently in order to recognize the whole syllable. By contrast, Boothroyd and Nittrouer (1988) reported a mean j of 2.46 when these same listeners were asked to recognize CVC real words. That value was statistically different from the 3.07 obtained for nonsense syllables. Together, these findings support the use of the j factor as a metric of the number of independent channels of information needed for recognition. This metric has been used by others for several clinical purposes such as assessing the abilities of deaf children to use lexical context (Boothroyd, 1985). Benkí (2003) used the j factor to index the lexical context effect of high-frequency words compared to low-frequency words. In addition to demonstrating a stronger context effect for high- over low-frequency words (j's of 2.25 and 2.46, respectively), Benkí (2003) replicated Boothroyd and Nittrouer's finding of a j equal to 3.07 for nonsense syllables.

Sentence context similarly improves listeners' abilities to recognize words under conditions of signal degradation (e.g., Boothroyd and Nittrouer, 1988; Duffy and Giolas,

1974; Giolas *et al.*, 1970; Kalikow *et al.*, 1977; Nittrouer and Boothroyd, 1990). When examining the effects of sentence context on word recognition, investigators are generally not interested in the effect of lexicality, but rather in the effects of syntactic and semantic constraints. Those effects can be quantified using Eq. (3) by making p_w the probability of recognizing the whole sentence and p_p the probability of recognizing each word in the sentence.

Nittrouer and Boothroyd (1990) computed the number of independent channels of information needed by adults and children (4–6 years of age) listening to sentences in their native language with signals degraded by the addition of noise at two SNRs: −3 dB and 0 dB. Two kinds of sentences were used: sentences that had appropriate syntactic structure, but no useful semantic information (syntax-only sentences), and sentences that provided clear syntactic structure, as well as strong semantic constraints (syntax+semantics sentences). All were comprised of four monosyllabic words. Across the two SNRs employed in that study, the mean *j* for adults listening to the syntax+semantics sentences was 2.32, and the mean *j* for children listening to those same sentences was 2.59. This age-related difference was not statistically significant, and so it may be concluded that children as young as 4 years are able to use syntactic structure and semantic context as well as adults, at least for simple sentences.

The *j* factor has not been used to assess the contributions of sentence context to recognition by adults listening to a second language, but several investigators have used the speech in noise (SPIN) test to evaluate this effect. In the SPIN test, listeners hear two kinds of sentences and must report the last word in each. In one kind of sentence, the words prior to the last word provide semantic information that strongly predicts that word (e.g., *My clock was wrong so I got to school late; After my bath I dried off with a towel.*). In the other kind of sentence, the preceding words provide no information that would especially predict the final word (e.g., *Mom talked about the pie. He read about the flood.*). Mayo *et al.* (1997) presented these sentences at several SNRs to L1 English listeners and L2 listeners who had learned English at ages varying from birth to early adulthood. Results showed that listeners who had learned English after the age of 14 years were less capable than native and early L2 learners at using the highly predictable sentence context to aid recognition of the final word. In another study, Bradlow and Alexander (2007) presented these same sentences to native and L2 English listeners, but used samples that could be described as plain or clear speech. Listeners in both groups showed more accurate recognition for words in the high-predictability contexts, but for L2 listeners this advantage was restricted to clear sentences.

The results above suggest that L2 adults are poorer at using sentence context for speech recognition than L1 adults. However, those results were obtained using the SPIN test, and there are important differences between that task and the way *j* factors are computed. In the SPIN test, recognition of only words in sentence-final position is being evaluated, and the comparison is between sentences that have strong semantic constraints (high predictability) and those that have neutral semantic constraints (low predictability). With *j* factors, recognition scores for all words within the sentences are involved in computation, and every word serves as context for every other word. The difference between these two methods is highlighted by the fact that young children show diminished semantic context effects on the SPIN test compared to older children (Elliott, 1979), but young children have *j* factors equivalent to those of adults, as long as syntax and semantics are within their language competencies (Nittrouer and Boothroyd, 1990). On the SPIN test, sentences with high and low predictabilities tend to differ in syntactic structure, with high-predictability sentences generally having more complex structures. That could constrain the abilities of children and L2 listeners to use the sentence context if the syntactic constructions are not within their competencies. This is not a concern when *j* factors are computed, both because the sentences that have been used involve simple syntactic constructions and because wholes and parts are scored from the same responses. Because of these differences, we did not necessarily expect L2 listeners to show the same deficits in using sentence context compared to L1 listeners that have been reported earlier.

In summary, the current experiment was designed to examine how listeners of different ages and native-language experience perceptually organize two kinds of signals derived from speech: sine wave replicas and amplitude envelopes. We hypothesized that L2 adults would be equally poor at recognition with both kinds of signals, but that children would show an advantage for the sine wave replicas over the amplitude envelope signals because they heavily weight global spectral structure in their speech perception. We selected sentence materials that we thought would produce similar context effects across listeners, but planned to compute *j* factors as a test of that assumption.

II. METHOD

A. Participants

100 listeners participated in this experiment: 80 native speakers of American English (20 adults, 20 7-year-olds, 20 5-year-olds, and 20 3-year-olds) and 20 adult, native speakers of Mandarin Chinese who were competent L2 speakers of English. None of the participants had listened to sine wave replicas or amplitude envelope speech before. All L2 speakers had started learning English between the ages of 12 and 14 years in language classes in China. They all came to the United States between the ages of 21 and 26 years to study, and were in the United States as graduate students, postdoctoral fellows, or junior faculty members. Mean age of the L1 adults was 24 years (range 19–35 years), and mean age of the L2 adults was 26 years (range 22–36 years). To participate, all listeners had to pass a hearing screening and demonstrate appropriate abilities to comprehend and produce spoken English. 3-year-olds were also given the Goldman–Fristoe 2 test of articulation (Goldman and Fristoe, 2000) to screen for speech problems. They were required to score at or better than the 30th percentile for their age. Each group was comprised of 10 males and 10 females, except for 7-year-olds (9 females, 11 males) and L2 listeners (11 females, 9 males).

B. Equipment

All speech samples were recorded in a sound booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Laboratories Soundblaster analog-to-digital converter. Perceptual testing took place in a sound booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Samson headphone amplifier, and AKG-K141 headphones.

C. Stimuli

Seventy-two sentences were used as stimuli in this experiment: 12 for practice and 60 for testing. All were selected from the hearing in noise test for children (HINT-C) (Nilsson *et al.*, 1996; Nilsson *et al.*, 1994), and all were five words in length. The sentences selected for use are shown in the appendix. The HINT-C sentences were originally developed to test how well children can recognize speech in competing noise. They are five to eight words long, syntactically correct, and follow a subject-predicate structure. They are highly predictable semantically. Sentences from the HINT-C were used by Eisenberg *et al.* (2000) in an experiment looking at recognition of amplitude envelopes by adults and children. Those listeners, regardless of age, had close to 90% correct word recognition when listening to eight-channel envelope stimuli. With four-channel stimuli, adults had close to 70% correct word recognition and 5- to 7-year-olds had close to 35% correct word recognition. Thus, although our primary motivation for using four channels for the amplitude envelope stimuli was because Nittrouer *et al.* (2009) found that adults had equivalent word recognition for those stimuli and sine wave replicas, the Eisenberg *et al.* (2000) results supported the expectation that word recognition scores would be neither too close to 0 nor too close to 100%, at least for the amplitude envelope stimuli. Finally, we restricted our corpus to five-word sentences so that length would be consistent, and relatively brief, thereby minimizing concerns about memory.

The selected sentences were recorded at a 44.1-kHz sampling rate with 16-bit digitization by an adult male speaker of American English who is a trained phonetician. All sentences were equalized for mean rms amplitude across sentences before any processing was done, and all sentences were used to create sine wave (SW) and four-channel envelope signals. Because these latter stimuli preserve the amplitude envelopes in each of four channels, they are dubbed as AE stimuli in this report. For creating the SW stimuli, a PRAAT routine written by Darwin (http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS) was used to extract the center frequencies of the first three formants. Stimuli were generated from these formant frequencies, and spectrograms were compared to spectrograms of the original sentences to ensure that the trajectories of the sine waves matched those of the formant frequencies.

To create the AE stimuli, a MATLAB routine was written. All signals were first low-pass filtered with an upper cut-off frequency of 8000 Hz. For the four-channel stimuli, cut-off frequencies between bands were 800, 1600, and 3200 Hz. Each channel was half-wave rectified, and results used to modulate white noise, limited by the same band-pass filters as those used to divide the speech signal into channels. Resulting bands of modulated noise were low-pass filtered using a 160-Hz high-frequency cut-off, and combined.

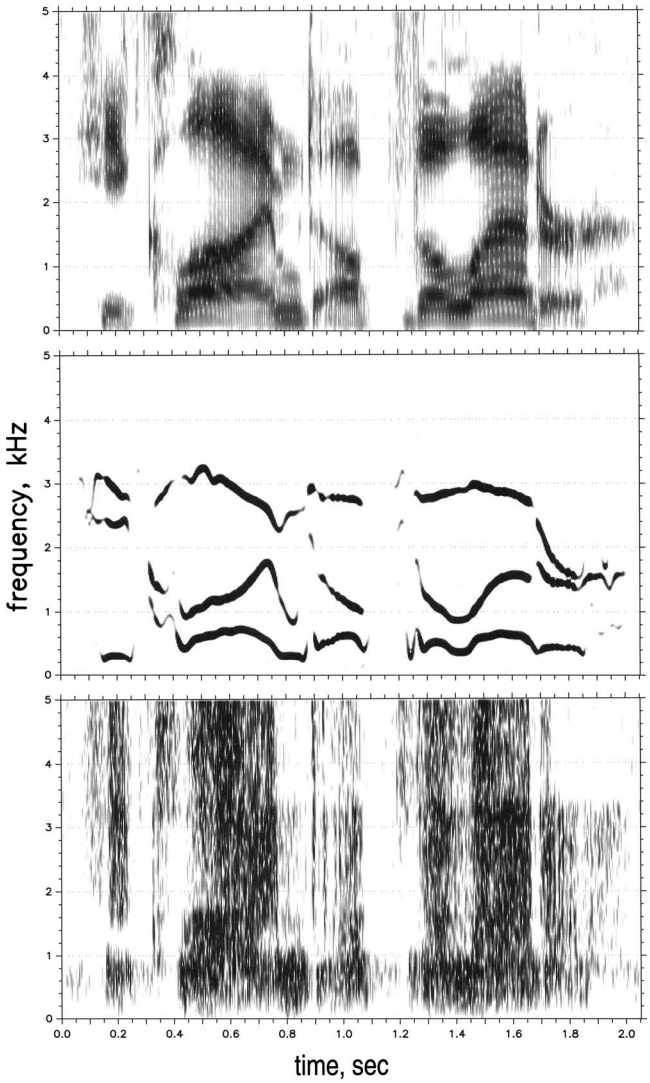

FIG. 1. Spectrograms of the sentence "He climbed up the ladder." (Top) Natural speech sample from which SW and AE stimuli were derived. (Middle) SW stimulus. (Bottom) Four-channel AE stimulus.

Figure 1 displays a natural speech sample of *He climbed up the ladder* (top panel), the SW replica (center panel), and the AE replica (bottom panel).

D. Procedures

All stimuli were presented under headphones at a peak intensity of 68 dB sound pressure level. For each participant, the software randomly selected 30 sentences to present as AE and 30 to present as SW prior to testing. Half of the participants heard all 30 AE sentences first, and then the SW sentences, and half of the participants heard all 30 SW sentences first, and then the AE sentences. Training was the same for each condition. For each of the six practice sentences, the

TABLE II. Mean percent correct words recognized by each group in each condition. Standard deviations are in parentheses. AE refers to amplitude envelope stimuli and SW refers to sine wave stimuli.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	79.53 (8.30)	43.85 (16.63)	30.73 (17.93)	16.23 (11.53)	33.47 (12.29)
SW	98.37 (1.38)	91.90 (3.03)	86.32 (6.11)	75.40 (15.22)	77.20 (8.78)

natural version was played first. The listener was instructed to listen and repeat it. Next, the listener was told "Now you will hear the sentence in a robot voice. Listen carefully so you can repeat it." The notion of a robot was invoked to personify the signal so that listeners would more readily hear the stimuli as speech. The children were specifically told that the SW sentences were spoken by a robot with a "squeaky voice," and the AE sentences were spoken by a robot with a "scratchy voice." For both the AE and SW conditions, listeners were asked at the end of the six-sentence practice if they were able to recognize the "robot's" productions as speech. It was made clear that the question was whether they could hear the stimuli as speech, rather than could they understand every word. If a listener either was unable to repeat any of the unprocessed sentences or failed to report hearing the processed signals as speech, he was dismissed. None of the listeners had difficulty hearing the AE sentences as speech. Two additional participants, one 3-year-old and one 5-year-old, were dismissed because they did not report hearing the SW sentences as speech.

During testing, the order of presentation of the sentences was randomized independently for each listener. Each sentence was played once, and the listener repeated it as best as possible. Participants could ask for a sentence to be replayed one time. Nittrouer *et al.* (2009) used three trials of each sentence, but listeners in all groups showed similar improvements across repetitions and means from the first to the third repetition never differed by more than 1 standard deviation. Consequently, the outcomes of that study would not have been different if only one repetition had been used, and so we had no motivation to use more repetitions in this study. The number of incorrect words for each sentence was entered into the program interface during testing. 5- and 7-year-olds moved a game piece along a game board every ten sentences, while 3-year-olds used a stamp to mark a paper grid after every five sentences, in order to maintain attention and give a sense of how much of the task remained.

After hearing the sentences in their processed forms, all sentences were played to listeners in their unprocessed forms. Listeners could get no more than 10% of the words wrong on this task. Data from two additional participants, one L2 adult and one 3-year-old, were excluded from analyses because they made too many errors repeating the unprocessed sentences. None of the other listeners had trouble comprehending the unprocessed sentences.

In addition to entering the number of words repeated incorrectly at the time of testing, participants were audiorecorded. A graduate student later listened to 15% of the recordings, and marked the number of words she heard as incorrect. Reliability coefficients were 1.00 for both AE and SW sentences for listeners of all ages.

III. RESULTS

A. Word recognition

The percentage of words correctly recognized across all sentences within any one condition served as the dependent measure. All statistics were performed using arcsine transforms because some listeners scored above 90% correct in the SW condition and others scored below 10% correct in the AE condition. As a preliminary test, analyses of variance (ANOVAs) were conducted on percent correct recognition scores for each group separately, based on whether they heard AE or SW sentences first. No significant order effect was observed for any group.

Next, the effect of English experience for the L2 adults was examined. Pearson product-moment correlation coefficients were computed between recognition scores for both the AE and SW sentences and the length of time the L2 adults had been speaking English based on when they began language classes. The correlation coefficient for AE sentences was -0.04, and for SW sentences it was -0.13. Neither was significant. In addition, correlation coefficients were computed between the numbers of years these participants had lived in the United States and their recognition scores. In this case, the correlation coefficient for AE sentences was -0.41, and for SW sentences, it was -0.27. Again, neither of these was significant; in fact, slightly negative correlations were found. The lack of an L2 experience effect matches findings by others: If learning starts after puberty, eventual proficiency is variable and unrelated to factors such as age of initial exposure or age of arrival in the new country (e.g., Johnson and Newport, 1989).

Table II shows mean correct word recognition for each group for each kind of stimulus. The greatest differences across groups occurred in the AE condition, with a difference of roughly 63% between the highest scoring group (L1 adults) and the lowest scoring group (3-year-olds). Groups performed more similarly in the SW condition, with a difference of only 23% between the highest and lowest scoring groups (again, L1 adults and 3-year-olds). This difference across conditions was due to 3-year-olds performing much better in the SW than in the AE condition, rather than to a decrement in performance for L1 adults in the AE condition. In fact, adults performed better in the SW than in the AE condition, just less so than children. When a two-way ANOVA was performed on these scores, with group as the between-subjects factor and signal type as the within-subjects factor, both main effects were significant: group, $F(4,95)=68.93$, $p<0.001$, and signal types, $F(1,95)=995.93$, $p<0.001$. Of particular interest, the group

TABLE III. Results of the statistical analyses of percent correct scores for group effects, using arc sine transforms. *F* ratios (shown in bold) are for one-way ANOVAs, with group as the between-subjects factor; t-tests are for *posthoc* comparisons. Degrees of freedom are 4, 95 for the ANOVAs, and 95 for the t-tests. Computed *p* values are shown. Bonferroni corrections for five multiple comparisons require a *p* of less than 0.005 for the comparison to be significant at the 0.05 level. Comparisons meeting this adjusted level are indicated by an asterisk.

	F ratio or *t*	*p*
AE sentences	**49.42**	<0.**001**
L1 adults vs 7-year-olds	7.22	<0.001*
L1 adults vs 5-year-olds	10.16	<0.001*
L1 adults vs 3-year-olds	13.31	<0.001*
L1 adults vs L2 adults	9.17	<0.001*
7- vs 5-year-olds	2.94	0.004*
7- vs 3-year-olds	6.09	<0.001*
7- vs L2 adults	1.95	0.054
5- vs 3-year-olds	3.16	0.002*
5- vs L2 adults	−0.99	NS
3- vs L2 adults	−4.14	<0.001*
SW sentences	**46.24**	<0.**001**
L1 adults vs 7-year-olds	5.03	<0.001*
L1 adults vs 5-year-olds	7.64	<0.001*
L1 adults vs 3-year-olds	11.60	<0.001*
L1 adults vs L2 adults	11.25	<0.001*
7- vs 5-year-olds	2.62	0.010
7- vs 3-year-olds	6.57	<0.001*
7- vs L2 adults	6.23	<0.001*
5- vs 3-year-olds	3.95	<0.001*
5- vs L2 adults	3.61	<0.001*
3- vs L2 adults	−0.34	NS

×signal type interaction was significant, $F(4,95)=12.64$, $p<0.001$, indicating that the variability in magnitude of the signal effect across groups was significant.

Table III shows the results of one-way ANOVAs and *posthoc* comparisons performed on word recognition scores for AE and SW stimuli separately. Significant main effects of group were found for both the AE and SW conditions, so *posthoc* comparisons among groups were examined. For the AE condition, L1 adults performed best, and mean correct recognition differed significantly from every other group. Differences were observed among all children's groups, with 7-year-olds performing the best, followed by 5-year-olds, and then 3-year-olds. Although 7-year-olds had better recognition scores than L2 adults, this difference was not significant once a Bonferroni correction was applied. However, this is the one statistical outcome that differed for untransformed and transformed data: When the ANOVA and *posthoc* comparisons were performed on untransformed scores, this comparison was unambiguously significant, $t(95)=2.38$, $p=0.019$, supporting the assertion that 7-year-olds performed more accurately than L2 adults. Five-year-olds performed similarly to L2 adults, but 3-year-olds performed significantly more poorly.

For the SW stimuli, L1 adults again performed the best, with a mean recognition score that differed significantly from every other group. 7- and 5-year-olds performed similarly, and the difference between their mean scores was not significant when a Bonferroni correction was applied. Both groups performed significantly better than 3-year-olds and L2 adults. 3-year-olds did not differ from L2 adults.

Prior to running this experiment, there was no reason to expect adults to be more accurate in their responses for one condition over the other: AE stimuli can be created with any number of channels, and we had selected four channels precisely because adults in Nittrouer *et al.* (2009) showed identical performance with four-channel AE and SW stimuli. 7-year-olds in that experiment were the only listener group to show better recognition for SW than for AE stimuli, and the difference between the two conditions was 12 percentage points. For this experiment, Table II indicates that, on the whole, listeners performed better with SW sentences than with AE sentences. That general result makes it important to investigate the magnitude of this condition-related difference across listener groups, and Table IV shows means of individual difference scores for each group. The L1 adults demonstrated the smallest difference across conditions, and the youngest children demonstrated the greatest difference. It may be that difference scores for L1 adults were constrained because many of these listeners showed near-perfect accuracy, especially with SW stimuli, but there can be no question that 3-year-olds showed an advantage for SW over AE stimuli. A one-way ANOVA done on these difference scores with group as the between-subjects factor revealed a significant effect, $F(4,95)=25.24$, $p<0.001$. *Posthoc* comparisons showed significant differences between L1 adults and all other groups. The comparison of 7- and 3-year-olds was significant. In addition, L2 adults showed significantly smaller difference scores than either 5- or 3-year-olds. Because the scores of these adults were not constrained by ceiling or floor effects, the finding suggests that the age-related difference for L1 listeners in condition effects is likely not entirely due to L1 adults scoring near the ceiling. Children were disproportionately more advantaged for the SW than the AE stimuli.

B. Top-down effects

Using the formula presented in the Introduction, *j* factors were computed for individual listeners using word and sentence recognition scores. Because this computation requires the use of sentence recognition scores, mean percentages of sentences recognized correctly for each group are shown in Table V, for AE and SW stimuli separately. Table VI shows mean *j* factors for each group. Scores for indi-

TABLE IV. Mean difference scores across the SW and AE conditions. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
Difference	18.83 (8.42)	48.05 (15.58)	55.58 (19.84)	59.17 (13.64)	43.73 (10.25)

TABLE V. Mean percent correct sentences recognized by each group in each condition. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	61.00 (12.66)	19.00 (10.93)	8.50 (8.89)	0.67 (1.37)	7.50 (6.66)
SW	93.50 (4.90)	78.50 (7.45)	63.83 (10.88)	44.33 (18.26)	43.65 (13.91)

vidual listeners were not included if the number of words or sentences recognized correctly was greater than 95% or less than 5%. For the SW sentences, 19 out of 20 L1 adults had better than 95% correct word recognition. For the AE sentences, all 20 3-year-olds had poorer than 5% correct sentence recognition.

One-way ANOVAs were computed on individual *j* factors, for the AE and the SW stimuli separately. For the SW stimuli, scores for L1 adults were excluded. For the AE stimuli, scores for the 3-year-olds were excluded. Neither analysis revealed a significant *F* ratio, so it may be concluded that listeners in all groups had similar context effects.

Table VII shows mean *j* factors only for participants who had word and sentence recognition scores between 5% and 95% correct in both conditions. A two-way ANOVA was computed on these scores, with group as the between-subjects factor and stimulus type as the within-subjects factor. The effect of stimulus type was significant, $F(1,33)=43.27$, $p<0.001$, but not the group effect. The stimulus type × group interaction was not significant. Overall then, these *j* factors were smaller for the AE sentences than for the SW sentences, and that difference was consistent across listener groups.

C. Perceptual organization or top-down effects

The main findings emerging from these data are that the younger children showed an advantage in their perceptual organization of the SW stimuli over the AE stimuli, and the L2 adults were poor at perceptually organizing either type of signal. However, two constraints in the results could dampen enthusiasm for those conclusions: First, L1 adults showed ceiling effects for the SW stimuli. Consequently, the true difference in scores between the SW and AE conditions may not have been obtained for that group. The second constraint was that *j* factors could not be computed for all listeners. Possibly the listeners for whom *j* factors could not be computed accounted for the observed differences in signal effects across groups.

Those concerns can be ameliorated by looking at results for only those participants who provided *j* factors. Mean correct word recognition for these individuals are shown in Table VIII, and when these results are compared to those for all listeners shown on Table II, it is apparent that the same trends across groups are observed. Analyses of covariance performed on these values, using *j* factor as the covariate, resulted in significant group effects, for both the AE stimuli, $F(3,54)=75.88$, $p<0.001$, and the SW stimuli, $F(3,72)=18.89$, $p<0.001$.

Further support for the finding that young children did indeed show an advantage in their perceptual organization of SW stimuli, compared to adults, is obtained by looking across results for 5-year-olds and L2 adults. Table III indicates that listeners in these two groups scored the same with the AE stimuli. Tables VI and VII show that top-down effects were similar for listeners in the two groups. The only result that differs for listeners in the two groups concerns their performance with the SW stimuli: 5-year-olds scored significantly better than L2 adults, as indicated by a significant *posthoc* comparison (Table III). Even when scores are examined for only those 5-year-olds and L2 adults for whom *j* factors could be computed (Table VIII), it is apparent that listeners in both groups performed similarly for the AE stimuli, but 5-year-olds scored better with the SW stimuli.

Finally, it is important to the conclusions reached here that context effects were less for the signal condition in which listeners scored better: that is, the SW condition. That outcome indicates that it was not simply the case that something about the SW stimuli permitted stronger context effects. Instead, the enhanced recognition that young children showed for the SW stimuli seem to be due to their abilities to perceptually organize these signals more handily.

IV. DISCUSSION

The experiment reported here was conducted largely to examine whether the way in which listeners perceptually organize speech signals depends on their native language experience, and on the amount of that experience. An additional question concerned whether listeners are better able to recover relevant linguistic structure with some kinds of processed signals than with others: In particular, we asked whether listeners are better able to recover structure from signals that preserve sensory information to which they are especially sensitive.

TABLE VI. Mean *j* factors for the AE and SW conditions. N/A signifies that participants in that group either all scored above 95% correct (adults, SW words) or below 5% correct (3-year-olds, AE sentences). Standard deviations are in parentheses. The numbers of participants with *j* scores for each group are given in italics.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	2.26 (0.34), *20*	2.26 (0.55), *18*	2.43 (0.60), *10*	N/A	2.43 (0.48), *11*
SW	N/A	2.98 (0.94), *17*	3.36 (0.92), *20*	3.43 (0.82), *20*	3.42 (0.53), *20*

TABLE VII. Mean *j* factors for the AE and SW conditions for participants who had *j* scores for both conditions. The numbers of participants are given in italics under the group name. Standard deviations are in parentheses.

	7-year-olds (*15*)	5-year-olds (*10*)	L2 adults (*11*)
AE	2.23 (0.59)	2.43 (0.60)	2.43 (0.48)
SW	3.03 (0.99)	3.65 (0.81)	3.49 (0.67)

Looking first at results for adults, it was found that the non-native listeners had poorer speech recognition for both kinds of signals than did the native listeners. All participants were required to pass a hearing screening before testing, and so all had equal access to the sensory information available in these signals. Furthermore, *j* factors were similar for both groups of adults, so it may be concluded that top-down cognitive influences on speech recognition were similar in this experiment: Although others have shown effects for L2 adults, none were found for these sentences with simple syntax. Consequently, we are left to posit the locus of effect squarely on how the two groups perceptually organized the sensory information that they received. The suggestion made here is that every language has its own unique global structure. The long-term spectra computed by Boysson-Bardies *et al.* (1986) for samples from speakers of four different languages demonstrated this fact for spectral structure. Presumably languages also differ in terms of global temporal and amplitude structure, given that languages vary in rhythm and stress. Here, it is suggested that adults are familiar with the global structure of their native language, and organize speech signals accordingly. This idea is commensurate with what is known about the phonetic perception of listeners for non-native languages. Listeners organize sensory information in the speech signal according to the phonetic structure of their native language (Beddor and Strange, 1982; Abramson and Lisker, 1967; MacKain *et al.*, 1981), so much so, in fact, that the effect is often described as L1 interference (e.g., Flege and Port, 1981). However, ideas regarding how L1 interferes with the perception of L2 traditionally focus on sensory information at the level of the acoustic cue. The signals presented to listeners in this experiment lacked that level of signal detail; there were no properties constituting acoustic cues. Nonetheless, these L2 speakers had more difficulty than L1 listeners recognizing linguistically relevant structure. That can only mean that they were unable to perceptually organize the sensory information in the signals in such a way as to recover global structure in the L2.

The children in this study were obviously not affected by having a native language interfere with how they perceptually organized the signals they heard because they were listening to processed versions of their native language. Nonetheless, these children had poorer speech recognition for the processed signals than L1 adults. This age-related finding likely reflects the fact that children have not had ample time to become as familiar with the global structure of their native language as adult listeners. Of course, it could just be that children show deficits at recognizing any impoverished signal, but results of another study suggest otherwise. Nittrouer and Lowenstein (2009) presented fricative-vowel syllables to adults and children for fricative labeling. The fricative noises were synthetic and spanned a continuum from one appropriate for /s/ to one appropriate for /ʃ/. The vocalic portions were either from natural speech samples or were sine wave replicas of that natural speech. Children's labeling functions were indistinguishable across the two kinds of stimuli. Therefore, it may be concluded that children are capable of using the information in processed speech signals for recognition, at least at the level of phonetic labeling.

Alternatively, the age-related differences observed here could be attributable to children having poorer access to the sensory information available in the speech signal, perhaps because of immature auditory systems (e.g., Sussman, 1993). That seems unlikely, however, given that L2 adults also performed more poorly than L1 adults, and no one would attribute that difference to poorer access to sensory information.

Moreover, it was found that children had similar *j* factors to adults, and so age-related differences in word recognition cannot be attributed to differences in top-down effects. Again, it seems that the source of the variability in recognition scores for L1 listeners of different ages must be placed squarely on differences in abilities to perceptually organize the sensory information in order to recover linguistically relevant structure.

Finally, it is informative to examine relative performance for the two kinds of processed signals used in this experiment. Children showed disproportionately better performance for the SW stimuli than for the AE stimuli. For example, 5-year-olds performed similarly to L2 adults for AE stimuli but significantly better for SW stimuli. 3-year-olds, on the other hand, performed more poorly than L2 adults for AE stimuli, but similarly with SW stimuli. Thus, all listeners were better able to recognize speech with the SW stimuli than for the AE stimuli, but this advantage was disproportionately greater for the younger children. We speculate that the reason for this outcome is related to the fact that in recognizing speech, young children rely particularly strongly on the dynamic spectral components, which are preserved by sine wave replicas.

In conclusion, this experiment was designed as an extension of Nittrouer *et al.* (2009), which showed that native language experience affected listeners' abilities to perceptu-

TABLE VIII. Mean percent correct words recognized by listeners for whom *j* factors could be computed. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	79.53 (8.30)	47.35 (13.37)	43.20 (10.08)	N/A	40.97 (9.78)
SW	N/A	91.14 (2.59)	86.32 (6.11)	75.40 (15.22)	77.20 (8.78)

ally organize processed speech signals, and that 7-year-olds showed greater benefits for SW over AE stimuli than adults. First, this experiment replicated the effect of native language experience: L2 adults performed more poorly than L1 listeners, in spite of having sufficient skills in their L2 to take graduate level course work. This difference in performance could not be attributed either to differences in sensory functioning or to differences in top-down cognitive effects. Second, this experiment extended the age-related findings of Nittrouer *et al.* (2009): Younger children than those who participated in that earlier study showed disproportionately enhanced performance for SW stimuli. 3-year-olds were able to recognize 75% of the words presented in SW sentences correctly. In sum, both native language experience and the extent of that experience have been found to influence listeners' abilities to organize sensory information in order to recover linguistically relevant structure.

These findings could have important implications for how we intervene with individuals, particularly children, who have hearing loss. At present, cochlear implants are not very good at preserving the kind of dynamic spectral information represented so well by the sine wave replicas used here. However, some individuals with severe-to-profound hearing loss have aidable hearing in the low frequencies. It may be worth exploiting that residual hearing in order to provide at least the first formant to these individuals through a hearing aid, and this practice could be especially beneficial for young children who rely heavily on dynamic spectral structure. Uchanski *et al.* (2009) found evidence to support this recommendation from one child with electric and acoustic hearing on the same side, and Nittrouer and Chapman (2009) found evidence from language outcomes of 58 children with hearing loss: Those who had some experience with combined electric and acoustic hearing fared better than children who had no such experience.

ACKNOWLEDGMENTS

We thank Mallory Monjot for her help with the reliability measures. This work was supported under research Grant No. R01 DC-00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to S.N.

APPENDIX

Sentences from the HINT-C that were used are as follows. The words in parentheses indicate that either words are acceptable answers. The underlined word is the one that is said in creating the stimuli.

(a) Practice sentences

1. The yellow pears taste good.
2. (A/the) front yard (is/was) pretty.
3. The pond water (is/was) dirty.
4. The ground (is/was) very hard.
5. They painted (a/the) wall white.
6. (A/the) little girl (is/was) shouting.
7. (An/the) ice cream (is/was) melting.
8. (An/the) apple pie (is/was) baking.
9. (A/the) boy forgot his book.
10. The two farmers (are/were) talking.
11. The sky (is/was) very blue.
12. (A/the) black dog (is/was) hungry.

(b) Test sentences

1. Flowers grow in (a/the) garden.
2. She looked in her mirror.
3. They heard (a/the) funny noise.
4. (A/the) book tells (a/the) story.
5. (A/the) team (is/was) playing well.
6. (A/the) lady packed her bag.
7. They waited for an hour.
8. (A/the) silly boy (is/was) hiding.
9. (A/the) mailman shut (a/the) gate.
10. (A/the) dinner plate (is/was) hot.
11. They knocked on (a/the) window.
12. He (is/was) sucking his thumb.
13. He grew lots of vegetables.
14. He hung up his raincoat.
15. The police helped (a/the) driver.
16. He really scared his sister.
17. He found his brother hiding.
18. She lost her credit card.
19. He wore his yellow shirt.
20. The young people (are/were) dancing.
21. Her husband brought some flowers.
22. The children washed the plates.
23. (A/the) baby broke his cup.
24. They (are/were) coming for dinner.
25. They had a wonderful day.
26. The bananas (are/were) too ripe.
27. She argues with her sister.
28. (A/the) kitchen window (is/was) clean.
29. (A/the) mother heard (a/the) baby.
30. (An/the) apple pie (is/was) good.
31. New neighbors (are/were) moving in.
32. (A/the) woman cleaned her house.
33. The old gloves (are/were) dirty.
34. (A/the) painter uses (a/the) brush.
35. The bath water (is/was) warm.
36. Milk comes in (a/the) carton.
37. (A/the) ball bounced very high.
38. School got out early today.
39. The rain came pouring down.
40. (A/the) train (is/was) moving fast.
41. (A/the) baby slept all night.
42. Someone (is/was) crossing (a/the) road.
43. (A/the) big fish got away.
44. (A/the) man called the police.
45. (A/the) mailman brought (a/the) letter.
46. He climbed up (a/the) ladder.
47. He (is/was) washing his car.
48. The sun melted the snow.
49. The scissors (are/were) very sharp.
50. Swimmers can hold their breath.
51. (A/the) boy (is/was) running away.
52. (A/the) driver started (a/the) car.
53. The children helped their teacher.
54. (A/the) chicken laid some eggs.
55. (A/the) ball broke (a/the) window.
56. Snow falls in the winter.
57. (A/the) baby wants his bottle.
58. (An/the) orange (is/was) very sweet.

59. (An/the) oven door (is/was) open.
60. (A/the) family bought (a/the) house.

Abramson, A. S., and Lisker, L. (**1967**). "Discriminability along the voicing continuum: Cross-language tests," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, pp. 569–573.

Beddor, P. S., and Strange, W. (**1982**). "Cross-language study of perception of the oral-nasal distinction," J. Acoust. Soc. Am. **71**, 1551–1561.

Benkí, J. R. (**2003**). "Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition," J. Acoust. Soc. Am. **113**, 1689–1705.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (**1989**). "Discovering phonetic coherence in acoustic patterns," Percept. Psychophys. **45**, 237–250.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (**2004**). "Neural correlates of sensory and decision processes in auditory object identification," Nat. Neurosci. **7**, 295–301.

Boothroyd, A. (**1968**). "Statistical theory of the speech discrimination score," J. Acoust. Soc. Am. **43**, 362–367.

Boothroyd, A. (**1985**). "Evaluation of speech production of the hearing impaired: Some benefits of forced-choice testing," J. Speech Hear. Res. **28**, 185–196.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Boysson-Bardies, B., de Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigations of cross-linguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterstrom, (Stockton, New York), pp. 113–126.

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Broadbent, D. E. (**1967**). "Word-frequency effect and response bias," Psychol. Rev. **74**, 1–15.

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design," Percept. Psychophys. **55**, 513–525.

Dau, T., Ewert, S., and Oxenham, A. J. (**2009**). "Auditory stream formation affects comodulation masking release retroactively," J. Acoust. Soc. Am. **125**, 2182–2188.

Dorman, M. F., Lindholm, J. M., and Hannley, M. T. (**1985**). "Influence of the first formant on the recognition of voiced stop consonants by hearing-impaired listeners," J. Speech Hear. Res. **28**, 377–380.

Duffy, J. R., and Giolas, T. G. (**1974**). "Sentence intelligibility as a function of key word selection," J. Speech Hear. Res. **17**, 631–637.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

Elliott, L. L. (**1979**). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Erber, N. P. (**1971**). "Evaluation of special hearing aids for deaf children," J. Speech Hear Disord. **36**, 527–537.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Flege, J. E., Schmidt, A. M., and Wharton, G. (**1996**). "Age of learning affects rate-dependent processing of stops in a second language," Phonetica **53**, 143–161.

Fu, Q., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Giolas, T. G., Cooker, H. S., and Duffy, J. R. (**1970**). "The predictability of words in sentences," J. Aud Res. **10**, 328–334.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).

Gottfried, T. L., and Beddor, P. S. (**1988**). "Perception of temporal and spectral information in French vowels," Lang Speech **31**, 57–75.

Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.

Howes, D. (**1957**). "On the relationship between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–307.

Johnson, J. S., and Newport, E. L. (**1989**). "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language," Appl. Cognit. Psychol. **21**, 60–99.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 148–154.

Liberman, A. M., Cooper, F. S., Harris, K. S., and MacNeilage, P. F. (**1962**). "A motor theory of speech perception," in Proceedings of the Speech Communication Seminar, Stockholm, pp. 1–12.

Lotto, A. J., Hickok, G. S., and Holt, L. L. (**2009**). "Reflections on mirror neurons and speech perception," Trends Cogn. Sci. **13**, 110–114.

MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (**2003**). "The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception," J. Speech Lang. Hear. Res. **46**, 1184–1196.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.

Nilsson, M., Soli, S. D., and Gelnett, D. J. (**1996**). *Development and Norming of a Hearing in Noise Test for Children* (House Ear Institute, Los Angeles, CA).

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.

Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.

Nittrouer, S., and Chapman, C. (**2009**). "The effects of bilateral electric and bimodal electric-acoustic stimulation on language development," Trends Amplif. **13**, 190–205.

Nittrouer, S., and Lowenstein, J. H. (**2009**). "Does harmonicity explain children's cue weighting of fricative-vowel syllables?," J. Acoust. Soc. Am. **125**, 1679–1692.

Nittrouer, S., Lowenstein, J. H., and Packer, R. (**2009**). "Children discover the spectral skeletons in their native language before the amplitude envelopes," J. Exp. Psychol. Hum. Percept. Perform. **35**, 1245–1253.

Nittrouer, S., and Miller, M. E. (**1997a**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.

Nittrouer, S., and Miller, M. E. (**1997b**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **102**, 572–580.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–949.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Revoile, S. G., Holden-Pitt, L., and Pickett, J. M. (**1985**). "Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or with normal hearing," J. Acoust. Soc. Am. **77**, 1263–1265.

Rosenzweig, M. R., and Postman, L. (**1957**). "Intelligibility as a function of frequency of usage," J. Exp. Psychol. **54**, 412–422.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Stelmachowicz, P. G., Lewis, D. E., Kelly, W. J., and Jesteadt, W. (**1990**). "Speech perception in low-pass filtered noise for normal and hearing-impaired listeners," J. Speech Hear. Res. **33**, 290–297.

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.
Sussman, J. E. (**1993**). "Auditory processing in children's speech perception: Results of selective adaptation and discrimination tasks," J. Speech Hear. Res. **36**, 380–395.
Uchanski, R. M., Davidson, L. S., Quadrizius, S., Reeder, R., Cadieux, J., Kettel, J., and Chole, R. A. (**2009**). "Two ears and two (or more?) devices: A pediatric case study of bilateral profound hearing loss," Trends Amplif. **13**, 107–123.
Van Engen, K. J., and Bradlow, A. R. (**2007**). "Sentence recognition in native- and foreign-language multi-talker background noise," J. Acoust. Soc. Am. **121**, 519–526.
Watson, J. M. M. (**1997**). "Sibilant-vowel coarticulation in the perception of speech by children with phonological disorder," Ph.D. thesis, Queen Margaret College, Edinburgh.
Zampini, M. L., Clarke, C., and Green, K. P. (**2000**). "Language experience and the perception of stop consonant voicing in Spanish: The case of late English-Spanish bilinguals," in *Spanish Applied Linguistics at the Turn of the Millennium*, edited by R. P. Leow and C. Sanz, (Cascadilla, Somerville, MA), pp. 194–209.