

# Learning to perceive speech: How fricative perception changes, and how it stays the same

Susan Nittrouer

*Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131*

(Received 25 May 2001; accepted for publication 28 May 2002)

A part of becoming a mature perceiver involves learning what signal properties provide relevant information about objects and events in the environment. Regarding speech perception, evidence supports the position that allocation of attention to various signal properties changes as children gain experience with their native language, and so learn what information is relevant to recognizing phonetic structure in that language. However, one weakness in that work has been that data have largely come from experiments that all use similarly designed stimuli and show similar age-related differences in labeling. In this study, two perception experiments were conducted that used stimuli designed differently from past experiments, with different predictions. In experiment 1, adults and children (4, 6, and 8 years of age) labeled stimuli with natural /f/ and /θ/ noises and synthetic vocalic portions that had initial formant transitions varying in appropriateness for /f/ or /θ/. The prediction was that similar labeling patterns would be found for all listeners. In experiment 2, adults and children labeled stimuli with initial /s/-like and /ʃ/-like noises and synthetic vocalic portions that had initial formant transitions varying in appropriateness for /s/ or /ʃ/. The prediction was that, as found before, children would weight formant transitions more and fricative noises less than adults, but that this age-related difference would elicit different patterns of labeling from those found previously. Results largely matched predictions, and so further evidence was garnered for the position that children learn which properties of the speech signal provide relevant information about phonetic structure in their native language. © 2002 Acoustical Society of America.

[DOI: 10.1121/1.1496082]

PACS numbers: 43.71.An, 43.71.Ft, 43.71.Es [CWT]

## I. INTRODUCTION

Perception is the extraction of information about the affordances of things in the world, according to Eleanor Gibson (1991b). The information to be extracted is always there, Gibson tells us, “like truth” (1977/1991a, p. 474). This does not mean, however, that all perceivers extract the same information about events. What is extracted depends upon several factors, with Gibson listing three as most important: the species of the organism, developmental maturity, and learning. When it comes to understanding how children learn to perceive the speech signal, we are most interested in the role that learning plays in the acquisition of competency because learning is the only one of the three factors that we can influence. Research with speakers/listeners of different languages has demonstrated robustly that the information extracted from the speech signal is highly dependent on the native language of the individual (Strange, 1995, offers a substantive review of the literature on cross-linguistic speech perception). Clearly, then, children must learn what information should be extracted in their native language. But, what determines the information that the child needs to learn to extract? Gibson tells us that a guide to answering this question is to think about the goal of perception. For example, the goal of perception for wine and tea tasters is to recognize various flavors, and so they must learn to extract information about subtle variances in the substances creating those flavors. In the case of speech, the goal of perception is to apprehend linguistic structure. For the purposes of most speech perception studies, the structure of interest is phonetic struc-

ture, and so it is here. In matters of speech perception, children must learn to extract the information that will permit access to phonetic structure in their native language.

One final principle described by Gibson (1977/1991a) that guides our investigation of how children learn to perceive speech is the notion that there is increasing specificity of correspondence between the information in the world that is extracted and in what is perceived. According to this account, events can be perceived from birth, but without as much differentiation among them as there will be later. As the level of differentiation increases, as it must, the information extracted from the environment becomes more specifically related to the events being perceived: that is, the child learns what information is relevant.

The hypothesis that has emerged to capture these developmental changes as they relate to speech perception has been termed the “developmental weighting shift” (or DWS). This hypothesis suggests simply that the informational aspects of the signal that get weighted greatly (or, attended to greatly) change as children gain experience listening to and speaking their native language. Initially, children seem to attend largely to those acoustic properties that specify a changing vocal tract, which means the dynamic spectral changes known as formant transitions. Then, as the child becomes more skilled, attention shifts to acoustic properties that do not involve spectral change, properties such as silent gaps (specifying periods of vocal-tract closure), differences in voicing duration (specifying the voicing of syllable-final

stops), or periods of stable spectral information (specifying place of consonantal constrictions).

One series of experiments tracing this developmental shift has focused on /s/-vowel and /ʃ/-vowel syllables (Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987). For adult English speakers, the decision about which of these sibilants occurred syllable initially is largely based on the spectrum of the noise (e.g., Harris, 1958). As would be predicted from the above account, young children base their responses not so much on the noise spectrum (when compared to adults), but more on the formant transitions that arise from movement as the vocal tract goes from the consonantal constriction to the vocalic constriction. A developmental change has been observed in the relative amounts of attention (or weight) paid to these acoustic properties for children between 3½ and 7½ years of age. However, these experiments all used similarly designed stimuli that had fricative noises varying along acoustic continua and formant transitions largely appropriate for initial /s/ or /ʃ/. The purpose of the experiments reported here was to test predictions arising from the DWS using stimuli designed differently.

First, the DWS hypothesis predicts that adults and children should perform similarly for a labeling decision that adults make based primarily on formant transitions. That is, if perceptual strategies early in life put particular weight on formant transitions, we should be able to find phonetic contrasts that require no modification from those early strategies because paying attention to transitions is the most efficient strategy for apprehending phonetic structure. Harris (1958) provides evidence of just such a decision. In that experiment, noises from unvoiced /f/, /θ/, /s/, and /ʃ/, as well as their voiced counterparts, were cross spliced with vocalic portions that had formant transitions appropriate for one of these syllable-initial fricatives. Harris showed that adult listeners make decisions about whether the fricative was /f/ or /θ/ based largely on the formant transitions, while decisions about whether it was /s/ or /ʃ/ are based largely on the noise spectrum. Thus, the decision about whether a syllable-initial fricative is /f/ or /θ/ should be one for which competent English speakers rely on perceptual strategies that do not require modification from their earliest state, and so adults and children should show similar results on labeling tasks. To test this prediction, stimuli were used in which formant transitions varied along continua from those appropriate for a syllable-initial /θ/ to those appropriate for a syllable-initial /f/, and with either a /θ/ or /f/ noise. This arrangement of acoustic properties was, in a sense, opposite to that of the earlier /ʃ/ versus /s/ labeling experiments. In those earlier experiments, the fricative noises formed the continuum, and formant transitions were set appropriately for one or the other fricative condition. However, that arrangement of acoustic properties across stimuli is not possible in experiments designed to investigate /f/ versus /θ/ perception because there simply is not enough of a difference between the spectra of these noises to form a continuum.

The second prediction of the DWS tested in this study arose precisely from this difference in stimulus design. According to the DWS, even if we reverse the arrangement of

acoustic properties in an /ʃ/ versus /s/ labeling experiment, children should continue to show evidence of weighting formant transitions more, and fricative noises less, than adults. This prediction was tested in a second experiment reported here, in which the arrangement of acoustic properties across stimuli was different from past experiments examining developmental changes in /ʃ/ versus /s/ perception (Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987). Here, acoustic properties were arranged as they were in the /f/ versus /θ/ experiment, with formant transitions varying along a continuum and two fricative noises set to favor one or the other fricative label.

In summary, two experiments were conducted to test rigorously the predictions of the DWS. In the first experiment examining /f/ versus /θ/ perception, listeners of all ages should show the same weighting strategies. In the second experiment examining /ʃ/ versus /s/ perception, a developmental decrease in the weight assigned to formant transitions should be found, with a developmental increase in the weight assigned to the fricative-noise spectra.

## II. ANALYSIS OF PREVIOUS STUDY

Nittrouer and Miller (1997b) serves as an example of a study tracing the emergence of mature perceptual weighting strategies for /s/-vowel versus /ʃ/-vowel decisions. In the three experiments reported there, the fricative noise varied along a nine-step continuum from one appropriate for /ʃ/ to one appropriate for /s/. Four vocalic portions were used in each experiment: two /a/ portions and two /u/ portions, with formant transitions appropriate for either a syllable-initial /ʃ/ or /s/. In experiment 1 these vocalic portions were natural. In experiments 2 and 3 they were synthetic, either with *F2* alone varying as appropriate for syllable-initial /ʃ/ or /s/ (experiment 2), or with both *F2* and *F3* varying as appropriate for syllable-initial /ʃ/ or /s/ (experiment 3). Listeners labeled stimuli as starting with /ʃ/ or /s/, and responses were plotted as the proportion of “s” responses at each level of fricative noise, for each vocalic portion separately. Cumulative normal distributions from experiment 3 of that study are shown in Fig. 1 for adults, 7-, 5-, and 3½-year-olds. Information about the relative weighting of the fricative noises and formant transitions can be gleaned from the shapes of the labeling functions and from the separation between those functions based on formant transitions: The steeper the functions, the more weight that was assigned to the fricative-noise spectrum; the greater the separation between functions for stimuli with /ʃ/ and /s/ transitions, the more weight that was assigned to the formant transitions.

Of course, estimating the relative weights assigned to the fricative noises and formant transitions in this way does not provide a precise metric. Another way to index these weights is by doing regression analyses, and that was done here. For data from each listener, the proportion of “s” responses given at each level of the two acoustic properties manipulated (fricative-noise spectrum and formant transitions) was correlated with these two acoustic properties, using the arcsine transforms of the proportions themselves. These analyses were done for the first and third experiments of Nittrouer and Miller (1997b) because these two experi-

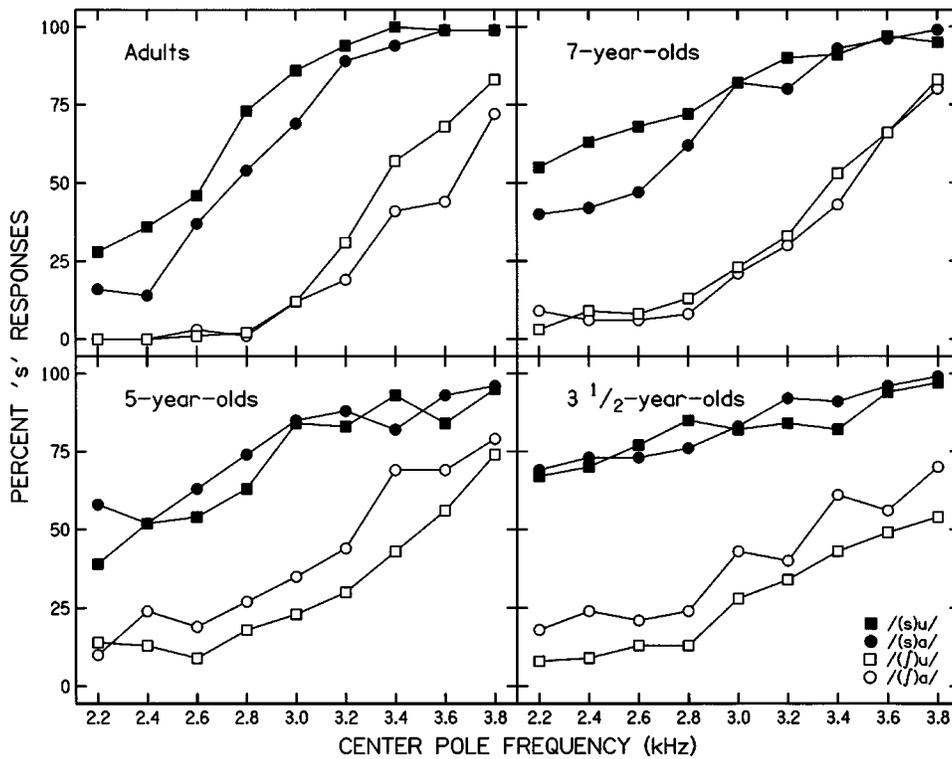


FIG. 1. Labeling functions for /f/-vowel and /s/-vowel stimuli in which noise spectra varied along a nine-step continuum and formant transitions varied dichotomously. (Reprinted from Nittrouer and Miller, 1997b, experiment 3).

ments had both  $F2$  and  $F3$  varying as appropriate for syllable-initial /f/ or /s/. Tables I and II show mean partial correlation coefficients derived for each age group in the two experiments, with standard deviations (s.d.'s) given in parentheses. Although the exact correlations for any age group differ across the studies, a general trend is easily discernible: With increasing age, correlation coefficients for noise increase while those for formant transitions decrease. Welch's equality of means tests were performed on correlation coefficients for each acoustic property to test the significance of these trends. This test is similar to analyses of variance (ANOVAs), but does not assume homogeneity of variance. For experiment 1 of Nittrouer and Miller, the effect of age was significant for correlations involving both noise,  $W(2,35)=10.83$ ,  $p<0.001$ , and formant transitions,  $W(2,38)=5.16$ ,  $p=0.01$ . For experiment 3, the age effect was also significant for both kinds of correlations: noise,  $W(3,22)=5.67$ ,  $p=0.005$ ; formant transitions,  $W(3,18)=7.70$ ,  $p=0.002$ . In summary, these correlational analyses capture nicely the fact that experiments have reliably demonstrated a developmental decrease in the weight assigned to formant transitions and a developmental increase in the

weight assigned to fricative-noise spectra in decisions of fricative identity for syllable-initial /f/ and /s/.

### III. EXPERIMENT 1: /f/ VERSUS /θ/ PERCEPTION

The purpose of this experiment was to test the prediction that adults and children would show similar perceptual strategies in labeling /f/-vowel and /θ/-vowel syllables because listeners of all ages would be found to weight formant transitions heavily, but not the noise spectra. This similarity in weighting strategies should give rise to correlation coefficients that do not differ across age groups.

#### A. Method

##### 1. Listeners

To participate, listeners were required to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz, presented at 25 dB HL. American English was the native language of all adults participating, and all children lived in monolingual English home environments. Children needed to score at or above the 30th percentile on the Goldman-Fristoe Test of Articulation (Goldman and Fristoe, 1986), and they needed to have negative histories of otitis media, defined as less than six episodes during the first 2 years of life.

TABLE I. Mean partial correlation coefficients for each group in experiment 1 of Nittrouer and Miller (1997b). Standard deviations (s.d.'s) are given in parentheses.

	4½-year-olds	7-year-olds	Adults
Experiment 1			
Fricative noise	0.683 (0.126)	0.770 (0.101)	0.826 (0.061)
Formant transitions	0.459 (0.135)	0.366 (0.139)	0.324 (0.133)

TABLE II. Mean partial correlation coefficients for each group in experiment 3 of Nittrouer and Miller (1997b). s.d.'s are given in parentheses.

	3½-year-olds	5-year-olds	7-year-olds	Adults
Experiment 3				
Fricative noise	0.367 (0.182)	0.581 (0.135)	0.599 (0.200)	0.664 (0.147)
Formant transitions	0.774 (0.139)	0.589 (0.162)	0.619 (0.167)	0.554 (0.148)

Adults were required to demonstrate at least an 11th grade reading level on the Wide-Range Achievement Test-Revised (Jastak and Wilkinson, 1984). Meeting these requirements were seventeen 4-year-olds, thirteen 6-year-olds, fourteen 8-year-olds, and 13 adults. However, four of these 4-year-olds were subsequently dismissed because they were unable to label reliably the best exemplars of the category labels. The mean ages (and ranges) of the children included in the study, given in years; months, were 4; 3 (3; 11 to 4; 6), 6; 1 (5; 11 to 6; 4), and 8; 2 (7; 11 to 8; 5). The mean age of adults was 30 years, with the range between 21 years and 40 years.

## 2. Equipment and materials

All testing took place in a sound-attenuated 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 by way of a Data Translation DT-2801A digital-to-analog converter, a Frequency Devices 901-F filter, a Crown D-75 amplifier, and AKG-K141 headphones. Responses were recorded with a keyboard connected to the computer. Four hand-drawn pictures (8 in. × 8 in.) were used to represent each response label (“tha,” “thu,” “fa,” or “fu”). These pictures were of animate objects, given the names of the response labels. Recorded stories were presented via a Nakamichi MR-2 audiocassette player, with a Tascam PA-30B amplifier and AKG-K141 headphones. Hand-drawn story books accompanied the taped stories, and the object named with the response label figured prominently on every page. 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 high-graphics 14-in. color monitor after completion of each block of stimuli. A bell sounded while the color graphics were being presented and served as additional reinforcement.

## 3. Stimuli

Stimuli consisted of natural /f/ and /θ/ noises combined with synthetic /a/ and /u/ vocalic portions. In order to create stimuli that matched natural productions of /fa/, /θa/, /fu/, and /θu/, five tokens of each of these syllables were recorded by four male speakers. None of these speakers had any history of speech or language problems, and all passed a hearing screening of the frequencies 0.5, 1.0, 2.0, and 4.0 kHz presented at 20 dB HL. All were native speakers of American English, with Midwestern dialects. Five measurements were obtained from these samples: relative amplitude of the noise portion of the syllable, compared to the vocalic portion (dB), duration of the noise (ms), the first spectral moment of the noise (kHz), *F*2 for the first pitch period (Hz), and *F*3 for the first pitch period (Hz). Means (and s.d.’s) of each of these measures for each syllable type are shown in Table III.

Four fricative noises were used for stimulus creation because /f/ and /θ/ noises separated from natural fricative-/a/ syllables were combined with the synthetic /a/ portions, and /f/ and /θ/ noises separated from natural fricative-/u/ syl-

TABLE III. Mean measurements across tokens produced by four adult, male speakers of the syllables /θa/, /fa/, /θu/, and /fu/ (5 tokens of each syllable per speaker). Five measurements were made: relative amplitude of the noise to the vocalic portion (dB), duration of the noise (ms), the first spectral moment (kHz), *F*2 for the first pitch period (Hz), and *F*3 for the first pitch period (Hz). s.d.’s are given in parentheses.

	/a/		/u/	
	/θ/	/f/	/θ/	/f/
Relative amplitude	-27.3 (3.0)	-29.5 (3.3)	-25.9 (3.5)	-27.9 (3.1)
Duration	176 (20)	191 (34)	212 (39)	198 (27)
First moment	8.00 (0.61)	7.68 (0.68)	7.88 (0.39)	7.86 (0.47)
<i>F</i> 2	1365 (67)	1092 (46)	1496 (180)	1337 (216)
<i>F</i> 3	2517 (187)	2423 (220)	2391 (143)	2268 (141)

lables were combined with the synthetic /u/ portions. The four noises used were all from the same speaker and closely approximated mean amplitude and first moment measurements for their respective fricative-vowel combination from the acoustic analysis. Spectra of these four noises are shown in Fig. 2. Each was truncated to 190 ms for the purpose of stimulus creation.

Eighteen synthetic vocalic portions were made using the Sensyn Laboratory Speech Synthesizer: nine each of /a/ and /u/. These vocalic portions were created at a 20-kHz sampling rate to match the sampling rate of the noises. Stimuli for both continua varied from one appropriate for /θ/-vowel to one appropriate for /f/-vowel. Parameter settings were based on the acoustic measures of the natural vocalic portions in the acoustic analysis. For all vocalic portions, duration was 300 ms, and *f*0 started at 120 Hz, falling throughout

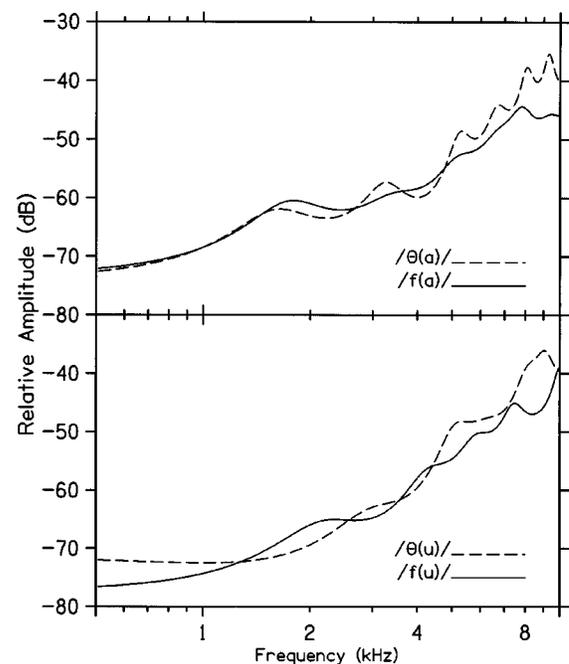


FIG. 2. Spectra of /f/ and /θ/ noises from /a/ and /u/ contexts, spoken by a male adult. These noises were used in experiment 1.

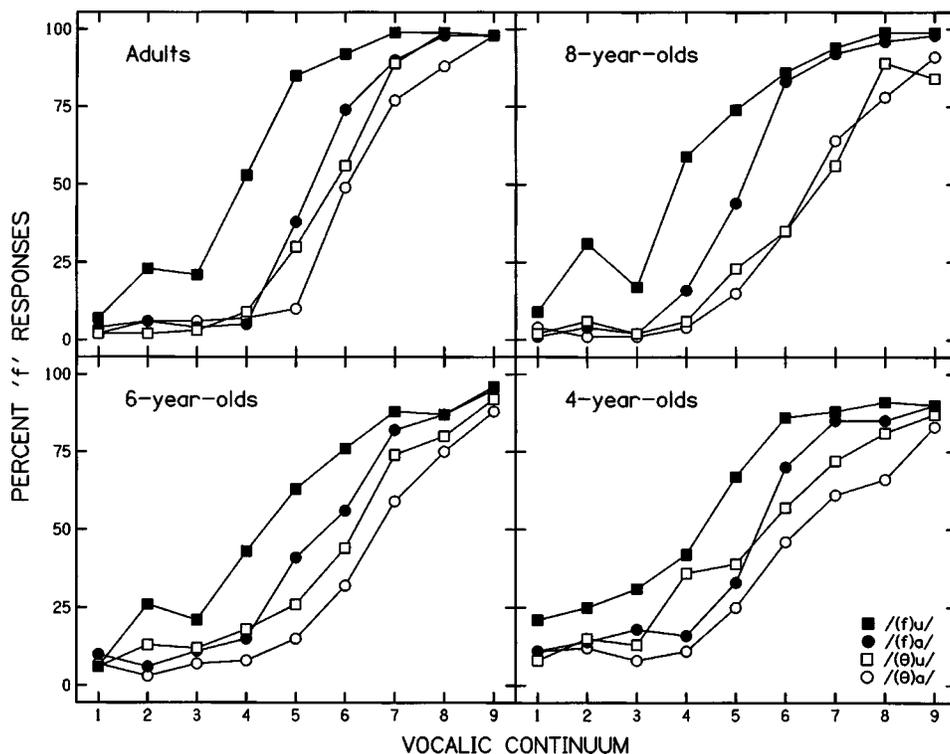


FIG. 3. Labeling functions for /f/-vowel and /θ/-vowel stimuli in which formant transitions varied along continua and noise spectra varied dichotomously (experiment 1).

the portion to an offset frequency of 90 Hz. The vocalic portion was always 22 dB greater in amplitude than the noise portion. For /a/ portions,  $F1$  began at 560 Hz and rose to a steady-state frequency of 750 Hz over the first 70 ms.  $F2$  onset formed a continuum varying in 40-Hz steps from 1410 Hz (most /θ/-like) to 1090 Hz (most /f/-like).  $F2$  changed over the first 70 ms from its onset frequency to a steady-state frequency of 1180 Hz.  $F3$  onset varied in nine steps from a frequency of 2535 Hz (most /θ/-like) to 2415 Hz (most /f/-like) in 15-Hz steps. This formant changed over the first 70 ms from the onset frequency to a steady-state frequency of 2480 Hz. Each of these nine portions was combined with each of the /f/ and /θ/ noises taken from the /a/ context, making 18 stimuli.

For /u/ portions,  $F1$  was 350 Hz at onset and fell very gradually throughout the vocalic portion to an offset frequency of 330 Hz.  $F2$  onset varied in 50-Hz steps from 1520 Hz (most /θ/-like) to 1120 Hz (most /f/-like).  $F2$  fell throughout the vocalic portion to an offset frequency of 940 Hz.  $F3$  onset varied in 20-Hz steps from 2400 Hz (most /θ/-like) to 2240 Hz (most /f/-like).  $F3$  changed throughout the portion to an offset frequency of 2250 Hz. These vocalic portions were combined with the /f/ and /θ/ noises taken from the /u/ context.

#### 4. Procedures

All screening tasks were completed first. For the 4- and 6-year-olds the next step was the presentation of tape-recorded stories about the labels they would be using in the labeling task (either “fa” and “tha” or “fu” and “thu”). Each of these four stories was roughly 3 min long, and the label was the name of an animate object that served as the main character in the story. These stories were heard twice, once with natural speech and once with synthetic speech.

The stories served two purposes. First, they allowed children greater opportunity to learn the stimulus labels. Second, they allowed children to acclimate to synthetic speech prior to testing.

Next, the labeling task was presented. Stimuli with /a/ and /u/ portions were presented separately, making the listener’s task a binary choice. All participants first heard ten practice items (five each of /f/-vowel and /θ/-vowel), which were the best exemplars of each category. For example, the best exemplar of “fa” was the stimulus with the natural /f/ noise and the synthetic vocalic portion in which  $F2$  and  $F3$  onsets were 1090 and 2415 Hz, respectively. To proceed to testing, subjects were required to label correctly at least nine of these ten practice stimuli.

During testing, the stimuli in each vowel set were presented ten times each in randomized blocks of 18. To have their data included in the analyses, participants needed to achieve at least 80% accurate responses for the best exemplars. This requirement insured that only participants who maintained attention during the task were included. Children kept track of their progress by moving the marker to the next space on the gameboard after they completed each block of stimuli. The order of presentation of the /a/ and /u/ stimuli was randomized across listeners.

#### B. Results

Figure 3 shows the labeling functions for all four groups. Clearly, there is less of a developmental change in the labeling of these /f/-vowel and /θ/-vowel stimuli than in the labeling of the /s/-vowel and /ʃ/-vowel stimuli seen in Fig. 1, from experiment 3 of Nittrouer and Miller (1997b): The functions across age groups appear similarly steep, with similar separations among the functions.

TABLE IV. Mean partial correlation coefficients for each group in experiment 1 (/f/ versus /θ/). s.d.'s are given in parentheses.

	4-year-olds	6-year-olds	8-year-olds	Adults
Fricative noise	0.202 (0.099)	0.200 (0.099)	0.293 (0.111)	0.173 (0.070)
Formant transitions	0.822 (0.067)	0.836 (0.065)	0.828 (0.50)	0.875 (0.045)

Table IV shows the mean partial correlation coefficients for each age group in experiment 1, for fricative noise and formant transitions. The coefficients for formant transitions appear similar across age groups, and the statistical analysis supports that observation: there is no significant effect of age.<sup>1</sup> For fricative noise, however, the 8-year-olds show larger correlation coefficients than the other three groups. In keeping with that observation, a significant main effect of age was found,  $W(3,27) = 3.62$ ,  $p = 0.026$ . Because of this finding, *post hoc* *t*-tests were done on these data. Results of these tests (using separate variances) are shown in Table V. Here, we see that 8-year-olds did indeed show a weighting of the fricative noise that was greater than that of all three other groups.

### C. Discussion

The prediction for this experiment was that there would be no differences in the weights assigned to the fricative-noise spectra and formant transitions by listeners of different ages because all listeners would base responses largely on formant transitions, and not very greatly on fricative-noise spectra. To a great extent, this prediction was met. The one finding that did not meet the prediction was that 8-year-olds weighted the fricative noise more than any of the other three groups. This result may be aberrant in this study, and not representative of 8-year-olds' strategies in general. However, it might also represent an instance of overgeneralization. That is, the overarching hypothesis being examined by this work is that initial weighting strategies for speech perception focus on acoustic properties associated with movement: i.e., formant transitions. Then, as the child gains experience with a native language, these strategies are modified. One kind of modification, it is proposed, involves learning to look to acoustic detail at the syllable margins for clues to phonetic structure. In this instance it may be that children overgeneralize this strategy as it is developing, using it in contexts that would best be served by the strategy that they had used ear-

TABLE V. Results of *post hoc* *t*-tests for experiment 1.

	Noise		
	<i>df</i>	<i>t</i>	<i>p</i>
4- vs 6-year-olds	23	0.06	ns
4- vs 8-year-olds	24	-2.24	0.034
4- vs adults	21	0.84	ns
6- vs 8-year-olds	24	-2.30	0.030
6-year-olds vs adults	21	0.78	ns
8-year-olds vs adults	22	3.30	0.003

lier. Certainly overgeneralization of newly learned linguistic forms is recognized in language development (e.g., de Villiers and de Villiers, 1978; Slobin, 1973).

Another interesting result from this experiment was that, although not statistically significant, adults had the highest correlation coefficients for formant transitions, out of the four age groups, and the lowest correlation coefficients for fricative-noise spectra. Although it was anticipated that adults would base judgments of fricative identity for /f/ versus /θ/ greatly on formant transitions, rather than on fricative noises, the prediction had been that the relative weights would be the same as those of young children. Instead, the weighting strategies of adults favored formant transitions slightly more than those of the children. Perhaps this finding reveals the extent to which highly honed, mature perceptual strategies match the weight assigned to each acoustic property to the amount of information it provides about phonetic structure in the language.

### IV. EXPERIMENT 2: /s/ VERSUS /ʃ/ PERCEPTION

Experiment 1 showed that adults and children perform similarly when the phonetic judgment to be made is one for which we expect adults and children to use similar weighting strategies. This finding lends support for the conclusion that results of earlier studies using /s/-vowel and /ʃ/-vowel stimuli (showing that children weight formant transitions more and fricative-noise spectra less than adults) represent real perceptual differences across listener groups. However, it is difficult to make direct comparisons between experiment 1 (with /f/ and /θ/) and the earlier experiments (with /s/ and /ʃ/) because the arrangement of acoustic properties differed across the two kinds of studies. Specifically, the fricative noise was the property varied along a continuum in the earlier work, with formant transitions coming from natural syllables or from synthetic vocalic portions fashioned after natural syllables. As a result, formant transitions were always set dichotomously. In experiment 1, vocalic portions were synthesized with formant transitions varying along continua, and natural noises were used, and so were intrinsically dichotomous. The purpose of this experiment was to devise a similar experiment using /s/ and /ʃ/ so that results could be more readily compared.

Of course, realizing the utility in designing such an experiment is one thing, actually doing so is another. Extensive pilot testing showed that it was not possible to use stimuli with completely natural /s/ and /ʃ/ noises (and vocalic portions varying along a continuum) because to do so resulted in many listeners (particularly adults) responding according to the noise alone. Besides, it would not necessarily be desirable to use natural /s/ and /ʃ/ noises, with the large spectral difference that there is between them. Instead, a test more equivalent to the /f/ versus /ʃ/ experiment would be had if stimuli were designed with roughly the same acoustic difference between the /s/ and /ʃ/ noises as there was for the /f/ and /θ/ noises in experiment 1. That is, there were just a few dB difference between the major spectral peaks of those /f/ and /θ/ noises, and so stimuli in this experiment were designed with just a few dB difference between the major peaks of these /s/ and /ʃ/ noises. As in experiment 1, vocalic portions

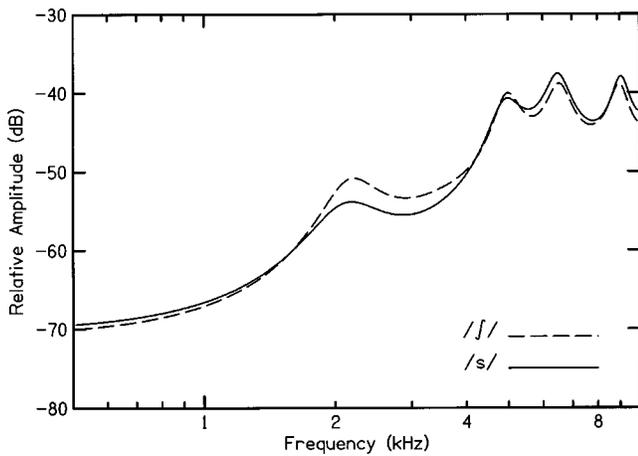


FIG. 4. Spectra of the noises used in experiment 2.

were synthesized so that the formant transitions at onset spanned a continuum: in this case, from those appropriate for a preceding /j/ to those appropriate for a preceding /s/.

## A. Method

### 1. Listeners

Listeners of the same age groups as those in experiment 1 participated in this experiment, and were required to meet the same criteria. Meeting these criteria were 12 4-year-olds, 13 6-year-olds, 19 8-year-olds, and 12 adults. However, four 4-year-olds and one 8-year-old were subsequently dismissed because they were unable to label reliably the best exemplars of each category. The mean ages (and ranges) of children were 4; 7 (4; 3 to 4; 9), 6; 3 (5; 9 to 6; 8), and 8; 0 (7; 10 to 8; 7). The mean age of adults was 31 years, with the range between 25 years and 39 years.

## 2. Equipment and materials

The same equipment was used in this experiment as in experiment 1.

## 3. Stimuli

Hybrid stimuli consisting of fricative noises, created from natural /s/ and /j/, and synthetic vocalic portions were used. The noises used in this experiment were not completely appropriate for /j/ and /s/. Instead, one of the noises used in this experiment had a spectrum more similar (than the other noise) to /j/, while the other noise had a spectrum more similar to /s/. These two noises were selected from a continuum created from natural /j/ and /s/ noises, and used in experiment 1 of Nittrouer and Miller (1997a). These noises were created by adjusting the amplitudes of natural /j/ and /s/ noises to desired levels (depending on how /j/-like or /s/-like the resulting noise should be), and then adding them together. In this case, the more /j/-like noise was created by combining the original /s/ and /j/ noises at equal amplitudes; the more /s/-like noise was created by combining the original /s/ and /j/ noises at a 2/1 (respectively) amplitude ratio. The spectra of these two resulting noises are shown in Fig. 4. The first moment of the /j/-like noise was 6.80 kHz, and the first moment of the /s/-like noise was 7.21. Both were 100 ms long.

Eighteen synthetic vocalic portions were created using Sensyn Laboratory Speech Synthesizer: nine each of /a/ and /u/. These vocalic portions were created at a 20-kHz sampling rate, to match the sampling rate of the noises. For all vocalic portions, duration was 260 ms, and  $f_0$  started at 120 Hz, falling throughout the portion to an offset frequency of 90 Hz. The vocalic portion was always 11 dB greater in amplitude than the noise portion. This amplitude difference

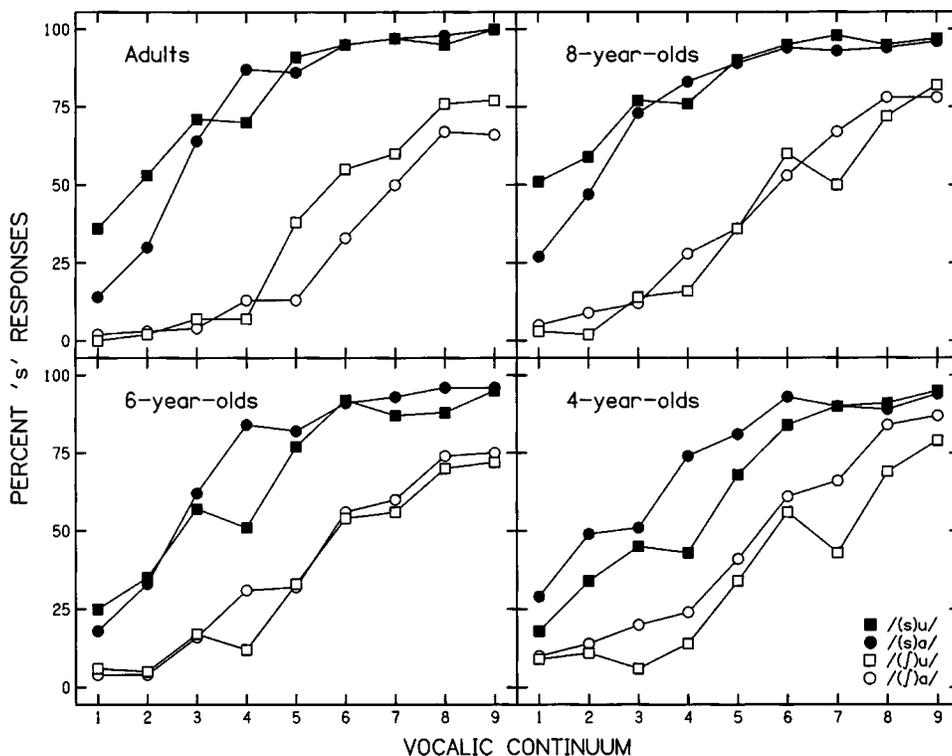


FIG. 5. Labeling functions for /s/-vowel and /j/-vowel stimuli in which formant transitions varied along continua and noise spectra varied dichotomously (experiment 2).

TABLE VI. Mean partial correlation coefficients for each group in experiment 2 (/ʃ/ versus /s/). s.d.'s are given in parentheses.

	4-year-olds	6-year-olds	8-year-olds	Adults
Fricative noise	0.370 (0.094)	0.498 (0.120)	0.582 (0.157)	0.604 (0.151)
Formant transitions	0.772 (0.065)	0.710 (0.079)	0.657 (0.139)	0.646 (0.141)

was the mean difference found for speech samples collected from speakers as part of an earlier analysis (Nittrouer, 1995; Nittrouer, Studdert-Kennedy, and Neely, 1996). For all /a/ portions,  $F1$  began at 450 Hz and rose to a steady-state value of 650 Hz over the first 50 ms.  $F2$  onset formed a nine-step continuum varying in 40-Hz steps from 1570 Hz (most /ʃ/-like) to 1250 Hz (most /s/-like).  $F2$  changed over the first 100 ms from its onset value to a steady-state value of 1130 Hz.  $F3$  onset varied in nine 58-Hz steps from a value of 2000 Hz (most /ʃ/-like) to 2464 Hz (most /s/-like). This formant changed over the first 100 ms from the onset value to a steady-state value of 2300 Hz.

For /u/ portions,  $F1$  was 250 Hz throughout.  $F2$  onset varied in 40-Hz steps from 1800 Hz (most /ʃ/-like) to 1480 Hz (most /s/-like).  $F2$  fell throughout the vocalic portion to an offset value of 850 Hz.  $F3$  onset varied in 40-Hz steps from 2200 Hz (most /ʃ/-like) to 2520 Hz (most /s/-like).  $F3$  fell through the first 130 ms to an offset frequency of 2100 Hz.

#### 4. Procedures

Procedures were the same for this experiment as for experiment 1.

#### B. Results

Figure 5 shows the labeling functions for all four groups, and generally gives the impression that the separation between functions, depending on fricative noise, increases with increasing age. Functions appear similar across age groups in terms of steepness. Table VI shows mean partial correlation coefficients for each age group, for both fricative noise and formant transitions. As with the partial correlation coefficients reported for experiments 1 and 3 of Nittrouer and Miller (1997b), we find a developmental increase in the weight assigned to the fricative noise, and a developmental decrease in the weight assigned to the formant transitions. The statistical analyses supported these observations: significant age effects were found both for the fricative noise,  $W(3,22)=7.55$ ,  $p=0.001$ , and for the for-

mant transitions,  $W(3,23)=3.60$ ,  $p=0.029$ . To investigate these results further, *post hoc t*-tests were conducted, and these results are shown in Table VII. The only clearly significant differences among age groups are found for 4-year-olds versus every other group. The comparison of 6-year-olds versus adults for noise was marginally significant. Nonetheless, there are linear developmental trends across age groups for both fricative-noise spectrum and formant transitions.

#### C. Discussion

The predictions for this experiment were that a developmental increase in the weight assigned to the fricative-noise spectrum and a developmental decrease in the weight assigned to formant transitions would be found. Both predictions were met.

#### V. GENERAL DISCUSSION

These experiments were designed to test specific predictions arising from the hypothesis that children's perceptual weighting strategies change as they learn what information they should be extracting from the speech signal in their native language in order to apprehend phonetic structure. Specifically, the DWS proposes that young children initially focus their perceptual attention on general movement of the vocal tract, as conveyed by patterns of changing formant frequencies. This perceptual strategy would meet the needs of novice language users who are just learning the fundamentals of how to move their own vocal tracts for the purposes of communication: Early vocalizations do not show the detail, particularly at syllable margins, that mature speech demonstrates (e.g., Oller, 1986). The perceptual strategy suggested here is consistent also with the suggestion that children's word representations are not as precisely phonetic as those of adults, resulting in more globally organized lexicons and poorer retention of linguistic material in working memory (e.g., Charles-Luce and Luce, 1990, 1995; Nittrouer and Miller, 1999). As children gain experience with their native language, it is proposed, perceptual attention (i.e., weight) gradually shifts to take advantage of the properties of the acoustic signal that are most informative regarding phonetic structure. At least one aspect of this developmental shift, it is suggested, entails progressively greater weight being assigned to acoustic properties that convey information about details of production, such as precise shapes of consonantal constrictions. Consequently, this maturing perceptual

TABLE VII. Results of *post hoc t*-tests for experiment 2.

	Noise			Formant transitions		
	<i>df</i>	<i>t</i>	<i>p</i>	<i>df</i>	<i>t</i>	<i>p</i>
4- vs 6-year-olds	15	-2.63	0.019	14	1.88	0.081
4- vs 8-year-olds	18	-4.13	<0.001	21	2.82	0.010
4- vs adults	16	-4.17	<0.001	16	2.65	0.017
6- vs 8-year-olds	28	-1.68	0.104	27	1.36	ns
6-year-olds vs adults	21	-1.93	0.067	16	1.39	ns
8-year-olds vs adults	24	-0.39	ns	23	0.21	ns

trend not only helps the child access the phonetic structure of the message, it also helps the child refine her own productions.

In general, predictions for this study were met. For the /f/-vowel and /θ/-vowel stimuli, listeners of all ages showed similar weighting strategies for both the fricative-noise spectra and formant transitions. This result was predicted based on the finding of Harris (1958) showing that adults heavily weight formant transitions in making this decision. For the /s/-vowel and /ʃ/-vowel stimuli in which the formant transitions formed the acoustic continua and noise spectra varied dichotomously, weighting strategies resembled those of earlier studies in which the stimuli were designed differently: With increasing age came an increase in the weight assigned to the fricative-noise spectra and a decrease in the weight assigned to the formant transitions. Consequently, the conclusion seems warranted that language perceivers learn what information is usually available in the signal, and look to the relevant signal properties for that information.

In conclusion, the experiments described here provide general support for the hypothesis that children's perceptual strategies for speech are modified as they gain experience with a native language. Specifically, children learn what information in the signal must be extracted in order to apprehend phonetic structure in the language they are acquiring. Thus, learning to perceive speech in a mature manner can be viewed within a framework of general perceptual learning.

#### ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. The author gratefully acknowledges the assistance of Marnie E. Miller and Sandy Estee in the collection and analysis of data.

<sup>1</sup>Throughout this manuscript, precise statistical results will be given when  $p < 0.10$ . Otherwise, results will be reported simply as nonsignificant.

- 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.
- deVilliers, J. G., and deVilliers, P. A. (1978). *Language Acquisition* (Harvard University, Cambridge, MA).
- Gibson, E. J. (1991a). "How perception really develops: A view from outside the network," in *An Odyssey in Learning and Perception* (MIT Press, Cambridge, MA), pp. 411–491. [Reprinted from *Basic Processes in Reading: Perception and Comprehension*, edited by D. Laberge and S. J. Samuels (Erlbaum, Hillsdale, NJ, 1977), pp. 155–173.]
- Gibson, E. J. (1991b). *An Odyssey in Learning and Perception* (MIT Press, Cambridge, MA).
- Goldman, R., and Fristoe, M. (1986). *Goldman Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).
- Harris, K. S. (1958). "Cues for the discrimination of American English fricatives in spoken syllables," *Lang. Speech* **1**, 1–7.
- Jastak, S., and Wilkinson, G. S. (1984). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).
- Nittrouer, S. (1992). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," *J. Phonetics* **20**, 1–32.
- Nittrouer, S. (1995). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *J. Acoust. Soc. Am.* **97**, 520–530.
- Nittrouer, S., and Miller, M. E. (1997a). "Developmental weighting shifts for noise components of fricative-vowel syllables," *J. Acoust. Soc. Am.* **102**, 572–580.
- Nittrouer, S., and Miller, M. E. (1997b). "Predicting developmental shifts in perceptual weighting schemes," *J. Acoust. Soc. Am.* **101**, 2253–2266.
- Nittrouer, S., and Miller, M. E. (1999). "The development of phonemic coding strategies for serial recall," *Appl. Psycholing.* **20**, 563–588.
- 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., Studdert-Kennedy, M., and Neely, S. T. (1996). "How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables," *J. Speech Hear. Res.* **39**, 379–389.
- Oller, D. K. (1986). "Metaphonology and infant vocalization," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton, New York), pp. 21–36.
- Slobin, D. I. (1973). "Cognitive prerequisites for the development of grammar," in *Studies of Child Language Development*, edited by C. A. Ferguson and D. I. Slobin (Holt, Rinehart, and Winston, New York), pp. 175–208.
- Strange, W. (editor). (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Baltimore).