

# Does harmonicity explain children's cue weighting of fricative-vowel syllables?<sup>a)</sup>

Susan Nittrouer<sup>b)</sup> and Joanna H. Lowenstein

*Department of Otolaryngology-Head & Neck Surgery, Ohio State University, Cramblett Suite 4A, 456 West 10th Avenue, Columbus, Ohio 43210*

(Received 18 June 2008; revised 3 December 2008; accepted 4 December 2008)

When labeling syllable-initial fricatives, children have been found to weight formant transitions more and fricative-noise spectra less than adults, prompting the suggestion that children attend more to the slow vocal-tract movements that create syllabic structure than to the rapid gestures more closely aligned with individual phonetic segments. That explanation fits well with linguistic theories, but an alternative explanation emerges from auditory science: Perhaps children attend to formant transitions because they are found in voiced signal portions, and so formants share a common harmonic structure. This work tested that hypothesis by using two kinds of stimuli lacking harmonicity: sine-wave and whispered speech. Adults and children under 7 years of age were asked to label fricative-vowel syllables in each of those conditions, as well as natural speech. Results showed that children did not change their weighting strategies from those used with natural speech when listening to sine-wave stimuli, but weighted formant transitions less when listening to whispered stimuli. These findings showed that it is not the harmonicity principle that explains children's preference for formant transitions in phonetic decisions. It is further suggested that children are unable to recover formant structure when those formants are not spectrally prominent and/or are noisy. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3056561]

PACS number(s): 43.71.An, 43.71.Ft [MSS]

Pages: 1679–1692

## I. INTRODUCTION

The primary focus of the study of human speech perception has traditionally been on the question of which signal properties listeners use in making phonetic judgments. An early example of this focus was a study published by [Harris \(1958\)](#) demonstrating that adult, native speakers of English base their decisions about the sibilant fricatives /s/ and /ʃ/ primarily on the spectral structure of the noise itself, but base their decisions about the weak fricatives /f/ and /θ/ on the pattern of formant transitions in the voiced syllable portion. That experiment was done using natural, spliced speech samples, but much of the work on human speech perception has been conducted with synthetic speech. Typically a single acoustic property was manipulated along a continuum spanning settings that evoke one phonetic judgment or another, assessing how those settings affected the choice of which phoneme was heard. The properties manipulated in such experiments came to be known as “acoustic cues” ([Repp, 1982](#)).

Although generally informative, one flaw in this approach was that acoustic evidence clearly showed, even then, that the qualities of many discrete regions of the signal varied as a function of surrounding phonetic structure. For example, the noise spectrum of syllable-initial /s/ or /ʃ/ differs depending on the vowel that follows ([Kunisaki and Fujisaki, 1977](#); [Heinz and Stevens, 1961](#)). It was reasonable that the effects of this sort of variation on phonetic decisions should

be examined, and so experimental paradigms were adjusted to incorporate a second acoustic cue, manipulated in a dichotomous manner, that is, set to be appropriate for the phoneme represented at one end or the other of the acoustic continuum formed by manipulating the first cue. For example, [Mann and Repp \(1980\)](#) found that when voiced formant transitions were set to be appropriate for either a preceding alveolar or palatal constriction, the /ʃ/-/s/ phoneme boundary on an acoustic continuum was shifted. This sort of interaction was termed cue trading or trading relations. These sorts of results showed that the relation between acoustic cues and phonetic judgments is more complicated than at first envisioned.

Later [Nittrouer and Studdert-Kennedy \(1987\)](#) set out to explore whether or not children (3–7 years old) exhibit the sort of cue trading reported by [Mann and Repp \(1980\)](#). For that purpose fricative-vowel syllables similar to those of [Mann and Repp \(1980\)](#) were constructed with synthetic noises spanning a spectral /ʃ/-to-/s/ continuum, and vocalic portions taken from natural speech samples with formant transitions appropriate for either an initial /ʃ/ or /s/. Within this design the spectral structure of the fricative noise was presumed to be the “primary” cue, defined as the one that would be given the greatest perceptual weight in the fricative decision. That designation followed from the work of [Harris \(1958\)](#) showing that indeed adult listeners primarily base their sibilant decisions on characteristics of the noise itself. It was presumed that the role of formant transitions in this decision would be secondary, and likely learned through years of experience hearing the patterns of formant change arising from the vocal tract moving from a fricative target to a vowel

<sup>a)</sup> Portions of this work were presented at the 154th Meeting of the Acoustical Society of America, New Orleans, LA, November 2007.

<sup>b)</sup> Electronic mail: nittrouer.1@osu.edu

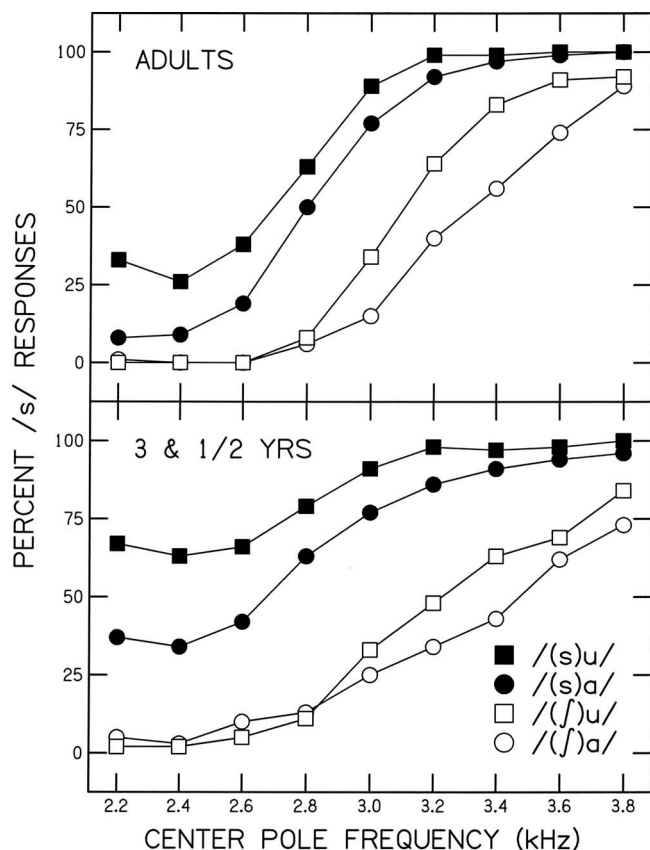


FIG. 1. Labeling functions for adults (top) and 3½-year-olds (bottom) from Nittrouer (1992). Center frequencies of the synthetic fricative noises are shown on the abscissa, and percents responses are shown on the ordinate. A separate function is plotted for each syllable type.

target (Stevens, 1975). The specific prediction was that the youngest children would show a greatly attenuated trading of cues, with a developmental increase in the magnitude of the trade. In fact, precisely the opposite pattern of results was obtained: The youngest listeners showed the greatest influence of formant transitions on fricative labeling, with a developmental decrease in effect. These findings were supported by subsequent research, which demonstrated that a developmental change occurs in the relative amount of attention paid to these acoustic cues for children between roughly 3 and 7 years of age (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b). Figure 1, from Nittrouer (1992), illustrates this trend: Children’s labeling functions are more separated as a function of whether formant transitions are appropriate for a preceding /s/ or /f/ closure. This pattern indicates that children’s responses were more strongly based on vocalic formant transitions than were those of adults. At the same time, children’s functions are shallower than those of adults, indicating that they paid less attention to the spectra of fricative noises in making their decisions. This developmental change in the amount of attention, or weight, given to separate acoustic cues came to be known as the “developmental weighting shift” (e.g., Nittrouer, 1996; Nittrouer *et al.* 1993; Nittrouer and Miller, 1997a, 1997b).

The finding that children weight acoustic cues differently from adults in their linguistic decision-making comple-

mented emerging evidence that speakers, and so listeners, of different languages weight acoustic cues differently (e.g., Beddor and Strange, 1982; Crowther and Mann, 1992, 1994; MacKain *et al.*, 1981). The way that the acoustic speech signal is shaped by gestures of the vocal tract in order to provide for the recognition of phonetic structure varies across languages (e.g., Crowther and Mann, 1992; Flege and Port, 1981; Lisker and Abramson, 1964). A primary chore of the child learning a language is to discover which properties of the signal provide linguistically significant variation, and so must be attended to, and which properties contain variation that is irrelevant to deriving meaning, and so may safely be ignored. It is explicitly this language-related discovery process that the developmental weighting shift seeks to capture. But while this notion may explain why weighting strategies change through childhood, an unresolved question concerns why children’s initial strategies are the way they are. Why would initial weighting strategies be so dependent on formant transitions within the voiced portion of the signal? The answer to this question has theoretical and clinical implications.

Several related explanations have been offered for why children pay particular attention to formant transitions in their phonetic decisions. In general it has been speculated that these relatively slow modulations in spectral structure (i.e., formant transitions) attract children’s attention early in development precisely because they are dynamic. Research into the development of visual perception shows that children attend more to moving properties than to static properties in object recognition (Ruff, 1982; Spelke *et al.*, 1989). It is reasonable that they would do the same in auditory perception. Another possible explanation for children’s early weighting strategies has a purely linguistic basis: It is commonly asserted that young children use more global processing strategies than adults to recognize words and syllables (e.g., Cole and Perfetti, 1980; Charles-Luce and Luce, 1990, 1995; Jusczyk, 1992; Nittrouer, 1996; Walley *et al.*, 1986). Although exactly what the term “global processing strategies” is referring to has not always been specified in these accounts, it is reasonable to suggest that the term references children’s enhanced attention to signal properties that are syllabic in nature, rather than specific to individual phonemes. The idea that children pay special attention to formant transitions is compatible with the notion. Finally, it has been suggested that these formant transitions reflect the general movements of the vocal tract. In production, children appear to master general articulatory placements and patterns of movement between these targets before they learn how to make precise constriction shapes (e.g., Goodell and Studdert-Kennedy, 1993; MacNeilage and Davis, 1991; McGowan and Nittrouer, 1988). In sum, several hypotheses have been offered to explain children’s preferential weighting of formant transitions.

Yet another hypothesis for why children focus so strongly on formant transitions emerges from the field of general auditory perception. To appreciate this hypothesis, it is necessary to first ask ourselves why variation across formant frequencies would provide the listener with any information at all. There is no reason to expect three disparate

spectral components, as these formant frequencies are, to cohere, providing a single perceptual impression. So, why do listeners hear these separate but co-occurring patterns as one entity, providing information about the structure of the signal? Again, different explanations have been offered to answer that question, with one set of explanations involving strictly auditory principles. This set of explanations is best exemplified by the work of [Bregman \(1990\)](#) on Auditory Scene Analysis (ASA). These principles suggest that separate spectral components cohere if they have synchronous onsets and offsets, or if they have the same harmonic structure, a principle known as harmonicity. Because formant transitions in fricative-vowel syllables come from the voiced portion of the syllable they share a harmonic structure, and so the principle of harmonicity may explain the coherence of formants in speech perception.

Another kind of explanation for why formant transitions may cohere perceptually is that they all arise from a single vocal tract, a fact that is apparent to the listener. This position is best explained and supported by [Remez \*et al.\* \(1994\)](#) who methodically tested each principle of ASA and serially rejected each as an explanation for why the various components of the speech signal cohere. Of course, those experiments were done with adult listeners. In this current study, we entertained the possibility that children, being less experienced with speech, may require that signals strictly adhere to ASA principles if they are to integrate separate spectral portions and use them for speech perception. Perhaps children's greater weighting of formant transitions in fricative-vowel syllables, compared to that of adults, arises from nothing more than the fact that these signal portions adhere to ASA principles. In the work reported here, the specific possibility examined was that children may attend strongly to formant transitions in voiced signal portions because they have a preference for signals that adhere to the harmonicity principle.

To test this possibility, stimuli were needed that retained formant transitions but eliminated harmonic structure. This requirement necessitated that speech signals be processed in some way, which is always a risky business. While there are many ways to modify a speech signal to eliminate harmonicity, each creates concomitant acoustic changes that may affect perception in unpredictable ways. For that reason, two processing algorithms were used in this study: Both disrupted the principle of harmonicity, but each did so in different ways. Sine-wave speech was used because it lacks harmonic structure, yet preserves the dynamic information provided by formant transitions. Whispered speech was used because it also lacks harmonic structure and preserves formant transitions. It is, in addition, familiar to young children, and thus not as "artificial" as sine-wave speech might be viewed as being.

Sine-wave speech uses three sinusoidal tones to replicate the first three formants of speech. Since the stimuli consist of sine waves, there are no harmonic relations between the components. There is ample evidence that adult listeners can understand sine-wave replicas of sentences when they have been instructed that what they are listening to is speech ([Barker and Cooke, 1999](#); [Remez \*et al.\* 1981](#); [Remez \*et al.\*,](#)

[1994](#)), but results for children are mixed. A recent study by [Vouloumanos and Werker \(2007\)](#) reported that 1- to 4-day-old infants preferred natural speech stimuli to sine-wave analogs of the same stimuli. Older children, however, have been shown to comprehend sine-wave sentences as well as adults ([Nittrouer \*et al.\*, 2009](#)). In that latter experiment adults and 7-year-olds were asked to repeat four-word sentences presented as sine-wave analogs or as 4- or 8-channel noise vocoded analogs. The 7-year-olds performed similarly to adults for the sine-wave stimuli, but much more poorly for the noise vocoded stimuli. [Serniclaes \*et al.\* \(2001\)](#) presented sine-wave analogs of stimuli from a /ba/-/da/ continuum to dyslexic and normal-reading 13-year-olds as a discrimination task. The stimuli were first presented as nonspeech whistles, and then presented as speech. When the sine-wave syllables were presented as speech, both the dyslexics and normal readers showed a phoneme boundary effect that was absent when the sine-wave syllables were presented as nonspeech whistles. Thus the evidence indicates that by 7 years of age, at the latest, children treat sine-wave analogs as speech.

Whispering is something that most children are familiar with. In whispered speech, turbulent noise produced by air flow through partially adducted vocal folds is shaped by the supralaryngeal vocal tract, resulting in speech that is very understandable, but that lacks vocal fold vibrations (and thus a fundamental). Magnetic resonance imaging and endoscopic studies of whispered vowels provide evidence that during whispering the supraglottal structures shift downwards to inhibit vocal fold vibration ([Matsuda and Kasuya, 1999](#); [Tsunoda \*et al.\*, 1997](#)). Acoustic studies of whispered speech have shown that formant frequencies, particularly F1, tend to be higher in whispered than in voiced vowels ([Kallail and Emanuel, 1984a, 1984b](#)). Formant bandwidths appear broader in whispered than in voiced speech, as well. In terms of perception, [Tartter \(1989\)](#) found that untrained adults could accurately label consonant-vowel syllables (which consisted of the vowel [a] preceded by voiced and voiceless labial, alveolar, and velar stops, nasals, liquids, and fricatives) 64% of the time, with the primary confusions being for normally voiced consonants. Whispered vowels in [hVd] context were identified correctly 82% of the time ([Tartter, 1991](#)). In sum we would expect recognition of unvoiced consonant-vowel syllables to be highly accurate for whispered speech.

Although sine-wave analogs and whispered speech both provide ways of examining speech perception when the harmonicity principle is violated, they differ from each other in ways that should provide useful insights beyond the central question of this study. For example, sine waves have effective bandwidths of 1 Hz, and so the spacing between formants is much greater than in natural speech: Sine-wave formants are quite prominent. Whispered speech, on the other hand, has broadened bandwidths, and so formants are not as prominent as in natural speech. Investigators have worried in the past that the broader auditory filters associated with hearing loss might make it hard for listeners with hearing loss to resolve individual formants, and so to understand speech ([Summerfield \*et al.\*, 1981](#); [Turner and Holte, 1987](#); [Turner and Van Tasell, 1984](#)). Although that work with adults

showed that it is generally not a problem, the possibility exists that children might have a stronger need for more spectrally prominent formants. Evidence for this situation would be obtained if children showed no change from natural speech in their weighting of formant transitions when presented with sine-wave stimuli, but had difficulty with whispered speech.

Another difference between the two kinds of signals used in this study was that when whispered vocalic portions were appended to synthetic fricative noises, the result was signals consisting completely of noise. Previous studies examining perceptual weighting strategies for fricative-vowel stimuli not only found that children give more weight to formant transitions than adults in their fricative decisions but that there is a reciprocal decrease in the weight given to fricative-noise spectra (Nittoer and Miller, 1997a, 1997b; Nittoer, 1992; Nittoer and Studdert-Kennedy, 1987). It may simply be that children are not as skilled at using information in noisy signal portions as adults are.

We stand to improve our understanding of the nature of perceptual development in general by exploring the questions asked by this work. In addition, understanding the basis of children's perceptual preference for dynamic signal components in speech perception may affect what we do in fitting auditory prostheses for children with hearing loss. Pediatric audiologists worry that fitting children with hearing aids using the same algorithms as those used with adults is not appropriate (e.g., Pittman *et al.*, 2003), and, in particular, they worry about preserving children's abilities to recognize sibilants because of their importance in grammatical marking (Stelmachowicz *et al.*, 2007). Therefore, it is appropriate to ask why children show the perceptual weighting strategies that they do, and what conditions support or harm their abilities to use these strategies. Results showing either that children require formants to be spectrally prominent or that they avoid noisy signal components would provide useful information for the future design and fitting of auditory prostheses.

In summary, the two experiments in this study were conducted to test the hypothesis that children greatly weight formant transitions because they share harmonic structure. Two types of stimuli that lack harmonic structure were used: sine-wave and whispered speech. If children were found to pay less attention to formant transitions with both kinds of nonharmonic stimuli, support would be provided for the hypothesis that children attend especially to signal portions that adhere to ASA principles. On the other hand, if children were found to perform similarly with natural speech and both kinds of nonharmonic stimuli, that finding would provide evidence that children attend to these signal portions for other reasons, likely the language-related reasons described above. If children demonstrated modifications in their perceptual weighting strategies for one, but not the other, processed signal, we would have to conclude that it was some concomitant change in signal structure that accounted for the perceptual change, rather than the lack of harmonicity.

## II. EXPERIMENT 1: VOICED AND SINE-WAVE VOCALIC PORTIONS

The primary goal of this experiment was to test the hypothesis that children's strong weighting of formant transitions in phonetic judgments of syllable-initial fricatives is accounted for by the fact that voiced syllable portions adhere to the harmonicity principle, whereas noise syllable portions do not. Support for this hypothesis would be obtained if children decreased the weight assigned to formant transitions when the harmonicity principle was disrupted by creating sine-wave analogs of the voiced syllable portions.

### A. METHOD

#### 1. Subjects

Twenty adults between the ages of 18 and 40 years, 20 7-year-olds, 22 5-year-olds, and 23 3-year-olds participated. The mean age of 7-year-olds was 7 years, 1 months (7;1), with a range from 6;11 to 7;5. The mean age of 5-year-olds was 5;1, with a range from 4;11 to 5;5. The mean age of 3-year-olds was 3;9, with a range of 3;5 to 3;11. None of the listeners (or their parents, in the case of children) reported any history of hearing or speech disorder, and they all seemed typical to the experimenter doing the testing. Nonetheless, objective criteria were used to document that they were indeed typical. In particular, all listeners passed hearing screenings consisting of the pure tones of 0.5, 1, 2, 4, and 6 kHz presented at 25 dB HL to each ear separately. Children were given the Goldman Fristoe 2 Test of Articulation (Goldman and Fristoe, 2000) to screen for speech problems and were required to score at or better than the 30th percentile for their age. An additional constraint on these scores was that no child could exhibit errors in producing /ʃ/ or /s/. All children were free from significant histories of otitis media, defined as six or more episodes during the first three years of life. Adults were given the reading subtest of the Wide Range Achievement Test-Revised (Jastak and Wilkinson, 1984) as a way to screen for language problems and were required to demonstrate at least an 11th-grade reading level.

#### 2. Equipment and materials

All testing took place in a soundproof booth, with the computer that controlled the experiment in an adjacent room. Hearing was screened with a Welch Allyn TM262 audiometer using TDH-39 headphones. Stimuli and stories were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. Two drawings (on 8 × 8 in.<sup>2</sup> cards) were used to represent each response label in each experiment, such as "Sue" (a girl) and a "shoe," or "Sa" (a space alien) and "Sha" (a king). Gameboards with ten steps were also used with children; they moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

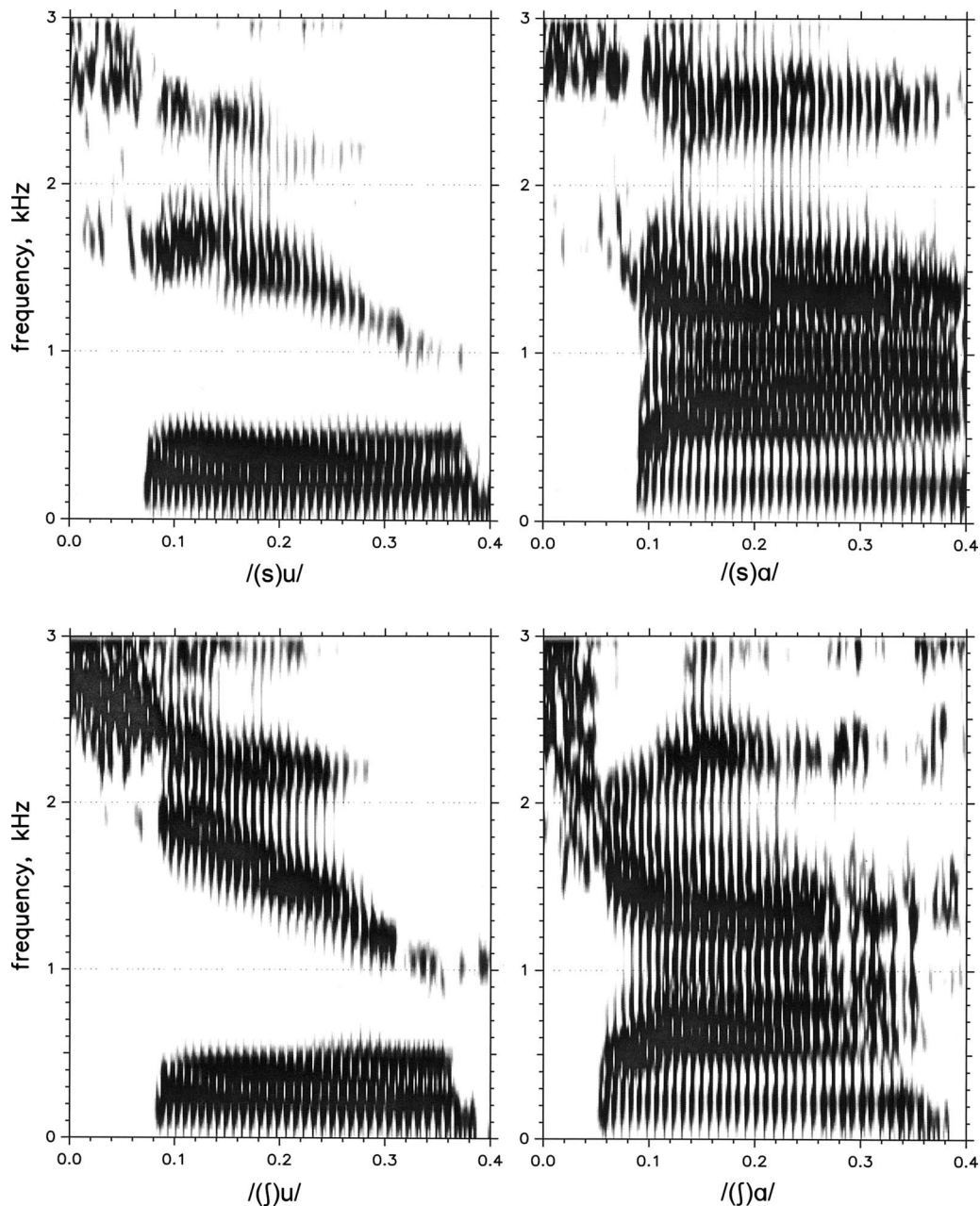


FIG. 2. Spectrograms displaying examples of the natural vocalic portions used in the VOICED stimuli of experiments 1 and 2. The fricative label in parentheses indicates the fricative context from which the vocalic portion was separated.

### 3. Stimuli

Thirty-six separate stimuli were created for each of the VOICED and SINE-WAVE conditions: nine fricative noises ranging along a continuum from one that is readily identified as /ʃ/ to one that is readily identified as /s/, combined with each of four vocalic portions. These vocalic portions differed in vowel quality (/a/ or /u/) and in whether formant transitions at the start of the portion were appropriate for a preceding /ʃ/ or /s/. Design of stimuli was based on stimuli that have been used previously in several experiments (Nittrouer 1992, 1996; Nittrouer and Miller, 1997a), and so we had general notions of what to expect in responses from children and adults. The VOICED stimuli in this experiment were exactly those used in the earlier experiments. In these stimuli, the fricative portions were single-pole, synthetic

noises, with the center frequency ranging from 2.2 to 3.8 kHz in nine 200 Hz steps. The vocalic portions of these VOICED stimuli were taken from a male speaker saying /su/, /sa/, /ʃu/, and /ʃa/. The recordings were sampled at 22.05 kHz and low-pass filtered below 11.025 kHz. Three tokens of each syllable were used, and Fig. 2 shows spectrograms of one of each kind of vocalic portion. Table I displays mean values for some phonetically relevant properties of the vocalic portions. For all /u/ portions, F1 remained stable throughout the vocalic portion, and F2 fell in frequency through the entire portion to a mean ending frequency of 930 Hz. For all /a/ vocalic portions, F1 rose in frequency over roughly the first 60 ms to a mean steady-state frequency of 703 Hz, and F2 fell in frequency over the first 60–90 ms to a mean steady-state frequency of 1309 Hz. For all stimuli,

TABLE I. Means across the three tokens for the phonetically relevant acoustic properties of the vocalic portions of the VOICED stimuli used in experiments 1 and 2.

Portion	$f_0$ (Hz)	Dur (ms)	F1-onset (Hz)	F2-onset (Hz)	F3-onset (Hz)
(s)a	93	333	457	1365	2457
(f)a	95	337	379	1532	2367
(s)u	99	347	320	1520	2496
(f)u	97	348	308	1706	2288

F3 fell in frequency over roughly the first 100 ms to a mean steady-state value of 2320 Hz for /u/ and 2365 Hz for /a/.

Natural vocalic portions were separated from the fricative noises using a waveform editor, and recombined with each of the synthetic fricative noises, yielding 36 kinds of VOICED stimuli (9 noises  $\times$  4 vocalic portions), with 3 tokens of each. The center panel of Fig. 3 shows a spectrogram of a representative VOICED / $\int$ a/ stimulus with the 3.0 kHz noise. RMS amplitude was equal for the noise and vocalic syllable portions.

The SINE-WAVE stimuli used the same synthetic fricative noises as the VOICED stimuli. The vocalic portions were created based on recordings of natural /su/, /sa/, /ju/, and /ja/ from the male speaker. Three sine waves were used to approximate the first three formants of each vowel and were adjusted until they most closely resembled the natural vowels in Fig. 2. The SINE-WAVE stimuli were all 250 ms in length and were generated at a sampling rate of 22.05 kHz and low-pass filtered at 11.025 kHz using TONE software (Tice and Carrell, 1997). The sine-wave vowels were combined with each of the synthetic noises, resulting in 36 SINE-WAVE stimuli (9 noises  $\times$  4 sine-wave vowels). The left

panel of Fig. 3 shows a spectrogram of a representative SINE-WAVE / $\int$ a/ stimulus with the 3.0 kHz noise. RMS amplitude was equal for the noise and sine-wave vocalic portions.

#### 4. Procedures

The screening tasks were presented first, followed by the labeling task. Stimuli with each vowel (/u/ or /a/) and stimulus type (VOICED or SINE-WAVE) were presented separately, making four conditions. Adults, 7-year-olds, and 5-year-olds participated in two test sessions. For these listeners order of presentation of conditions was randomized across listeners, with the restrictions that conditions with the same vowel or stimulus type were not presented in the same session and that order of presentation of the VOICED and SINE-WAVE conditions be different across sessions. For example, if the first condition, randomly chosen, consisted of VOICED /u/ stimuli, the second condition presented at that session would necessarily consist of SINE-WAVE /a/ stimuli. At the next session the listener would be presented with SINE-WAVE /u/, followed by VOICED /a/. The same ordering procedures were followed for 3-year-olds, except that they were spread out over four test sessions because pilot testing had revealed that 3-year-olds were more likely to complete testing if it was spread out like that.

During testing, stimuli were presented ten times each, in blocks of 18. Because there were three tokens of each VOICED stimulus, the program randomly selected one of the tokens to present during the first block, and then repeated this random selection during the next block without replacement. After the first three blocks, this process was repeated until testing was done. The listener's task during testing was

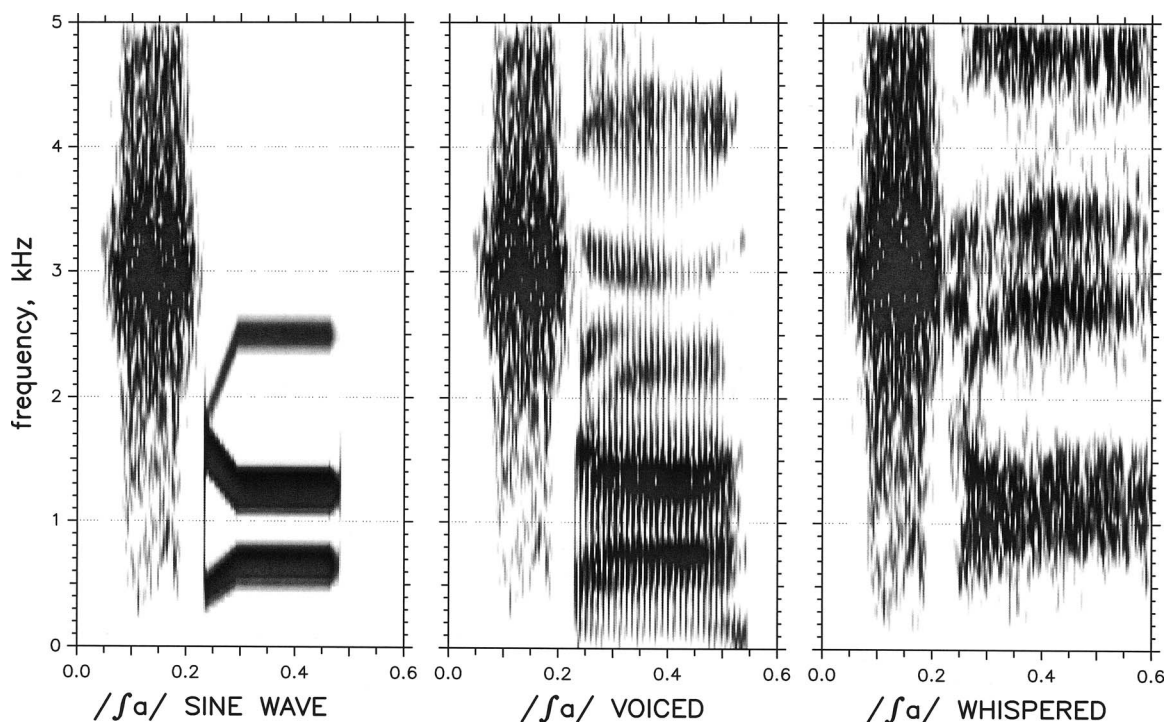


FIG. 3. Spectrograms displaying examples of the SINE-WAVE, VOICED, and WHISPERED / $\int$ a/ stimuli from experiments 1 and 2.

to point to the picture representing their response choice, and say the word that corresponded to that choice.

Children were introduced to the response labels (sa, sha, sue, or shoe) before testing with brief, digitized stories accompanied by pictures. Each story was presented once with natural speech for the VOICED condition. For the SINE-WAVE condition, each story was presented twice, once with natural speech and once with sine-wave speech. These stories served both to familiarize children with the response labels and to provide experience listening to sine-wave speech. During pilot testing it was found that presenting these stories to adults did not affect their responding. Consequently, we did not play them to adults.

All listeners heard 12 practice items (6 each of /sV/ and /ʃV/) before testing, using the best exemplars of each category (i.e., the syllable consisting of the lowest-frequency fricative noise combined with a vocalic portion with /ʃ/ formant transitions or the syllable consisting of the highest-frequency fricative noise combined with a vocalic portion with /s/ formant transitions). Each listener had to label 11 of the 12 best exemplars correctly to proceed to testing. This criterion ensured that all listeners could perform the required task, and knew the response labels. 3-year-olds had one additional familiarization task: Before hearing the practice items, they were presented with completely natural tokens of each syllable and had to label 11 of 12 correctly to proceed to the next practice. This extra step ensured that 3-year-olds recognized natural exemplars of these stimuli correctly. When it came to testing, listeners were required to give at least 80% accurate responses to the best exemplars, thereby ensuring that data were included only from listeners who maintained attention to the task.

To summarize, slightly different procedural steps were followed for each group: Adults received practice with the best exemplars, followed by testing. Children who were 5- or 7-years-old heard stories, followed by practice with the best exemplars, and then testing. The 3-year-olds heard the stories, followed by practice with natural tokens, then practice with the best exemplars, and finally testing.

For analysis purposes, partial correlation coefficients were computed between the proportion of “s” responses and the acoustic properties of formant transitions (appropriate for /s/ or /ʃ/) and fricative noise (step on the continuum), across vowels. These correlation coefficients served as weighting coefficients, explaining the proportion of variance in response labels associated with formant transitions and with fricative-noise spectra (Nittrouer, 2002).

## B. Results

### 1. VOICED condition

Five 3-year-olds were dismissed because they did not meet the criterion of getting 11 of the 12 syllables correct during practice with completely natural tokens. All of the remaining 18 3-year-olds were able to complete the VOICED /u/ condition with 80% accurate responses on best exemplars or better, but only 12 of them were able to complete the VOICED /a/ condition.

Figure 4 displays the labeling functions for adults,

7-year-olds, 5-year-olds, and 3-year-olds for the VOICED condition. Adults and 7-year-olds performed similarly. 3-year-olds and 5-year-olds showed greater separation in functions based on whether the vocalic portion originally started with /ʃ/ or /s/, and generally shallower functions than adults or 7-year-olds. These two age-related differences in response patterns reflect greater weighting of formant transitions and lesser weighting of fricative-noise spectra, respectively, by the younger listeners (Nittrouer, 2002). When Fig. 4 is compared to Fig. 1 it is seen that response patterns for adults and 3-year-olds in this experiment are very similar to what has been found in previous studies (Nittrouer, 1992, 1996; Nittrouer and Miller, 1997a, 1997b).

Table II shows mean weighting coefficients for the VOICED condition. The response patterns observed in Fig. 4 are apparent in these coefficients: 3- and 5-year-olds weighted transitions more than adults or 7-year-olds, and weighted the fricative noises less. When these weighting coefficients are compared to weighting coefficients derived from responses of listeners hearing the same stimuli in the past (e.g., Nittrouer and Miller, 1997b), very similar results are obtained: Weighting coefficients of formant transitions for adults, 7-year-olds, and 4½-year-olds were, respectively, 0.32, 0.36, and 0.46; for fricative noises weighting coefficient were 0.82, 0.77, and 0.68 (Nittrouer, 2002). Thus, we are assured that the listeners in this experiment are similar to those in earlier experiments.

One-way analyses of variance (ANOVAs) with age as the main effect and *post hoc* comparisons were done on weighting coefficients for formant transitions and fricative noises obtained in this experiment. Results are shown in Table III. Here, and in the remainder of this paper, precise statistical results are given only when  $p < 0.10$ . Otherwise, results are reported as not significant (ns). These results show significant overall age effects for both formant transitions and fricative noises. The *post hoc t*-tests confirm that 3-year-olds and 5-year-olds weighted formant transitions more and fricative noises less than 7-year-olds or adults. It is also observed that the four age groups form a dichotomy in response patterns, such that the two youngest groups did not differ in weighting strategies from each other, and the two oldest groups did not differ from each other.

### 2. SINE-WAVE condition

Three-year-olds found the SINE-WAVE condition very difficult: Only 8 out of 18 of them (44%) responded to the SINE-WAVE practice items and best exemplars during testing at levels accurate enough to have their data included in analyses. When less than half of the listeners in a group are

TABLE II. Mean weighting coefficients in the VOICED condition of experiment 1. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds	3-year-olds
Transition	0.30 (0.17)	0.32 (0.15)	0.48 (0.19)	0.50 (0.15)
Noise	0.82 (0.11)	0.81 (0.11)	0.71 (0.14)	0.66 (0.13)

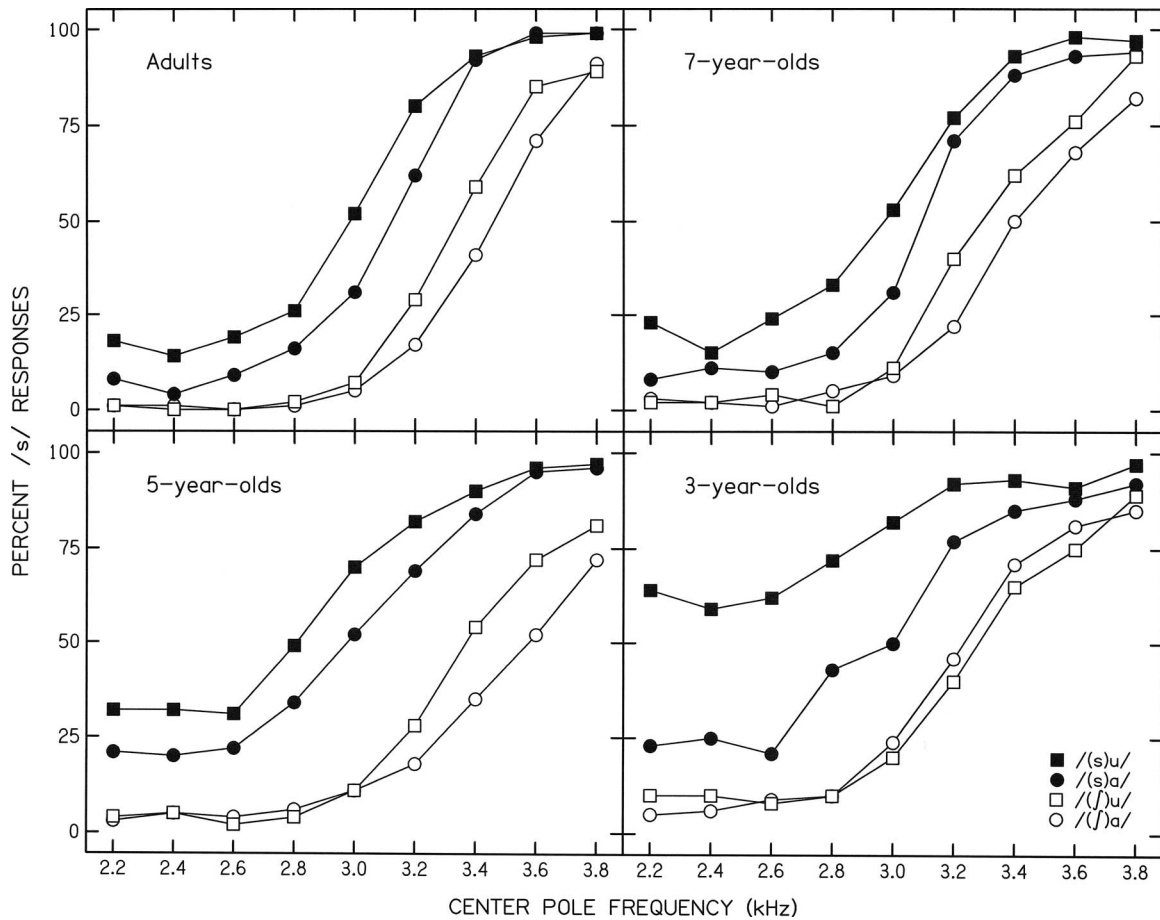


FIG. 4. Labeling functions for the VOICED condition of experiment 1 for adults, 7-year-olds, 5-year-olds, and 3-year-olds. See legend for Fig. 1 for more details.

able to perform a task, their performance on that task cannot be taken as representative of the group as a whole. Therefore, data from the 3-year-olds were not analyzed for this condition. On the other hand, 80% of 7-year-olds and 75% of

TABLE III. Results of ANOVAs and *post hoc t* tests done on weighting coefficients in the VOICED condition of experiment 1. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	<i>df</i>	<i>F</i> or <i>t</i>	<i>p</i>	Bonferroni significance
Transition				
Age effect	3,76	8.11	<0.001	
Adults vs 7-year-olds	76	0.38	ns	
Adults vs 5-year-olds	76	3.53	<0.001	0.01
Adults vs 3-year-olds	76	3.80	<0.001	0.01
7- vs 5-year-olds	76	3.14	0.002	0.05
7- vs 3-year-olds	76	3.46	0.001	0.01
5- vs 3-year-olds	76	0.46	ns	
Noise				
Age effect	3,76	7.61	<0.001	
Adults vs 7-year-olds	76	-0.21	ns	
Adults vs 5-year-olds	76	-2.79	0.007	0.05
Adults vs 3-year-olds	76	-3.98	<0.001	0.001
7- vs 5-year-olds	76	-2.58	0.02	0.10
7- vs 3-year-olds	76	-3.78	<0.001	0.01
5- vs 3-year-olds	76	-1.35	ns	

5-year-olds were able to complete the SINE-WAVE condition. Data from these listeners were analyzed.

Table IV shows mean weighting coefficients for the SINE-WAVE condition. In comparing the numbers shown on Table IV to those on Table II, the greatest difference appears for adults: They seem to have weighted the fricative noises similarly, but to have weighted formant transitions less in this condition than in the VOICED condition. To a lesser extent it appears as if 5-year-olds weighted fricative noises less in this SINE-WAVE condition than in the VOICED condition.

ANOVAS, with *post hoc t*-tests, were done on weighting coefficients for formant transitions and fricative noises separately, and results are shown in Table V. There was a significant age effect for both transitions and fricative noises. The *post hoc t*-tests confirm that 5-year-olds continued to weight formant transitions more and fricative noises less in their

TABLE IV. Mean weighting coefficients in the SINE-WAVE condition of experiment 1. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.21 (0.10)	0.31 (0.16)	0.50 (0.19)
Noise	0.85 (0.06)	0.80 (0.12)	0.65 (0.20)



TABLE V. Results of ANOVAs and *post hoc t*-tests done on weighting coefficients in the SINE-WAVE condition of experiment 1. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	<i>df</i>	<i>F</i> or <i>t</i>	<i>p</i>	Bonferroni significance
Transition				
Age effect	2,59	20	<0.001	
Adults vs 7-year-olds	59	2.16	ns	
Adults vs 5-year-olds	59	6.21	<0.001	0.001
7- vs 5-year-olds	59	4.00	<0.001	0.001
Noise				
Age effect	2,59	12.50	<0.001	
Adults vs 7-year-olds	59	-1.03	ns	
Adults vs 5-year-olds	59	-4.73	<0.001	0.001
7- vs 5-year-olds	59	-3.68	<0.001	0.01

fricative decisions than 7-year-olds or adults. These older two groups performed similarly to each other, as they had for the VOICED stimuli.

### 3. Across-conditions comparison

Of course, the major focus of this experiment was to examine whether listeners, particularly children, modify their weighting strategies when stimuli lack harmonicity. For that purpose, weighting coefficients obtained for the VOICED and SINE-WAVE stimuli were compared for each group separately, using matched *t*-tests. Results are shown in Table VI. These tests confirm that 7-year-olds assigned similar weighting coefficients to both properties in each condition. The tests further confirm that adults weighted formant transitions less in the SINE-WAVE than in the VOICED condition. Results for 5-year-olds revealed that they weighted fricative noises slightly less in the SINE-WAVE than in the VOICED condition, but the difference fell just short of statistical significance.

### C. Discussion

The purpose of this experiment was to test the hypothesis that listeners would change their perceptual weighting strategies when sine-wave replicas replaced natural vocalic portions, consequently eliminating the harmonicity of the stimuli. In particular, we wanted to examine whether chil-

TABLE VI. Results of matched *t*-tests done on within-group weighting coefficients across the VOICED and SINE-WAVE conditions for experiment 1.

	<i>df</i>	<i>t</i>	<i>p</i>
Transition			
Adults	19	2.16	0.04
7-year-olds	19	0.16	ns
5-year-olds	21	-0.80	ns
Noise			
Adults	19	-1.18	ns
7-year-olds	19	0.16	ns
5-year-olds	21	1.89	0.07

dren would show diminished weighting of formant transitions. Clearly that was not found: 7-year-olds' weighting of the two acoustic cues examined were consistent across testing with these stimulus types, and 5-year-olds decreased the weight they assigned to the fricative noises only (not to the formant transitions) in the SINE-WAVE condition, compared to the VOICED condition. Consequently, no support was garnered for the hypothesis that children weight formant transitions strongly in their phonetic decisions regarding syllable-initial fricatives because they adhere to the harmonicity principle. In fact, it is interesting that adults were the only listeners to decrease the amount of weight they gave to formant transitions in making these decisions when the vocalic portions were sine waves.

The results reported here were obtained in spite of the fact that sine-wave stimuli are very unnatural in quality. Regardless, children showed similar weighting of formant transitions for both sets of stimuli.

## III. EXPERIMENT 2: VOICED AND WHISPERED VOCALIC PORTIONS

As with experiment 1, the purpose of this second experiment was to test the hypothesis that children's strong weighting of formant transitions in judgments of syllable-initial fricatives is explained by the fact that voiced syllable portions, where formant transitions reside, adhere to the harmonicity principle. Unlike experiment 1, removing harmonicity from vocalic syllable portions in this second experiment meant creating syllables that consisted entirely of noise. This was accomplished by presenting stimuli with whispered vocalic portions. This modification to the speech signal meant that the stimuli were more natural in quality than were the sine waves of experiment 1, but it also meant that formant bandwidths were broader and stimuli consisted entirely of noise.

### A. Method

#### 1. Subjects

Twenty adults between the ages of 18 and 40 years, 20 7-year-olds, and 20 5-year-olds participated. The mean age of 7-year-olds was 7;3, with a range from 7;1 to 7;5. The mean age of 5-year-olds was 5;2, with a range from 4;11 to 5;5. All listeners met the same criteria as in experiment 1. 3-year-olds were not tested in this second experiment both because they had performed similarly to 5-year-olds in the VOICED condition of experiment 1 and because they did so poorly with the stimuli that were not VOICED in that experiment. It seemed neither necessary nor advisable to include them.

#### 2. Equipment and stimuli

The same equipment and VOICED stimuli were used as in experiment 1.

The WHISPERED stimuli used the same synthetic fricative noises as the VOICED and SINE-WAVE stimuli. In these stimuli, however, natural whispered vocalic portions were combined with those synthetic noises. These portions were taken from a male speaker whispering /su/, /sa/, /ʃu/, and /ʃa/. The recordings were sampled at 22.05 kHz and low-

TABLE VII. Means across the three tokens for the phonetically relevant acoustic properties of the vocalic portions of the WHISPERED stimuli used in experiment 2.

Portion	Dur (ms)	F1-onset (Hz)	F2-onset (Hz)	F3-onset (Hz)
(s)a	375	718	1617	2770
(ʃ)a	405	574	1680	2656
(s)u	387	517	1769	2727
(ʃ)u	341	445	2010	2627

pass filtered below 11.025 kHz. Three tokens of each vocalic portion were used. Table VII presents mean values for duration and formant-frequency onsets for each whispered vocalic portion. For /ʃu/ and /s)u/, F1 remained stable throughout the vocalic portion, F2 fell through the entire portion to a mean ending frequency of 1572 Hz, and F3 fell over roughly the first 100 ms to a mean steady-state frequency of 2469 Hz. For /ʃ)a/ and /s)a/, F1 rose over the first 60–90 ms to a steady-state frequency of 969 Hz, and F2 fell to a mean steady-state frequency of 1314 Hz. For /s)a/, F3 remained fairly stable through the entire vocalic portion; for /ʃ)a/, F3 rose over the first 100 ms to a mean steady-state frequency of 2756 Hz. Compared to the voiced vocalic portions, formant frequencies were generally higher in these whispered portions, which matches the findings of (Kallail and Emmanuel 1984a, 1984b). A WHISPERED /ʃ)a/ stimulus with a 3.0 kHz fricative noise is shown on the right-hand side of Fig. 3. It is apparent from this spectrogram that the formants were much less prominent in the whispered stimuli, compared to the sine wave and natural stimuli. Furthermore, in this token, where the frequency of F1 is high, F1 does not appear attenuated. However, in /u/ tokens, where F1 frequency is lower, F1 is greatly attenuated. This is shown in Fig. 5, where a whispered /ʃ)u/ is displayed next to a voiced /ʃ)u/. This figure also more clearly shows that a subglottal

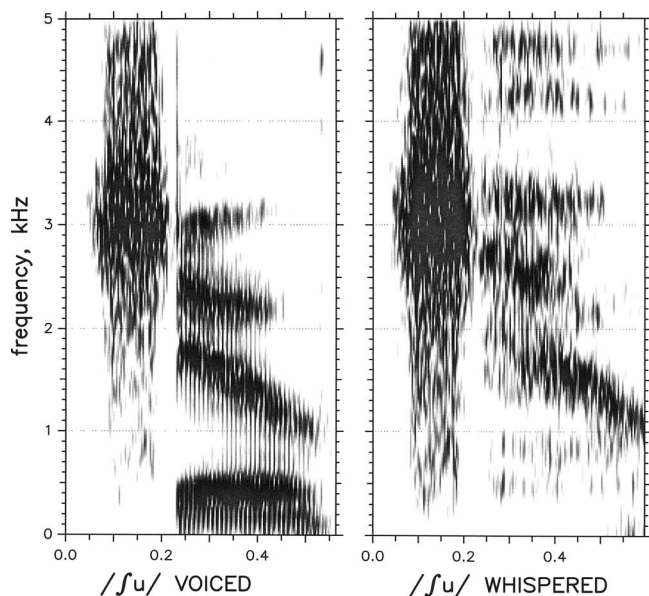


FIG. 5. Spectrograms displaying representative tokens of voiced and whispered /ʃ)u/.

TABLE VIII. Mean weighting coefficients in the VOICED condition of experiment 2. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.37 (0.19)	0.37 (0.17)	0.51 (0.12)
Noise	0.78 (0.12)	0.78 (0.14)	0.65 (0.14)

resonance was present at roughly 1500 Hz in the whispered tokens, making the measurement of F2 frequency tricky.

One issue that arose in creating the WHISPERED stimuli concerned the amplitude ratio of the fricative noise and vocalic syllable portions. In natural speech, whispered vocalic portions are much lower in amplitude than the fricative noises. When we tried to combine fricative and whispered portions with equal amplitudes, as we had done for the other stimuli, the two syllable portions did not cohere perceptually. Instead, it was necessary to combine them with the whispered vocalic portion 6 dB lower than the fricative-noise portion.

### 3. Procedures

The same procedures were used as in experiment 1.

## B. Results

### 1. VOICED condition

Table VIII shows mean weighting coefficients for the VOICED condition. As in experiment 1, there appear to be differences in cue weightings for 5-year-olds compared to listeners in the two older groups: It appears that 5-year-olds weighted formant transitions more and fricative noises less than adults or 7-year-olds did. To examine these apparent findings, one-way ANOVAs, with age as the main effect, were done on weighting coefficients for transitions and fricative noises, separately, for results from the VOICED condition in experiment 2. Results are shown in Table IX. There was a significant main effect of age for both formant transitions and fricative noises. *Post hoc t* tests confirmed that

TABLE IX. Results of ANOVAs and *post hoc t*-tests done on weighting coefficients in the VOICED condition of experiment 2. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	<i>df</i>	<i>F</i> or <i>t</i>	<i>p</i>	Bonferroni significance
Transition				
Age effect	2,57	4.85	0.011	
Adults vs 7-year-olds	57	0.02	ns	
Adults vs 5-year-olds	57	2.71	0.009	0.05
7- vs 5-year-olds	57	2.69	0.009	0.05
Noise				
Age effect	2,57	6.23	0.004	
Adults vs 7-year-olds	57	-0.01	ns	
Adults vs 5-year-olds	57	-3.06	0.003	0.01
7- vs 5-year-olds	57	-3.05	0.004	0.05

TABLE X. Mean weighting coefficients in the WHISPERED condition of experiment 2. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.35 (0.26)	0.21 (0.18)	0.35 (0.22)
Noise	0.75 (0.15)	0.81 (0.12)	0.69 (0.18)

5-year-olds weighted formant transitions more and fricative noises less than did 7-year-olds or adults.

Next we examined if there were any differences between the VOICED conditions in experiments 1 and 2. To accomplish that goal two-group *t*-tests were performed for each age group separately. There was no significant difference in weighting of formant transitions or of fricative noises across experiments for any age group. Consequently we may conclude that the listeners in this experiment were similar to those in experiment 1 in terms of their perceptual weighting strategies.

## 2. WHISPERED condition

75% of 7-year-olds and 50% of 5-year-olds were able to complete the WHISPERED condition with 80% accurate responses on best exemplars during testing. These percentages were lower than the 80% of 7-year-olds and 75% of 5-year-olds who were able to complete the SINE-WAVE condition in experiment 1. Apparently whispered speech was more difficult than sine-wave speech for children to recognize.

Table X presents mean weighting coefficients for the WHISPERED condition. When these coefficients are compared to those in Table VIII, for the VOICED condition, it is apparent that adults performed similarly across the two conditions. Both 7- and 5-year-olds, however, showed greatly diminished weighting of the formant transitions; the weights they assigned to fricative noises appear similar across conditions.

ANOVAs were done on weighting coefficients for transitions and fricative noises separately, and results are shown in Table XI. The main effect of age for transition was significant. *Post hoc t*-tests revealed that for the first time 7-year-olds performed differently from adults, having greatly decreased the weights they assigned to formant transitions in these WHISPERED stimuli. Now 7-year-olds weighted formant transitions less than adults. 5-year-olds similarly decreased the weights they assigned to formant transitions, but in this case that change brought them inline with the weights used by adults.

Although no great changes in weighting coefficients for fricative noises were observed, the small shifts across groups led to changes from previous conditions in statistical outcomes. In particular, adults' slight decrease and 5-year-olds' slight increase in weighting coefficients meant that there was no statistically significant difference between these groups. As before, there was a significant difference between 7- and 5-year-olds.

TABLE XI. Results of ANOVAs and *post hoc t*-tests done on weighting coefficients in the WHISPERED condition of experiment 2. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	<i>df</i>	<i>F</i> or <i>t</i>	<i>p</i>	Bonferroni significance
Transition				
Age effect	2,57	3.29	0.044	
Adults vs 7 year-olds	57	-2.21	0.032	0.10
Adults vs 5-year-olds	57	0.03	ns	
7- vs 5-year-olds	57	2.24	0.029	0.10
Noise				
Age effect	2,57	4.21	0.02	
Adults vs 7-year-olds	57	1.50	ns	
Adults vs 5-year-olds	57	-1.40	ns	
7- vs 5-year-olds	57	-2.90	0.005	0.05

## 3. Across-conditions comparison

As in experiment 1, the most relevant analysis for the main hypothesis being tested was the one for each group separately, across stimulus conditions. Matched *t*-tests were done comparing performance of each group across the VOICED and WHISPERED conditions in this second experiment, and the results are shown in Table XII. Here we see a significant effect of stimulus type for the weights assigned to formant transitions for both groups of children, but not for adults. In fact, the magnitude of change in those weights across conditions is remarkably similar for 7- and 5-year-olds. For example, if we compute effect sizes for the changes in weights using Cohen's *d* (Cohen, 1988), we find  $d=0.91$  for 7-year-olds and  $d=0.90$  for 5-year-olds. Clearly these young listeners had difficulty recovering and/or using these noisy formant transitions in their phonetic decisions.

The slight changes in weights assigned to fricative noises across conditions were not statistically significant for any group.

## C. Discussion

The purpose of this experiment was to test the hypothesis that listeners would change their perceptual weighting strategies when whispered vocalic portions replaced natural vocalic portions, consequently eliminating the harmonicity of the stimuli. In particular, we wanted to examine whether

TABLE XII. Results of matched *t*-tests done on within-group weighting coefficients across the VOICED and WHISPERED conditions for experiment 2.

	<i>df</i>	<i>t</i>	<i>p</i>
Transition			
Adults	19	0.36	ns
7-year-olds	19	3.62	0.002
5-year-olds	19	4.31	<0.001
Noise			
Adults	19	0.93	ns
7-year-olds	19	-0.97	ns
5-year-olds	19	-1.21	ns

children would show diminished weighting of formant transitions when these whispered vocalic portions were used.

Adults performed similarly in the VOICED and WHISPERED conditions, while children performed differently in the two conditions. In stark contrast to what was found in experiment 1, where harmonicity was disrupted by using sine-wave speech, children were found to greatly diminish the weights they assigned to formant transitions in this experiment with whispered speech. Of course, a lack of harmonicity could not readily explain this observed shift in children's weighting strategies: Stimuli in experiment 1 similarly lacked harmonicity, and children's weighting strategies were not affected. That being the case, the question arose as to what could explain these observed changes in perceptual weighting.

This second experiment showed that children greatly reduced the weights they assigned to formant transitions when whispered vocalic portions were used in stimulus construction. Either children were unable to recover the patterns of global spectral change in these noisy signals or they chose not to tune their perceptual attention to these signal components. Whatever the reason might be, whispered speech is very natural, so the difficulty could not have been due to a lack of naturalness. We suspect that instead the difficulty was due to the broadened formant structure of whispered speech. As Figs. 3 and 5 illustrate, whispering results in formants that are less prominent than in voiced speech; energy is spread more evenly across the spectrum. This lack of prominence for formants could make it difficult for children to recover the dynamic spectral patterns associated with changing vocal-tract cavity shapes and sizes. In some sense, listening to whispered speech is like listening to speech through the widened auditory filters of hearing loss. Turner and Van Tassel (1984) reported that adult listeners did not encounter difficulty listening to speech under such conditions. Our results for adults with normal hearing listening to whispered speech match those results for adults with hearing loss: They did not modify their weighting strategies for the whispered stimuli. In fact, it is interesting that adults in this current study actually decreased their weighting of formant transitions when sine-wave replicas of the vocalic portions were used in stimulus construction. This finding suggests that adults might be hindered slightly by processing algorithms that emphasize the spectral prominences of formants. A study by Hillenbrand and Houde (2002) in which adult listeners showed slightly poorer consonant and vowel recognition for spectrally enhanced speech signals supports that suggestion, although a different study by Summerfield *et al.* (1981) did not find that adults were hindered in their speech recognition when formants were made more spectrally prominent.

But regardless of whether adults are hindered or not by algorithms that enhance the spectral prominences (i.e., formants) of speech, it is clear that they are not hurt by broadened bandwidths. Our results for children, on the other hand, suggest quite a different situation: Children did encounter difficulty when the formants were not spectrally prominent, as occurred with the whispered stimuli. This finding matches other results coming from this laboratory. In particular, Nittrouer *et al.* (2009) found that 7-year-old children were

poorer at recognizing noise vocoded sentences than adults, but were similarly skilled at recognizing the same sentences when they were presented as sine waves. Even more so than whispered stimuli, noise vocoded speech broadens the bandwidths of the spectral prominences. If children with normal hearing and typical speech development demonstrate these constraints in their processing of such signals, it is reasonable to expect that children with hearing loss would also have difficulty processing signals with reduced spectral prominences. Perhaps children with hearing loss would benefit from algorithms that track formants in the signal and shape the spectral output to make those formants more prominent.

Of course, signals in this second experiment, as well as those in Nittrouer *et al.* (2009), shared another attribute. In both cases they were constructed entirely of noise. That means that possibly the shift away from weighting formant transitions strongly demonstrated by this experiment and by the poor sentence recognition found in Nittrouer *et al.* (2009) for noisy stimuli might be attributable to children failing to attend to noise components in speech signals. The current study cannot really separate the contributions made to children's results of broadened bandwidths and noisy signals. Future investigations will need to examine that distinction.

#### IV. GENERAL DISCUSSION

Previous research has revealed that adults and children assign different weights to the various components of the acoustic speech signal. Adults, it seems, are quite attentive to the spectrotemporal details of the signal. These are properties that arise from precise timing relations among articulatory actions, or exact constriction shapes. In the production of a fricative, for example, many variables come into play: the place of the constriction in the vocal tract must be accurate; the degree of constriction between articulatory structures must be tight enough that turbulence noise is generated, but not so tight that stopping is introduced; the length and shape of the constriction must be right; and any secondary cavities that help mold the acoustic output, such as a sublingual cavity in the case of English /ʃ/, must be created (Perkell *et al.*, 1979). Learning to produce speech with all these parameters properly implemented takes time, as apparently does learning to pay attention to their acoustic consequences so that phonetic structure may be accurately and efficiently recovered.

Unlike adults, research into children's speech perception and word recognition has uncovered consistent evidence that children pay particular attention to global spectral patterns within the acoustic signal (i.e., ongoing change in formant frequencies), rather than to the details that capture adults' attention. Previously it was hypothesized that this perceptual strategy is preferred early in life because these global, recurrent patterns of structure help children isolate linguistically significant stretches, such as words (e.g., Jusczyk, 1993; Nittrouer *et al.*, 2009; Studdert-Kennedy, 1987). However, careful scientific inquiry dictates that all alternative explanations be explored before accepting just one. It may be that the enhanced weighting observed in children's speech perception

for global changes in formant frequencies have to do with nonlinguistic considerations. In the study reported here we explored the possibility that children's preferential weighting of formant transitions in decisions of syllable-initial fricatives may actually be explained by the fact that the portion of formant transitions involved in stimulus design in early experiments was always voiced, while the fricative noises were always just that—noise. Perhaps children attend to signal portions that adhere to principles of ASA, especially the harmonicity principle. At least that was the hypothesis tested.

Results of these experiments led us to reject that hypothesis: When the principle of harmonicity was disrupted in the stimuli by using sine-wave replicas of the vocalic portions, children assigned the same weights to formant transitions as they did for stimuli with natural vocalic portions. Consequently, the assertion that children rely on the slowly modulating formants in the speech signal, which arise from fluid changes in vocal-tract cavity shapes and sizes, remains supportable. But that was not the end of this story. It was also found that children experience difficulty attending to formant transitions when those transitions are not spectrally prominent and consist only of noise. It was suggested that if the fact that the whispered stimuli lacked spectrally prominent formants explained the lion's share of difficulty encountered by children it would mean that children have more difficulty than adults perceiving speech under the condition of sensorineural hearing loss.

In summary, the two experiments reported here were designed to investigate whether listeners, particularly children, would show decreased weighting of formant transitions when vocalic portions failed to provide common harmonic structure across the formants. Children were not found to decrease their weightings of formant transitions in the first experiment, and the suggestion was made that results of the second experiment may be explained by difficulty on the part of children in recovering formant transitions when formants are not spectrally prominent and/or consist only of noise. Future research will need to test the veracity of that suggestion.

## ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC-00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to S.N.

- Barker, J., and Cooke, M. P. (1999). "Is the sine-wave speech cocktail party worth attending?," *Speech Commun.* **27**, 159–174.
- Beddor, P. S., and Strange, W. (1982). "Cross-language study of perception of the oral-nasal distinction," *J. Acoust. Soc. Am.* **71**, 1551–1561.
- Bregman, A. S. (1990). *Auditory Scene Analysis* (MIT, Cambridge, MA).
- Charles-Luce, J., and Luce, P. A. (1990). "Similarity neighbourhoods of words in young children's lexicons," *J. Child Lang.* **17**, 205–215.
- Charles-Luce, J., and Luce, P. A. (1995). "An examination of similarity neighbourhoods in young children's receptive vocabularies," *J. Child Lang.* **22**, 727–735.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ).
- Cole, R. A., and Perfetti, C. A. (1980). "Listening for mispronunciations in a children's story: The use of context by children and adults," *J. Verbal Learn. Verbal Behav.* **19**, 297–315.
- Crowther, C. S., and Mann, V. (1992). "Native language factors affecting use of vocalic cues to final consonant voicing in English," *J. Acoust. Soc. Am.* **92**, 711–722.
- 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.
- Flege, J. E., and Port, R. (1981). "Cross-language phonetic interference: Arabic to English," *Lang Speech* **24**, 125–146.
- Goldman, R., and Fristoe, M. (2000). *Goldman-Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).
- Goodell, E. W., and Studdert-Kennedy, M. (1993). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study," *J. Speech Hear. Res.* **36**, 707–727.
- Harris, K. S. (1958). "Cues for the discrimination of American English fricatives in spoken syllables," *Lang Speech* **1**, 1–7.
- Heinz, J. M., and Stevens, K. N. (1961). "On the properties of voiceless fricative consonants," *J. Acoust. Soc. Am.* **33**, 589–593.
- Hillenbrand, J. M., and Houde, R. A. (2002). "Speech synthesis using damped sinusoids," *J. Speech Lang. Hear. Res.* **45**, 639–650.
- Jastak, S., and Wilkinson, G. S. (1984). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).
- Jusczyk, P. W. (1982). "Auditory versus phonetic coding of speech signals during infancy," in *Perspectives on Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities*, edited by J. Mehler, M. F. Garrett, and E. C. T. Walker (Erlbaum, Hillsdale, NJ), pp. 361–387.
- Jusczyk, P. W. (1993). "From general to language-specific capacities: The WRAPSA model of how speech perception develops," *J. Phonetics* **21**, 3–28.
- Kallail, K. J., and Emanuel, F. W. (1984a). "An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects," *J. Phonetics* **12**, 175–186.
- Kallail, K. J., and Emanuel, F. W. (1984b). "Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects," *J. Speech Hear. Res.* **27**, 245–251.
- Kunisaki, O., and Fujisaki, H. (1977). "On the influence of context upon perception of voiceless fricative consonants," *Ann. Bulletin of the RILP* **11**, 85–91.
- Lisker, L., and Abramson, A. S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 384–422.
- 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.
- MacNeilage, P. F., and Davis, B. L. (1991). "Acquisition of speech production: Frames, then content," in *Attention & Performance XIII*, edited by M. Jeannerod (Erlbaum, New York), pp. 453–476.
- Mann, V. A., and Repp, B. H. (1981). "Influence of preceding fricative on stop consonant perception," *J. Acoust. Soc. Am.* **69**, 548–558.
- Matsuda, M., and Kasuya, H. (1999). "Acoustic Nature of the Whisper," *Eurospeech '99*, pp. 133–136.
- 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.
- McGowan, R. S., and Nittrouer, S. (1988). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /sh/ and /s/," *J. Acoust. Soc. Am.* **83**, 229–236.
- 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. (1996). "The discriminability and perceptual weighting of some acoustic cues to speech perception by three-year-olds," *J. Speech Hear. Res.* **39**, 278–297.
- Nittrouer, S. (2002). "Learning to perceive speech: How fricative perception changes, and how it stays the same," *J. Acoust. Soc. Am.* **112**, 711–719.
- 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.
- 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.* in press.
- Nittrouer, S., Manning, C., and Meyer, G. (1993). "The perceptual weighting of acoustic cues changes with linguistic experience," *J. Acoust. Soc. Am.* **94**, S1865.
- Perkell, J. S., Boyce, S. E., and Stevens, K. N. (1979). "Articulatory and acoustic correlates of the [sʃ] distinction," *J. Acoust. Soc. Am.* **65**, S24.
- Pittman, A. L., Stelmachowicz, P. G., Lewis, D. E., and Hoover, B. M. (2003). "Spectral characteristics of speech at the ear: Implications for amplification in children," *J. Speech Lang. Hear. Res.* **46**, 649–657.
- 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.
- Ruff, H. A. (1982). "Effect of object movement on infants' detection of object structure," *Dev. Psychol.* **18**, 462–472.
- Serniclaes, W., Sprenger-Charolles, L., Carre, R., and Demonet, J. F. (2001). "Perceptual discrimination of speech sounds in developmental dyslexia," *J. Speech Lang. Hear. Res.* **44**, 384–399.
- Spelke, E., von Hofsten, C., and Kestenbaum, R. (1989). "Object perception in infancy: Interaction of spatial and kinetic information for object boundaries," *Dev. Psychol.* **25**, 185–196.
- Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (2007). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," *Ear Hear.* **28**, 483–494.
- Stevens, K. N. (1975). "The potential role of property detectors in the perception of consonants," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York), pp. 303–330.
- Studdert-Kennedy, M. (1987). "The phoneme as a perceptuomotor structure," in *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing*, edited by A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer (Academic, Orlando), pp. 67–84.
- Summerfield, Q., Tyler, R., Foster, J., Wood, E., and Bailey, P. J. (1981). "Failure of formant bandwidth narrowing to improve speech reception in sensorineural impairment," *J. Acoust. Soc. Am.* **70**, S108–S109.
- Tartter, V. C. (1989). "What's in a whisper?," *J. Acoust. Soc. Am.* **86**, 1678–1683.
- Tartter, V. C. (1991). "Identifiability of vowels and speakers from whispered syllables," *Percept. Psychophys.* **49**, 365–372.
- Tice, B., and Carrell, T. (1997) TONE v.1.5b (software).
- Tsunoda, K., Ohta, Y., Soda, Y., Niimi, S., and Hirose, H. (1997). "Laryngeal adjustment in whispering magnetic resonance imaging study," *Ann. Otol. Rhinol. Laryngol.* **106**, 41–43.
- Turner, C. W., and Holte, L. A. (1987). "Discrimination of spectral-peak amplitude by normal and hearing-impaired subjects," *J. Acoust. Soc. Am.* **81**, 445–451.
- Turner, C. W., and Van Tasell, D. J. (1984). "Sensorineural hearing loss and the discrimination of vowel-like stimuli," *J. Acoust. Soc. Am.* **75**, 562–565.
- Vouloumanos, A., and Werker, J. F. (2007). "Listening to language at birth: Evidence for a bias for speech in neonates," *Dev. Sci.* **10**, 159–164.
- Walley, A. C., Smith, L. B., and Jusczyk, P. W. (1986). "The role of phonemes and syllables in the perceived similarity of speech sounds for children," *Mem. Cognit.* **14**, 220–229.