INDIVIDUAL DIFFERENCES IN, AND A COMPARISON OF, IDENTIFICATION AND SIMILARITY JUDGMENTS OF CONTEXT-CONDITIONED /S/ AND /ʃ/ PHONEMES

Margaret F. Cheesman[1] and **Dianne J. Van Tasell**[2]
[1]Department of Communicative Disorders, The University of Western Ontario, London, ON Canada N6G 1H1
[2]Department of Communication Disorders, University of Minnesota, Minneapolis, MN USA

ABSTRACT

Speech segments strongly influence the perception of adjacent speech segments. Such context effects provide interesting evidence of the interaction of acoustic information in the perceptual system. Studies that have dealt with such phenomena have focused on the effect of context on the label assigned to a phoneme, so that little is known about the **within-class** perception of context-conditioned phonemes. In the present study, the effect of vowel context on the perception of synthetic /s/- and /ʃ/-like frication noises was examined in two experiments. A two-alternative forced-choice identification task confirmed that identification of the fricative in a set of consonant-vowel syllables was influenced by the vowel context. In a second experiment, the perceptual similarity of pairs of fricatives whose identity was influenced by the vowel was estimated in a triadic comparison task. INDSCAL analyses provided three dimensions that could account for 80.9% of the observed variance. However, individuals differed greatly on the contribution of each dimension to their similarity judgments. For some listeners, judgments of perceptual similarity were strongly related to their identification judgments. For other listeners, similarity of the fricatives was related to the physical differences between the fricatives, regardless of whether the fricatives had been identified as the same consonant or not. These results indicate that listeners differ in their abilities to perceive differences between phonemes that have been assigned the same label.

SOMMAIRE

Des segments de discours influencent fortement la perception des segments de discours adjacents. De tels effects de contexte produisent une intéressante mise en valeur de l'interaction de l'information acoustique dans le système perceptif. Les études portant sur un tel phénomène se sont concentrées sur l'effet du contexte sur l'étiquette collée à un phonème, de telle sorte que l'on en sait peu à propos de la perception des phonémes conditionnés par le contexte à l'intérieur de la classe. Dans la présente étude, l'effet du contexte vocalique sur la perception des sons fricatifs tels que /s/ et /ʃ/ synthétiques a fait l'objet de deux expériences. Le travail d'identification d'un choix binaire a confirmé que l'identification de la fricative dans un éventail de syllabes consonne-voyelle était influencé par le contexte vocalique. D'après une seconde expérience, la similarité de perception des paires de fricatives dont l'identité était influencée par la voyelle a été estimée dans un projet de comparaison ternaire. Les analyses de l'INDSCAL ont mis en évidence trois dimensions qui pouvaient comptabiliser 80,9% de la variance observée. Cependant, les individus ont différé énormément pour la contribution de chaque dimension à leurs jugements de similarité. Pour certains auditeurs, les jugements de similarité de perception étaient fortement liés à leurs jugements d'identification. Pour d'autres auditeurs, la similarité des fricatives était liée aux differérences physiques entre les fricatives, qu'elles aient été identifiées comme la même consonne ou non. Ces résultats indiquent que les auditeurs diffèrent dans leurs capacités à percevoir des différences entre les phonèmes auxquels on a assigné la même étiquette.

The acoustic information that characterizes a phoneme varies with the context of the other phonemes surrounding it. For example, the phoneme /d/, when produced at the onset of a syllable, contains a brief noise burst and periodic energy. When the following vowel has a high-frequency second formant (F_2), the F_2 transition rises at the onset of the syllable (e.g., in /di/). When the vowel has a low F_2 (e.g., in /du/), the initial F_2 transition falls. The onset frequency and extent of the transitions can serve as reliable cues to the identity of the consonant (Liberman, et al., 1967). This dependence of the acoustic characteristics of a phoneme on the context in which it occurs is called context-conditioned variability.

As a consequence of the variability in the acoustic content of speech segments, it is often impossible to predict the phonemic identity of a particular acoustic pattern without also knowing the acoustic information (context) that precedes or follows the segment. Because a given phonemic distinction may be cued by several types of acoustic information distributed in time, two acoustic cues may compensate for one another; a change in one cue may be "cancelled" by a change in the other, thereby maintaining a constant phonetic percept.

Over a limited range of values, such cancellation effects have been demonstrated with a number of speech contrasts. For example, the "say" - "stay" distinction may be cued both by the duration of silence following the fricative /s/ and by the frequency of the first formant at the onset of voicing. When either of these cues is ambiguous, the other will cue the presence of /t/. However, a lengthening of the silent interval in a word that is perceived as "say," which normally would change what a listener hears to "stay," can be compensated by increasing the onset frequency of the first formant (F_1) so that the perception of "say" persists (Best, et al., 1981). Likewise, for a limited range of F_1 values and silent intervals, a higher F_1 will not produce the "stay" percept if the silent duration is shortened.

Phonetic context effects have been studied extensively in identification tasks in which phonemes are labelled in a forced-choice task (Repp, 1982, provides a review of these studies). With such tasks, listeners must select from a limited set of phoneme labels for their identification responses, even if the labels are not particularly appropriate to the phonemes. Because of the limited set of responses permitted in the identification task, listeners may adopt response strategies that assign the same labels to phonemes that are perceptually noticeably dissimilar. Little is known regarding the degree of perceptual similarity (or dissimilarity) among phonemes that have been assigned the same (or different) labels. The question arises whether the effect of systematically changing the phonemic context along an acoustic continuum is to create a perceptual continuum, which is then artificially partitioned because of the nature of the forced-choice task used to study it. Alternatively, the underlying percept may indeed be categorical and phonemes labelled as belonging to one phonemic category may indeed be perceptually more similar than stimuli that lie in opposite sides of a category boundary. If this is indeed so, it would have interesting ramifications for theories of speech perception because stimuli on either side of a phonemic boundary could clearly be acoustically more similar than within-category stimuli.

Support for the latter hypothesis, that context-conditioned phonemes within a category are perceptually more similar than across category phonemes, comes from studies of the discriminability of context-dependent phonemes (Bailey, et al., 1977; Oller, et al., 1991; Repp, 1981). In these studies, discriminability of phonemes is usually better for phoneme pairs that cross category boundaries than for those that lie within a category. Such results lend support to the notion that within-category stimuli are more similar than between-category stimuli. However, Repp (1981) identified two subgroups of listeners that performed differently in a fricative discrimination task. One group demonstrated the good cross-boundary and poor within-category discrimination reported in earlier studies. The other, smaller, group of listeners demonstrated good discrimination of within-category stimuli. Repp postulated that this group of listeners who did not respond to the fricatives in a categorical manner were able to listen to these stimuli as auditory, rather than phonetic objects.

In the present paper, the context-conditioning of phonemes was studied using both a traditional forced-choice identification task and a triadic comparison procedure that yielded a direct measure of perceptual similarity. The effect of vowel context on the perception of /s/ and /ʃ/-like frication noises was studied.

1. PERCEPTION OF /S/ AND /ʃ/

The perception of context-conditioned /s/ and /ʃ/ segments has been studied by a number of investigators (Abbs & Minifie, 1969; Kunisaki & Fujisaki, 1977; Mann & Repp, 1980; Mann, et al., 1985; Nittrouer & Studdert-Kennedy, 1987; Repp, 1981; Whalen, 1981; Yeni-Komshian & Soli, 1981). Kunisaki and Fujisaki (1977) used synthetic syllables produced by combining a frication-noise continuum (representative of /s/- and /ʃ/-like frication) with /a, e, o, u/ vowels. Japanese listeners labelled the consonants as either /s/ or /ʃ/. The boundary between /s/- and /ʃ/-labelled stimuli was at different fricative frequencies for different vowel contexts. The boundary shifted to lower fricative frequencies before rounded vowels, which contain lower second and third formants than do unrounded vowels. These results have been replicated with English-speaking adults (Mann & Repp, 1980; Repp, 1981; Whalen, 1981), with children as young as three years of age (Nittrouer & Studdert-Kennedy, 1987), and with vowel contexts that do not occur in the listeners' native language (Whalen, 1981).

The context-dependent perception of /s/ and /ʃ/ segments provides an ideal stimulus set to investigate the relationship between identification and similarity judgments. Not only have the acoustic variables that influence the perception been extensively studied, but also individual differences in the ability to discriminate fricatives embedded in different vowel contexts have been described (Repp, 1981). Stimuli can be created that vary along the two independent acoustic dimensions of frication frequency and vowel quality; within a range of each of these acoustic dimensions, identification of the fricative will be dependent on both of these acoustic dimensions. The phoneme categories will therefore be bounded by stimuli that contain acoustically-identical fricatives on one or the other dimension. A comparison of the perceptual similarity of stimuli that span the boundary, yet contain identical frication noises, versus stimuli that lie to one side of the boundary therefore can be made.

2. TRIADIC COMPARISONS

Triadic comparison procedures have been used to estimate the perceptual similarity of musical intervals (Levelt, et al., 1966), of timbres (Plomp, 1970), and of vowels (Beck, et al., 1988; Pols, 1970; Pols, et al., 1969; Rakerd & Verbrugge, 1985). In this procedure, sets of three stimuli (triads) are compared by the subject, who must decide which two stimuli are **most similar** and which two are **most dissimilar**. This comparison is made for all possible triads of the stimulus set. The number of times that each pair of stimuli is selected as more similar than other pairs yields an index of the perceptual similarity of stimulus pairs.

An advantage of the triadic comparison task is that, unlike verbal scaling procedures, triadic comparisons do not force subjects to use verbal categories in order to obtain a similarity metric. Rather, the task permits the use of a simple instruction set that allows subjects to set their own criteria for similarity (Levelt, et al., 1966).

A disadvantage of the procedure is the rapid increase in the number of trials that is needed as the number of stimuli is increased. In order that every possible pair of stimuli is compared with all other pairs of stimuli, all possible stimulus triads must be included in the design. The total number of triads that can be created from N stimuli is N(N-1)(N-2)/6. Thus, for 45 stimuli, 14,190 triads can be formed; for 12 stimuli, there are 220 triads.

3. PURPOSE

In the present study, a set of synthetic fricative-vowel stimuli was constructed to demonstrate the effect of vowel context on fricative perception. A two-dimensional continuum was constructed, with frication-noise frequency comprising one dimension and vowel context the second dimension. These two dimensions were combined factorially to construct the stimulus set. In Experiment 1, identification data were obtained with this stimulus set to confirm that the vowel F_2 and F_3 frequencies did systematically affect the identification of the fricatives. In the second experiment, the perceptual similarity among the fricatives in a subset of 12 of the synthetic fricative-vowel syllables was estimated from the results of a triadic comparison task. This second stage of data collection provided perceptual similarity judgments and allowed a comparison to be made among: (a) the perceptual space occupied by the synthetic stimuli, (b) the labels assigned to these syllables in Experiment 1, and (c) their acoustic characteristics.

4. EXPERIMENT 1: IDENTIFICATION

4.1 Method

Stimuli. A synthetic, frication-noise continuum was paired with a synthetic /i/ - /u/ continuum to form a set of 45 consonant-vowel (CV) syllables (9 noises x 5 vowels). The fricative and vowel sounds were created separately and concatenated. All synthesis was performed with 12-bit resolution at a 14-kHz sample rate.

The synthetic fricatives were 150-ms noises; this duration is slightly longer than fricative durations in natural sentence production (Klatt, 1974), and is slightly shorter than the 175-ms durations for these fricatives produced in isolated CV syllables (Behrens & Blumstein, 1988).

The fricatives were synthesized with ILS (Interactive Laboratory System, Version 4.0) software. A wideband, flat-spectrum noise was created digitally. The noise waveform had a linear rise time of 75 ms from silence to full amplitude and a 30-ms linear fall from full to half-amplitude. This frozen noise was digitally filtered to form a continuum of nine noises (C_1 - C_9) in which the low-frequency cut-off increased from 1800 to 4000 Hz and the high-frequency cut-off increased from 3950 to 4950 Hz in equally-spaced steps. The filters were elliptical, third-order filters, which provided 40-dB attenuation in the stopbands.

The five vowels (V_1-V_5) were synthesized with an implementation of the Klatt cascade formant synthesizer (Jamieson, et al., 1989; Klatt, 1980). The vowels were 300 ms long. F_1 was fixed at 250 Hz. F_2 and F_3 contained transitions that increased in both duration and frequency from V_1 (/u/) to V_5 (/i/). Further details of the formant transitions are provided in Table 1.

Instrumentation. Stimulus generation and data collection were controlled with an IBM/AT computer and a DT2801A I/O board, followed by a Hewlett-Packard passive attenuator, a Kemo VB/25 programmable filter, and a Charybdis programmable attenuator (Model D). Stimuli were output at a digital-to-analog conversion rate of 14 kHz and low-pass filtered at 6 kHz with a rejection rate of 96 dB/octave.

Table 1
F_1 and F_2 Synthesis Parameters for the Stimuli Used in Experiment I

	F_2 transition			F_3
Vowel	Start (Hz)	Finish (Hz)	Duration (ms)	Freq (Hz)
V_1	1450	850	300	2200
V_2	1650	1200	240	2400
V_3	1850	1550	180	2600
V_4	2050	1900	60	2800
V_5	2250	2250	0	3000

Note: Entries are the centre frequency and the duration of the F_2 transitions, and the F_3 (stationary) centre frequency. Transition durations were selected to maintain perceptual continuity of the fricative with the vowel. For more /u/-like vowels, longer transitions were required.

The synthetic stimuli were presented monaurally via a TDH-49 earphone in an MX41/AR cushion while the subject was seated in a double-walled IAC sound-attenuating test booth. Stimuli were presented at a level at which the continuous, steady-state portion of one of the synthetic vowels, V_3, measured 65 dB SPL at the earphone in an NBS-9A coupler. Instructions were presented to the subject using a colour monitor, and the subject responded by pressing "keys" on a template placed over a Koala digitizing pad interfaced to the computer.

Subjects. Twelve adults, aged 20 to 41 years, served as subjects (S1 - S12). All had some phonetic training. All listeners had pure-tone thresholds better than 20 dB HL (ANSI, 1989) at 250, 500, 1000, 2000, 4000, and 6000 Hz in the test ear.

Procedure. Identification data were collected in a two-alternative forced-choice task. Subjects were instructed to indicate, after each stimulus presentation, whether the consonant sounded more like an /s/ or /ʃ/ by pressing one of two buttons labelled "ss" or "sh". Twenty identification judgments were made for each synthetic stimulus. Stimuli were presented in 20 blocks of 45 stimuli each; within each block, the order of stimulus presentation was randomized without replacement.

4.2 Results

Individual subjects' responses are displayed in Figure 1. The percentage /ʃ/ responses made for each stimulus is plotted as a function of the frication-noise portion (C_1-C_9) of the stimulus. The parameter is the vowel (V_1-V_5) with which each fricative was paired. Each point in these identification functions is based on 20 identification responses.

For all subjects and in each vowel context, subjects responded /ʃ/ for lower-frequency fricatives and /s/ for higher-frequency fricatives. The effect of the vowel on the identification of the

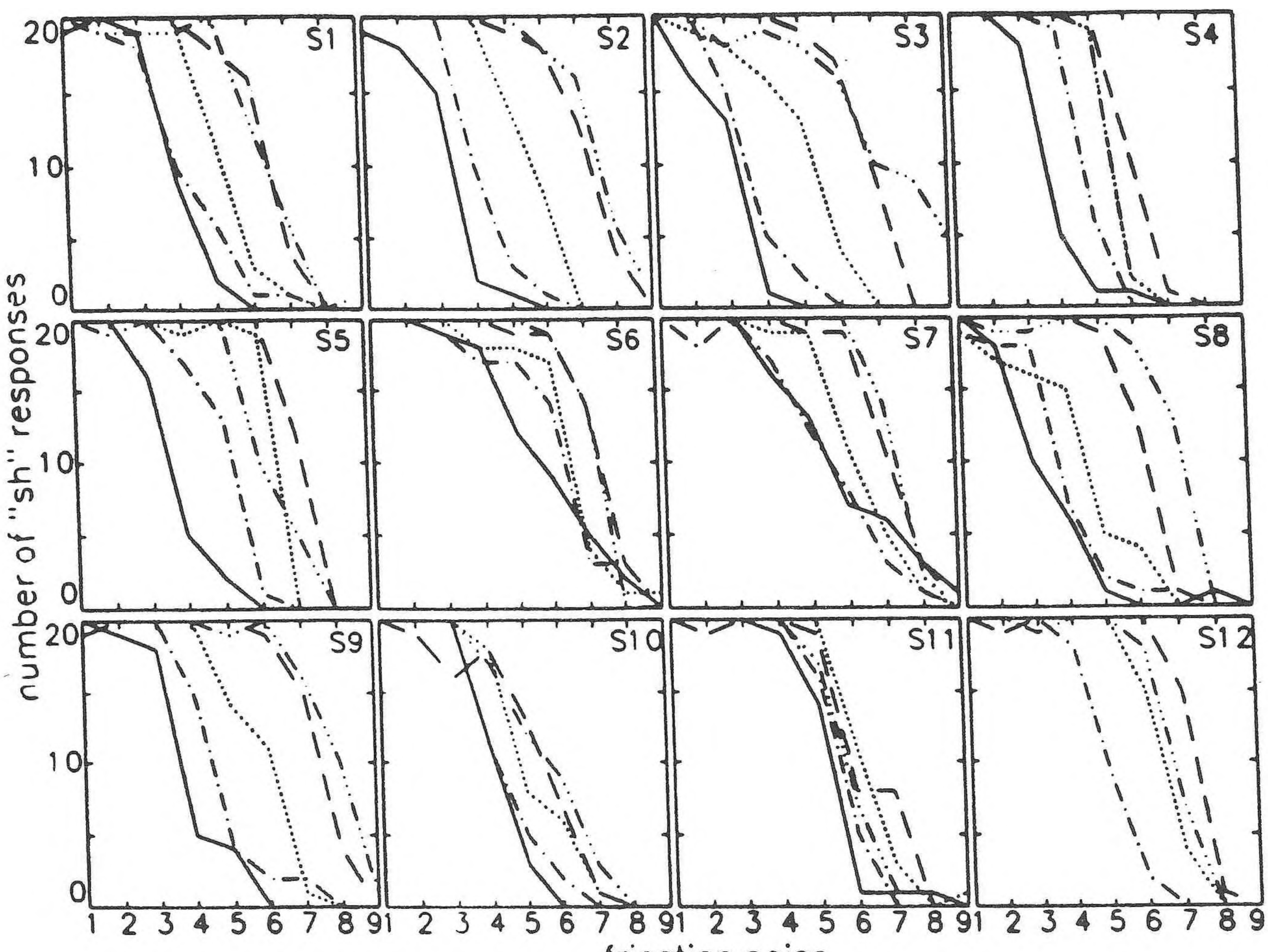

Figure 1. Identification responses for each of the 12 listeners as a function of frication noise and vowel quality obtained in Experiment 1. Each point represents the responses [on 20 stimulus presentations]. Solid line (V_1-/u/), dash-dot line (V_2), dotted line (V_3), dashed line (V_4), dash-dot-dot line (V_5-/i/).

fricative is shown by the shift of the identification functions to the right as the F_2 and F_3 frequencies increased from V_1 (/u/) to V_5 (/i/). The category boundary, as defined by the point at which there were 50% /s/ and /ʃ/ responses, shifted to higher-frequency noises as the vowel context changed from V_1 to V_5. Over the range of fricatives where this shift occurred, the same frication noise was labelled differently before different vowels. For example, Subject 1 labelled C_5 as /ʃ/ only 10% of the time when it was followed by V_1, yet identified the same frication noise as /ʃ/ 100% of the time when it was followed by V_5.

There were large individual differences in the fricative identification functions both in the location of the /ʃ/ - /s/ identification boundary and in the magnitude of the vowel context effect. For some subjects, the effect of increasing the formant frequencies did not extend across all vowels. Four subjects (S1, S3, S6 and S10) had shifts in the fricative identification boundaries for V_1-V_4 only. Four others (S4, S5, S11 and S12) showed an effect of the vowel context that was non-monotonic; the increase in second and third formant frequencies from V_4 to V_5 shifted the /ʃ/ - /s/ identification boundary back to lower-frequency fricatives. Such non-monotonic changes in identification functions for these context-conditioned phonemes have been observed by Mann and Liberman (1983) and may be influenced by perceptual "magnet" effects (Kuhl, 1991).

The effect of the vowel context and frication noise on the identification of the fricatives was examined using a repeated-measures analysis of variance. A significant effect of vowel ($F=39.0$, $df=4,44$, $p<.001$) and frication noise ($F=271.6$, $df=8,88$, $p<.001$) was obtained, as well as a significant interaction between these two factors ($F=22.1$, $df=35,352$, $p<.001$).

In summary, the effect of vowel context on the perception of fricatives, as reported by Mann and Repp (1980), Repp (1981), and Whalen (1981), was replicated with a set of synthetic fricative-vowel syllables. For all subjects, some fricatives were labelled as /ʃ/ when followed by vowels that had high F_2 and F_3 values and as /s/ when followed by vowels with lower formant frequencies, although there were substantial individual differences in the extent of the vowel influence.

5. EXPERIMENT 2: PERCEPTUAL SIMILARITY JUDGMENTS

5.1 Method

Stimuli. The stimulus set was selected from the set of synthetic syllables used in Experiment 1. Syllables that contained the five fricatives (C_{1-3} and C_{8-9}) whose identi-fication was not strongly influenced by the vowel context were not included. Two of the vowels were eliminated from the stimulus set in order to reduce the number of stimuli to be used in the triadic comparison task. Post-hoc analyses of the data from Experiment 1 indicated that V_4 and V_5 did not differ with respect to the number of /s/ and /ʃ/ responses that each elicited and, for some subjects, V_5 created a non-monotonic shift in the identification boundaries (cf. Figure 1). V_5 was therefore eliminated in favour of V_4 and V_2 was eliminated arbitrarily, to reduce further the stimulus set. The 3 vowels ($V_{1,3,4}$)that had transition durations of 300, 180, and 60 ms and a systematic effect on the perception of the frication noise were included. After eliminating these stimuli, a set of twelve syllables remained for the triadic comparison task (Roskam, 1979); those stimuli produced by combining C_{4-7} and $V_{1,3,4}$. The stimulus set was sufficiently small that a completely-balanced triadic comparison procedure could be completed in a single experimental session of reasonable duration.

The twelve syllables were combined to form all possible sets of three different syllables, or 220 triads. For the purpose of analysis, each triad can be treated as three pairs of stimuli, from which the subjects selected the most similar pair and the most dissimilar pair. Within the set of 220 triads, each stimulus occurred 110 times, and each pair of stimuli occurred 10 times.

Procedure. Instrumentation and subjects were as described for Experiment 1. Subjects participated in Experiment 2 during a second test session.

Each trial of the triadic comparison task consisted of an initial stimulus presentation sequence in which each of the three stimuli to be compared (Stimuli A, B and C) was presented once. A section of a video monitor corresponding to each stimulus and labelled "A", "B" or "C" was flashed in reverse video as each stimulus was presented. A 500-ms interval followed each stimulus presentation.

After the initial presentation sequence, subjects could listen repeatedly to any of the stimuli by pressing the labelled buttons on a touch-sensitive digitizing pad. Subjects were instructed to indicate which pair of consonants was most similar and which pair was most dissimilar by pressing the button on the digitizing pad that was labelled with the chosen stimulus pair. Stimuli could be repeated as many times as required in order to make a decision.

Ten practice trials were completed prior to starting the 220 trials. The order of the triads and the order of the stimuli within the triads were random.

5.2 Results

In order to generate a similarity matrix that included the entire stimulus set, the three possible pairwise combinations of stimuli within each triad were first rank-ordered with respect to similarity. Within each triad, a score of 2 points was assigned to the pair selected as the most similar (it was judged to be more similar than the two other pairs). Zero points were assigned to the pair selected as most dissimilar. One point was assigned to the pair that was not selected (it was judged to be more similar than one of the other pairs and less similar than the other). The points assigned to a pair were then summed over the 10 occurrences of that pair within the 220 triads, to yield a composite score indicating the number of comparison pairs that were judged to be less similar than that pair — that is, the **similarity value** of the pair.[1] The maximum similarity value that can be obtained with this procedure is 20 (i.e., when a pair was selected as the most similar every time that it was presented, regardless of the other stimulus in the triad) and the minimum is 0 (when a pair was selected as most dissimilar on every trial). Completely random responding yields an expected similarity value of 10 for each pair.

The obtained similarity values indicated that subjects were not responding randomly: the full range of possible similarity values (0-20 across different stimulus pairs) was obtained for several subjects and the smallest range of similarity values was from 1-18 (for subject 9). The summed similarity matrix for all 12 subjects is presented in Table 2. Each entry in the matrix is the similarity value for a pair of stimuli — that is, the total number of times that a stimulus pair was chosen as more similar than other pairs — summed across all 12 subjects.

Table 2

Summed similarity matrix for 12 subjects in the triadic comparison task. Each entry indicates the total number of times that each pair of stimuli was selected as more similar than other pairs. The maximum attainable value was 240; the minimum was 0.

		C4			C5			C6			C7		
		V1	V3	V4	V1	V3	V4	V1	V3	V4	V1	V3	V4
C4	V1		155	144	164	159	138	92	96	102	73	49	57
	V3			213	93	209	169	45	109	123	36	60	65
	V4				83	173	210	49	72	135	37	38	77
C5	V1					119	111	194	127	105	141	115	102
	V3						189	65	160	157	51	71	79
	V4							66	103	196	51	48	104
C6	V1								149	105	213	190	149
	V3									167	132	185	151
	V4									79	79	97	172
C7	V1											201	157
	V3												190
	V4												

[1] An alternative, ordinal-level interpretation of the similarity value is that the value comprises the sum of the ranks assigned to each pair in all occurrences in different triads. The most similar pair was given a rank of 2, the pair that was not selected was given a rank of 1 and the least similar pair was given a rank of 0.

Some general observations can be made concerning the patterns of similarity values observed. First, similarity values were generally smallest when the physical differences between two fricatives were the greatest (C_4 vs. C_7). This is indicated by the relatively small values contained in the upper right corner of the summed similarity matrix. Second, similarity values were generally largest when the fricatives were either identical or differed by just one step. This result is revealed in the large entries occurring near the main diagonal of Table 2. Thus, the physical distance between the fricative portions of two stimuli was inversely related to the subject's similarity judgments, regardless of other stimulus parameters, including vowel spectrum and consonant identity.

Fricative labels and perceptual similarity. To examine the relation between the labelling and similarity judgments, each stimulus was first classified as /s/ or /ʃ/ on the basis of the label given to it more than 50% of the time in Experiment 1. Table 3 presents mean similarity values for two groups of stimuli: (1) pairs of syllables that were labelled as the same fricative (either both identified as /s/ or both as /ʃ/) and (2) pairs of syllables that were labelled as different fricatives (one /s/ and the other /ʃ/).

Table 3

Comparison of the perceptual similarity values for pairs of syllables that had been assigned the same vs. different fricative labels in Experiment 1. Entries are: (1) the number of stimulus pairs included in each calculation, (2) the mean similarity value and (3) the standard deviation of the similarity values.

Subject	Same Label			Different Label		
	N	Mean	SD	N	Mean	SD
S1	31	12.55	5.2	35	7.74	4.4
S2	30	14.73	3.0	36	6.06	3.4
S3	31	13.26	4.4	35	7.11	4.3
S4	31	12.88	4.9	35	7.46	4.4
S5	31	11.77	5.5	35	8.42	5.2
S6	39	11.08	5.1	27	8.44	5.8
S7	39	11.36	4.6	27	8.04	4.0
S8	34	11.76	5.1	32	8.13	4.3
S9	31	12.10	4.4	35	8.14	3.8
S10	34	12.71	4.3	32	7.13	4.2
S11	31	13.48	3.9	35	6.91	4.6
S12	31	12.71	5.1	35	7.60	4.5
Mean		32.8	12.53*		33.3	7.60*
SD		3.17	1.01		3.17	0.71

* A matched-pairs t-test between the overall means indicated a significant difference ($t(11)=10.047$, $p<.001$).

Although the intersubject variability of these values was relatively large, in general judgments of perceptual similarity were related to the perceptual identity of the consonants, in that consonants that were labelled the same were judged to be perceptually more similar than consonants that were labelled differently.

The comparison made in Table 3 included two types of stimulus pairs: (1) stimuli in which the fricatives were physically different, and (2) stimuli in which the fricative portion was fixed, but the vowel portion differed. Because the interest in this study was to examine the perceptual context effect and because perceptual similarity judgments were related, in part, to the physical similarity of the consonants, it was important to isolate the relationship between the fricative label and the perceptual similarity, particularly in those cases where the fricatives were identical (so that it was the vowels that influenced the label). To achieve this isolation, the perceptual similarity analysis described above was repeated, restricting the data set to the cases where both stimuli in a pair contained the same frication noise.

Table 4 presents the results of this analysis, based on the 12 pairs that contained the same fricative noises in both stimuli of the pair (i.e., the 12 stimulus pairs that lie nearest the main diagonal in Table 2). It can be seen that mean similarity values for these stimuli are higher than those in the inclusive analysis in Table 3, indicating that, as expected, stimuli were judged to be more similar when the fricative portions of the stimuli within the pair were more similar, physically. Again, it can be seen that, on average, fricatives that were given the same label had higher similarity values than those that were given different labels, reflecting the fact that, for some subjects, similarity was judged on the basis of whether the fricatives belonged to the same phoneme class.

Although the pattern described above holds for the summarized results, there were large individual differences in response patterns. Tables 5 and 6 compare the similarity matrices for two subjects (S2 and S6) who responded quite differently. S6 obtained high similarity values for all stimulus pairs that were included in this analysis. For this subject, similarity judgments apparently were based on the physical differences among the fricatives, rather than on how the signals were labelled. A similar but less extreme dependence on physical differences was shown by subject 11, who also had very high similarity values for all stimulus pairs included in this analysis.[2]

Table 4

Comparison of the perceptual similarity values for pairs of syllables that had been assigned the same vs. different fricative labels in Experiment 1. Only pairs of syllables in which the frication components were physically identical (and the vowels differed) have been included. Entries are (1) the number of stimulus pairs included in each calculation, (2) the mean similarity value, and (3) the standard deviation of the similarity values.

Subject	Same Label			Different Label		
	N	Mean	SD	N	Mean	SD
S1	6	15.33	1.5	6	12.33	3.7
S2	4	14.75	3.6	8	7.13	2.2
S3	6	13.66	2.4	6	8.33	1.4
S4	6	17.00	2.5	6	12.17	3.9
S5	4	17.25	1.3	8	12.25	4.2
S6	8	17.50	1.4	4	18.75	1.9
S7	8	12.88	4.8	4	9.75	3.9
S8	6	11.33	4.8	6	8.00	3.8
S9	4	14.75	2.2	8	9.88	3.4
S10	10	15.70	3.3	2	11.50	2.1
S11	10	17.50	1.3	2	16.00	1.4
S12	4	17.00	1.4	8	10.00	5.1
Mean		6.3	15.39*		5.67	11.34*
SD		2.23	2.01		2.23	3.35

* A matched-pairs t-test between the overall means indicated a significant difference ($t(11)=5.888$, $p<.001$).

For the remainder of the subjects, judgments of perceptual similarity were related, at least partially, to how the fricatives were labelled, in the identification task of Experiment 1. Subject 2 was the most extreme of these subjects, showing a mean similarity rating of just 7.13 when the fricatives of the pair were labelled differently, vs. 14.75 when the fricatives were labelled the same. Other subjects fell between S2 and S6 in terms of the relative dependence of their similarity judgments on the physical differences between the signals and on the labels assigned to stimuli.

<u>Multidimensional scaling of similarity data</u>. In order to study further the dimensional structure of these perceptual similarity judgments, the data were subjected to nonmetric multidimensional scaling. SPSS-X ALSCAL (v.3.1) routines produced a three-dimensional solution which accounted for 80.9% of the variance. Adding a fourth dimension contributed little to the goodness of fit, accounting for just 1.0% more of the variance.

The three-dimensional solution is illustrated in Figure 2. Dimension 1 corresponds to the physical (acoustical) differences among fricatives and accounts for 49.5% of the variance. Dimension 2 accounts for 24.7% of the explained variance and appears to correspond to the perceptual identity of the fricative.

[2]Subjects 6, 7 and 11 reported that they were making similarity judgments based on the "pitch" of the fricatives. This is consistent with comments of Repp's (1981) non-categorical subjects. The data from subjects 6 and 11 suggest that they were making the judgment independently of the vowel context.

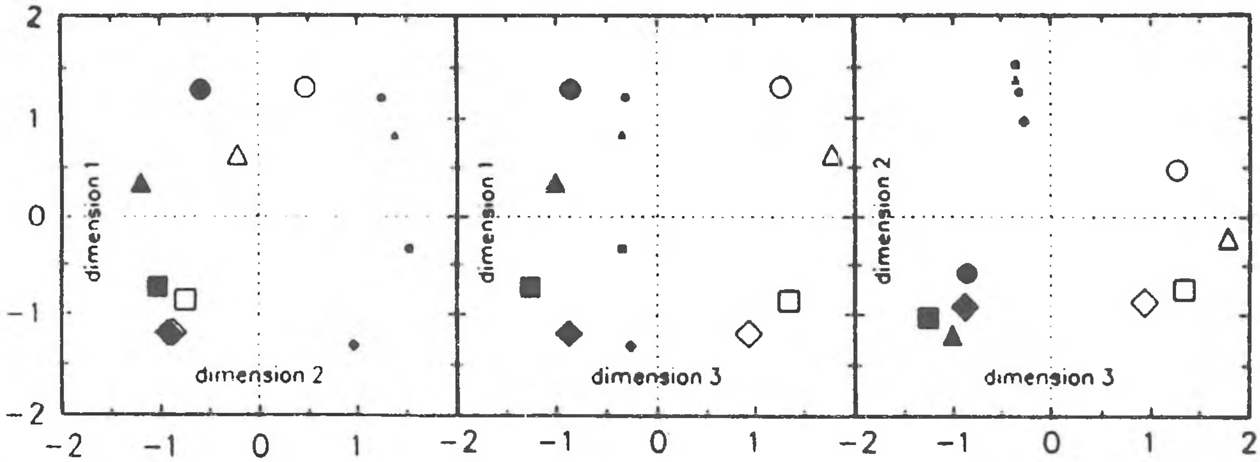

Figure 2. Three-dimensional scaling solution for similarity judgments obtained in Experiment 2. Each point represents the location of an individual stimulus in space. The solution accounted for 80.9% of the observed variance in the similarity judgments

Table 5

Perceptual similarity matrix for S2. Each entry indicates the total number of times each pair of stimuli was selected as more similar than other pairs. The maximum attainable value was 20; the minimum was 0.

		C4			C5			C6			C7		
		V1	V3	V4	V1	V3	V4	V1	V3	V4	V1	V3	V4
C4	V1		6	8	16	9	8	11	12	9	11	10	7
	V3			18	5	19	16	1	10	15	3	5	13
	V4				3	15	17	1	4	16	1	3	14
C5	V1					6	5	19	9	3	19	14	7
	V3						14	3	15	13	4	6	13
	V4							5	11	19	4	3	17
C6	V1								10	6	18	16	6
	V3									12	11	15	14
	V4										6	5	15
C7	V1											17	6
	V3												8
	V4												

Table 6

Perceptual similarity matrix for S6. Each entry indicates the total number of times each pair of stimuli was selected as more similar than other pairs. The maximum attainable value was 20; the minimum was 0.

		C4			C5			C6			C7		
		V1	V3	V4	V1	V3	V4	V1	V3	V4	V1	V3	V4
C4	V1		17	19	12	11	9	7	4	5	3	0	3
	V3			18	11	15	10	5	7	6	3	1	2
	V4				10	13	15	4	6	4	1	5	2
C5	V1					19	15	14	14	12	8	5	6
	V3						16	9	12	15	6	5	5
	V4							12	9	13	3	6	6
C6	V1								20	16	15	15	14
	V3									18	11	12	12
	V4										10	9	13
C7	V1											18	19
	V3												20
	V4												

Stimuli which were primarily labelled as "sh" were weighted negatively, and stimuli that were perceived as "ss" were weighted positively in this analysis. The third dimension, which distinguishes among the three vowels that were paired with the consonants, identifies the remaining 6.7% of the explained variance.

The differences between subjects are most visible in Figure 3, where the weights given by each subject for each dimension are displayed. This solution is consistent with the observation that the relative contribution of vowel and fricative information to perceptual similarity judgments varied from listener to listener. For several listeners (6, 10[shown by the symbol A in Figure 3] and 11[shown by B in Figure 3]), the first dimension (related to the physical properties of the frication noise) is weighted very highly, and dimension 2 (related to the mean group identification of the fricative) received little or no weight. For most other subjects, dimension 2 was weighted substantially, as was dimension 1.

A further reduction of the data for the two subjects (S2 and S6) who differed dramatically in the three-dimensional solution is presented in Figure 4. Figure 4a and 4b show the obtained perceptual similarity score for every pair of stimuli plotted as a function of the probability that the pair had been assigned the same fricative labels ($p["ss"]_{STIM1} * p["ss"]_{STIM2} + p["sh"]_{STIM1} * p["sh"]_{STIM2}$), for Subjects 2 and 6 respec tively. Figure 4c and 4d present the obtained perceptual perceptual similarity

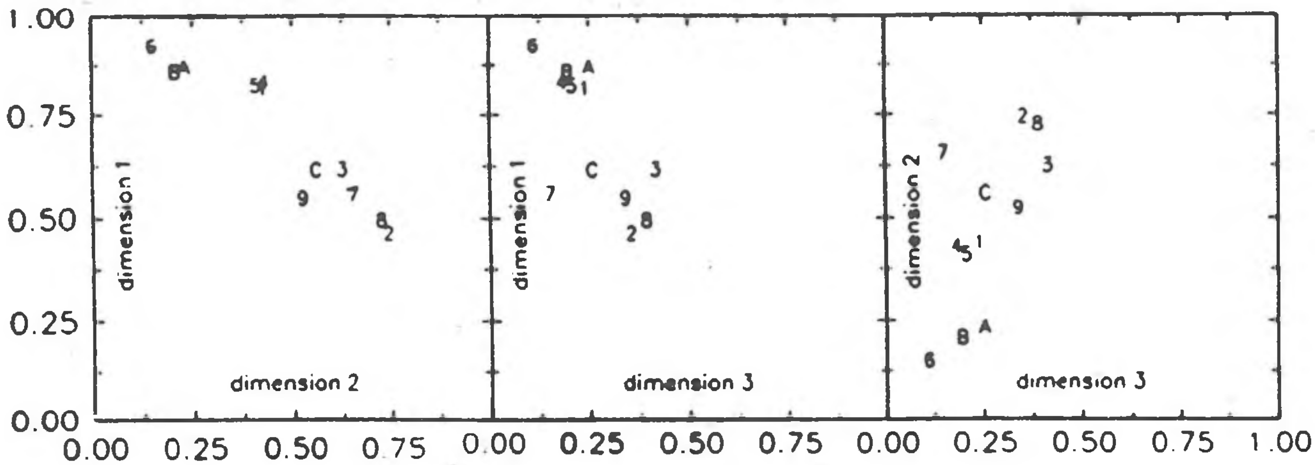

Figure 3. Three-dimensional scaling solution for similarity judgments. Individual listener's values for each dimension are indicated by subject numbers (S1-S9) and by A, B, and C for S10, S11, and S12 respectively.

score for every pair of stimuli plotted as a function of the difference in the centre frequency of the frication noise, again for Subjects 2 and 6, respectively. For S2, similarity judgments were related to the **labelling judgments** (r = 0.88; cf. Figure 4a) and not to the physical differences between the frication noises (r = -0.17; cf. Figure 4c). On the other hand, for S6, similarity judgments were strongly related to the **physical differences** between the consonants (r = -0.94; cf. Figure 4d), rather than to the labels that were assigned to stimuli (r = 0.30; cf. Figure 4b). The other subjects were distributed between these two extremes — making more or less use of both physical differences between the fricative portions of the signals and the labels given to the signals.[3]

6. GENERAL DISCUSSION

In these experiments the effect of vowel context on the perception of a preceding fricative consonant was studied using two experimental paradigms: a labelling task, in which listeners were forced to make a binary labelling decision, and a comparison task, in which listeners rated the perceptual similarity of pairs of fricatives. The results of the labelling task (Experiment 1) confirm the reliable occurrence of context-conditioning, where the label assigned to each stimulus reflected **both** the acoustic properties of the fricative sound and the acoustic vowel context in which the fricative information was presented. Listeners differed in the extent of the influence of vowel context on the perception of the fricatives, but all listeners showed the systematic influence of the vowels on the consonant identification over a range of frication frequencies. The results confirm those of Mann and Repp (1980) and Whalen (1981), with a new set of entirely synthetic syllables.

The results of the triadic comparison task (Experiment 2) showed that reliable patterns of similarity judgments could be obtained with context-conditioned signals. These similarity judgments were compared to: (a) the physical (acoustical) similarity of the fricative stimuli, and (b) the **predicted** similarity — derived from the identification (labelling) data obtained in Experiment 1. There was a continuum formed by the way in which individual listeners combined the two types of available information in making their similarity judgments. For **some** listeners, similarity judgments were made almost independently of the labels that were assigned to the fricatives; for other listeners, judged similarity was substantially a function of the fricative labels. Two subjects — S2 and S6 — bounded the extremes of this continuum, with S2's similarity responses being strongly linked to the labels assigned and S6's responses to the acoustical differences between the fricative portions of the signals.

Listeners who made greater use of the pitch of the fricative in making similarity judgments may be less strongly influenced by the vowel context in their identification judgments (for example S6 showed relatively small shifts in the phonemic boundary as a function of vowel context); alternatively, such subjects may be more "analytical" listeners than others and better able to "tune" their listening to one portion of the syllable while ignoring the rest. The former hypothesis is not supported by the data of Repp (1981) who did not observe a relation between the magnitude of the context effect and the ability of listeners to discriminate between within phone class fricatives.

The present findings seem to extend the traditional notion of "categorical perception" of phonemes (e.g., Liberman, et al., 1967), to view the perception of speech as being a continuum of abilities. Certainly, an extreme categorical view is not

[3]These factors are, of course, not completely independent, because identification itself was dependent on the frication frequency as shown in the results of Experiment 1.

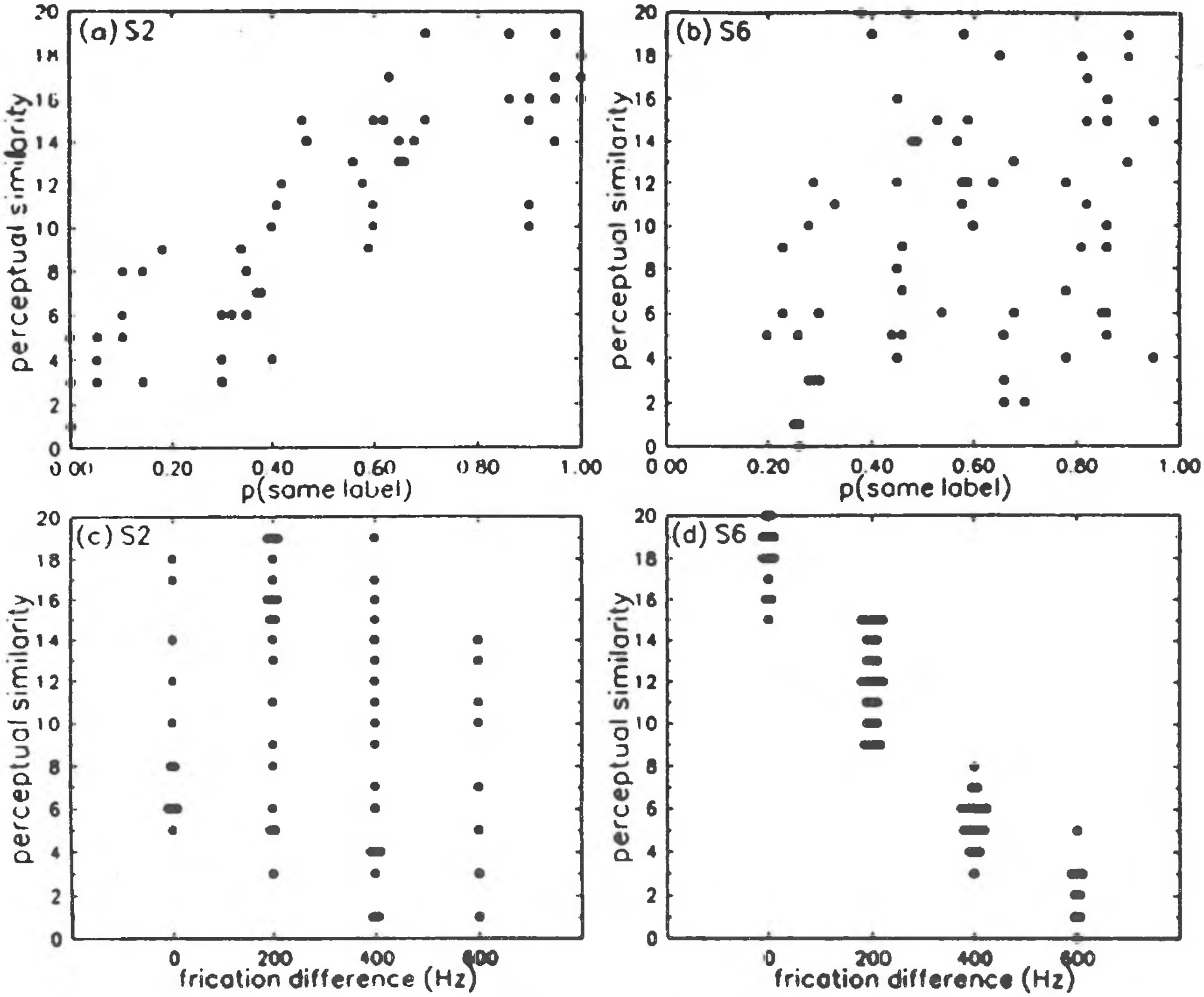

Figure 4. Judged similarity vs labelling similarity and fricative differences. In panels (a) and (b), perceptual similarity values are plotted as a function of the probability that a pair of fricatives would be assigned the same label for S2 and S6. In panels (c) and (d), perceptual similarity is plotted as a function of the difference in the centre frequency of the frication noises in each pair of consonants for S2 and S6. Horizontal jitter has been introduced to the data points in order that multiple data points do not mask each other.

consistent with these data: many of the listeners were more sensitive to differences between stimuli than a simple, binary-labelling view would admit. Indeed, some subjects clearly were able to make similarity judgments on the basis of physical similarities between the fricative noises, with little reference to the labels assigned to the stimuli. Implicit in this ability is the capacity to discriminate between phonemes that belong to the same identification category. Moreover, the continuous nature of the difference among subjects in the extent to which they relied on stimulus labels seems inconsistent with a view that subjects were responding in either a "speech" or an "auditory" mode (Liberman & Mattingly, 1985).

Of considerable interest is the origin of the individual differences in performance on the two tasks. One clear possibility is that, with further practice on the triadic comparison task, or with re-instruction to focus the listener's attention on the auditory as opposed to phonemic (linguistic) cues, the patterns of similarity judgments shown by a few listeners in the present study might be shown by all listeners.

7. REFERENCES

Abbs, MS & Minifie, FD (1969) Effect of acoustic cues in fricatives on perceptual confusions in preschool children, **J. Acoust. Soc. Am. 46**, 1535 1542.

American National Standard Specification for Audiometers (1989) **ANSI S3.6-1989.**

Bailey, PJ, Summerfield, Q & Dorman, M (1977) On the identification of sine-wave analogs of certain speech sounds, **Haskins Laboratories Status Report on Speech Research, 51/52**, 1-25.

Beck, WG, Leek, MR & Dorman, MF (1988) Quantifying perceptual distance among synthetic vowels, **Am. Speech Hear. Assoc. 30**, 87(A).

Behrens, SJ & Blumstein, SE (1988) Acoustic characteristics of English voiceless fricatives: a descriptive analysis, **J. Phon. 16**, 295-298.

Best, CT, Morrongiello, B & Robson, R (1981) Perceptual equivalence of acoustic cues in speech and nonspeech perception, **Percept. Psychophys. 29**, 191-211.

Jamieson, DG, Nearey, TM & Ramji, K (1989) CSRE: A Speech Research Environment, **Canadian Acoustics, 17**, 23-35.

Klatt, DH (1974) The duration of [s] in English words, **J. Speech Hear. Res. 17**, 51-63.

Klatt, DH (1980) Software for a cascade/parallel formant synthesizer, **J. Acoust. Soc. Am. 67,** 971-995.

Kuhl, PK (1991) Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not, **Percept. Psychophys. 50**, 93-107.

Kunisaki, O & Fujisaki, H (1977) On the influence of context upon perception of voiceless fricative consonants, **Annual Bulletin of the Tokyo Research Institute for Logopedics and Phoniatrics, 11**, 85-91.

Levelt, WJM, van de Geer, JP & Plomp, R(1966) Triadic comparisons of musical intervals, **Brit. J. Math. Stat. Psych. 19,** 163-179.

Liberman, AM, Cooper, FS, Shankweiler, DP & Studdert-Kennedy, M (1967) Perception of the speech code, **Psychol. Rev. 74**, 431-461.

Liberman, AM & Mattingly, IG (1985) The motor theory of speech perception revisited, **Cognition, 21**, 1-36.

Mann, VA & Liberman, AM (1983) Some differences between phonetic and auditory modes of perception, **Cognition, 14,** 211-235.

Mann, VA & Repp, BH (1980) Influence of vocalic context on perception of the [ʃ]-[s] distinction, **Percept. Psychophys. 28**, 213-228.

Mann, VA, Sharlin, HM & Dorman, M (1985) Children's perception of sibilants: The relation between articulation and perceptual development, **J. Exp. Child Psych. 39**, 252-264.

Nittrouer, S & 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.

Oller, DK, Eilers, RE, Miskiel, E, Burns, R & Urbano, R (1991) The stopglide boundary shift: modelling perceptual data, **Phonetica.**

Plomp, R (1970) Timbre as a multidimensional attribute of complex tones, in R Plomp & G. Smoorenburg (eds.) **Frequency Analysis and Periodicity Detection in Hearing**, 397-414.

Pols, LCW (1970) Perceptual space of vowel-like sounds and its correlation with frequency spectrum, in R Plomp & G Smoorenburg (eds.) **Frequency Analysis and Periodicity Detection in Hearing**, 463-473.

Pols, LCW, van der Kamp, LJT &Plomp, R (1969) Perceptual and physical space of vowel sounds, **J. Acoust. Soc. Am. 46,** 458-467.

Rakerd, B & Verbrugge, RR (1985) Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels, **J. Acoust. Soc. Am. 77,** 296-301.

Repp, BH (1981) Two strategies in fricative discrimination, **Percept. Psychophys. 30,** 217-227.

Repp, BH (1982) Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception, **Psych. Bull. 92**, 81-110.

Roskam, EE (1979) The method of triads for nonmetric multidimensional scaling, in JC Lingoes, EE Roskam & I Borg (eds) **Geometric Methods for Representations of Relational Data,** 497-510.

Whalen, DH (1981) Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary, **J. Acoust. Soc. Am. 69,** 275-282.

Yeni-Komshian, G & Soli, S (1979) Extraction of vowel information from frication spectra, **J. Acoust. Soc. Am. 65**, Suppl. 1, 57.

AUTHOR NOTES

This paper presents part of a doctoral dissertation submitted to the University of Minnesota. Portions of this work were reported at the 116th Meeting of the Acoustical Society of America (J. Acoust. Soc. Am. 84, S157).

This work was supported by NIH NS12125, a SSHRC doctoral fellowship , the Bryng Bryngelson Research Fund and NSERC.