Perception & Psychophysics
1981, 29 (3), 191-211

Perceptual equivalence of acoustic cues in speech and nonspeech perception

CATHERINE T. BEST
Haskins Laboratories, New Haven, Connecticut 06510

and

BARBARA MORRONGIELLO and RICK ROBSON
University of Massachusetts, Amherst, Massachusetts 01003

Trading relations show that diverse acoustic consequences of minimal contrasts in speech are equivalent in perception of phonetic categories. This *perceptual equivalence* received stronger support from a recent finding that discrimination was differentially affected by the phonetic cooperation or conflict between two cues for the /slIt/-/splIt/ contrast. Experiment 1 extended the trading relations and perceptual equivalence findings to the /sei/-/stei/ contrast. With a more sensitive discrimination test, Experiment 2 found that cue equivalence is a characteristic of perceptual sensitivity to phonetic information. Using "sine-wave analogues" of the /sei/-/stei/ stimuli, Experiment 3 showed that perceptual integration of the cues was phonetic, not psychoacoustic, in origin. Only subjects who perceived the sine-wave stimuli as "say" and "stay" showed a trading relation and perceptual equivalence; subjects who perceived them as nonspeech failed to integrate the two dimensions perceptually. Moreover, the pattern of differences between obtained and predicted discrimination was quite similar across the first two experiments and the "say"-"stay" group of Experiment 3, and suggested that phonetic perception was responsible even for better-than-predicted performance by these groups. Trading relations between speech cues, and the perceptual equivalence that underlies them, thus appear to derive specifically from perception of phonetic information.

Research with a variety of minimal segmental distinctions in synthetic speech has shown that perception of a phonetic contrast can be cued by appropriate change in the major acoustic property that differentiates that contrast in natural speech (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith, 1961; Liberman & Studdert-Kennedy, 1978; Studdert-Kennedy, 1976). However, minimal articulatory contrasts result in concurrent differences along more than one acoustic dimension. As this last fact might suggest, perceptual studies indicate that listeners make use of the various acoustic consequences of a given spoken segmental distinction. For example, voicing distinctions in initial, prestress position can be cued not only by changes in voice onset time (VOT—as an acoustic measure), with all else held constant, but also by changes in F1 onset frequency, F0 contour, or aspiration energy (e.g., Haggard, Ambler, & Callow, 1970; Lisker, 1975; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1979; Lisker, Note 1), all of which are acoustic correlates of laryngeal timing distinctions in stop production (Abramson & Lisker, 1965). A variety of acoustic consequences of articulatory distinctions have also been found to serve as cues for place (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977; Harris, Hoffman, Liberman, Delattre, & Cooper, 1958) and manner of articulation (e.g., Dorman, Raphael, & Liberman, 1979; Miller & Liberman, 1979; Repp, Liberman, Eccardt, & Pesetsky, 1978).

The fact that various acoustic properties serve as cues for a phonetic contrast suggests that they provide equivalent information about the distinction involved (cf. Dorman et al., 1977; Repp et al., 1978). If different cues do provide equivalent phonetic information, it should be possible to offset a "weakness"

This work was supported by NICHD Grant HD01994 to the Haskins Laboratories and NINCDS Postdoctoral Fellowship Grant NS5085 to the first author. We thank the following people for their valuable contributions: Alvin M. Liberman, for use of the Haskins facilities, and for helpful advice throughout the project; Rachel Clifton, for use of her auditory perception laboratory at the University of Massachusetts at Amherst and for advice during the initial stages of the project; Terry G. Halwes, for instruction on stimulus development; and Michael Studdert-Kennedy, Bruno H. Repp, Robert E. Remez, Robert Verbrugge, and Steven S. Braddon for helpful discussions and suggestions on earlier drafts of this paper. These findings were presented at the 98th Meeting of the Acoustical Society of America, Salt Lake City, Utah, November 1979. An earlier version of this paper also appears in the *Haskins Laboratories: Status Report on Speech Research*, 1980, SR-62.
The senior author is currently at the Neurosciences and Education Program, Box 142, Teachers College, Columbia University, New York, New York 10027.

0031-5117/81/030191-21$02.35/0

in one cue by strengthening the value of another (within limits possibly defined by the acoustic effects of natural articulations). Empirical work has supported this hypothesis—perceptual *trading relations* have been found among diverse cues for voicing (e.g., Summerfield & Haggard, 1977), place (e.g., Bailey & Summerfield, 1980; Hoffman, 1958), and manner distinctions (e.g., Dorman, Raphael, & Isenberg, 1980). However, while trading relations indicate that different acoustic properties may cue a given phonetic category, they cannot support a stronger claim that the cues yield *qualitatively equivalent percepts*. Trading-relation studies have typically used forced-choice identification tests, which merely assess whether any of several acoustic manipulations are *acceptable* as cues for a given contrast. Forced-choice tests do not measure whether the same-category percepts based on different acoustic cues are identical in quality. Qualitative equivalence between percepts based on diverse acoustic cues will be referred to as *perceptual equivalence*.

In a recent experiment on the /slIt/-/splIt/ contrast, Fitch, Halwes, Erickson, and Liberman (1980) conducted a more stringent test of perceptual equivalence between two cues for the medial stop /p/. A trading relation between two synthetic /slIt/-/splIt/ continua showed that when the formant transitions following silent closure were appropriate for a natural "slit" (/slIt/-biased continuum), listeners needed a significantly longer closure gap than when the transitions were appropriate for "split" (/splIt/-biased continuum) in order to hear "split" 50% or more of the time. Thus, additional silence compensated perceptually for "weakness" of the /slIt/-biased spectral cue. If the convergence of the two cues upon a unitary speech percept ("split") was tied to their common articulatory origin, Fitch et al. (1980) reasoned, then differently cued stimuli should be difficult to discriminate within a phonetic category (i.e., they should be perceptually equivalent).

Mere demonstration of poor within-category discriminability between cues would not support perceptual equivalence, however, since the null hypothesis cannot be proven. Therefore, Fitch et al. tested whether discrimination performance would be differentially affected by cooperation or conflict of the two cues along the phonetic dimension, using an oddity procedure that included three types of comparisons between stimuli from the two continua. In "two conflicting cues" comparisons, /slIt/-biased stimuli (spectral bias toward "slit") had longer closures (temporal bias toward "split") than /splIt/-biased stimuli, such that the two cues exactly *cancelled* one another phonetically. In the "two cooperating cues" comparisons, the two cues *complemented* each other phonetically—on all trials the /splIt/-biased stimuli had a longer closure gap (by the same amount of difference as in "two conflicting cues") than the /slIt/-biased stimuli. In "one-cue" comparisons, the stimuli contrasted only on the spectral dimension.

The "phonetic" hypothesis was that if the two cues showed perceptual equivalence along a single phonetic dimension, /slIt/-biased and /splIt/-biased stimuli would be quite difficult to discriminate when they belonged to the same phonetic category. This would be the case for all "conflicting cues" comparisons. In contrast, /slIt/-biased and /splIt/-biased stimuli should have been comparatively easy to discriminate when they belonged to different phonetic categories; this would be the case for those "one-cue" comparisons that straddled the category boundary. Enhancing the between-category differences should lead to the highest discrimination performance; this was accomplished by those "cooperating cues" comparisons that straddled the phonetic category boundary. The alternative "auditory" hypothesis was that the two cues might remain discriminable on an auditory basis. In that case, performance would be equally high across the board for both "two-cue" comparison types, since they contrasted along two acoustic dimensions, relative to performance on "one-cue" comparisons, which contrasted on only one acoustic dimension. The results clearly supported the "phonetic" hypothesis, indicating that the two acoustic cues were perceptually equivalent along a single dimension in speech.

The three-way oddity results may offer an important contribution to our knowledge about the conditions under which information from diverse acoustic dimensions is integrated in speech perception. However, other phonetic category cues should be explored to assess the extent and reliability of perceptual equivalence among phonetic cues (although, indeed, the many reported trading relations make it unlikely that perceptual equivalence is idiosyncratic to /slIt/-/splIt/). Experiment 1 of this paper extended the trading relations and perceptual equivalence findings to the /sei/-/stei/ contrast,[1] which is simpler than /slIt/-/splIt/ in phonetic, articulatory, and acoustic properties. /Sei/ and /stei/ are dynamically similar, in that each starts (/s/) with the tongue pressed against the inner sides of the upper teeth, tongue-tip nearly in contact with the alveolar ridge and/or the inner side of the juxtaposed front teeth, and each ends (/ei/) with a more open vocal tract. The result is that the vocalic formant transitions are very similar in the two words. The major acoustic consequences of *complete* linguoalveolar (or -dental) closure following /s/ (for /stei/) are the introduction of a silent gap and a lower vocalic F1 onset frequency. (For general accounts of stop closure properties, see Delattre, Liberman, & Cooper, 1955; Fant, 1962; Stevens, 1971, 1974.) In the /slIt/-/splIt/ contrast, on the other hand, bilabial juxta-

position and the consequent labial transitions of the upper formants occur only for /splIt/.

If wider support was to be found (in Experiment 1) for the suggestion that perceptual integration of diverse acoustic cues takes account of their common origin in speech production, then it might be that trading relations and perceptual equivalence between cues are specific to the perception of phonetic information. Experiments 2 and 3 were therefore designed to test two alternatives to the notion that such findings might be unique to phonetic perception.

First, the oddity procedure might not provide the optimal test for true equivalence of the *perceptual* qualities of phonetic cues. It is widely believed that the oddity procedure places heavy demands on auditory short-term memory, which may have encouraged listeners to categorize each stimulus in order to distinguish among the *categorizations*, rather than discriminate finer-grained acoustic qualities that might have been perceptually available prior to categorization. Moreover, since the discrimination test was administered after the forced-choice identification test, test order could also have biased the subjects to categorize stimuli before discriminating them. Experiment 2 minimized these problems by using a discrimination test with lower memory demands, and by collecting discrimination data prior to identification data.

Second, although it has been suggested that perceptual equivalence between diverse speech cues would occur only for perception of phonetic category information (cf. Fitch et al., 1980), no direct studies have been conducted with *non*speech sounds. Experiment 3 tested the "psychoacoustic" alternative that trading relations and perceptual equivalence between cues might occur for nonspeech sounds with complex acoustic properties like those used in our /sei/-/stei/ contrasts.

EXPERIMENT 1

Method

Subjects

The subjects were 15 undergraduate students. Ten subjects from Yale University were tested at Haskins Laboratories and paid $3/h for participation. The other five were from the University of Massachusetts at Amherst, and they received grade-credits toward their introductory psychology courses; they completed the tests in an auditory perception laboratory at their psychology department. All subjects reported having normal hearing in both ears (no diagnosed hearing losses).

Stimuli

Two 290-msec, three-formant vocalic syllables were created on the Haskins parallel-resonance synthesizer. They were stylized versions of the vocalic portions from natural, male utterances of "say" and "stay," and differed from one another only in F1 onset frequency (230 Hz vs. 430 Hz), as the acoustic analyses in the

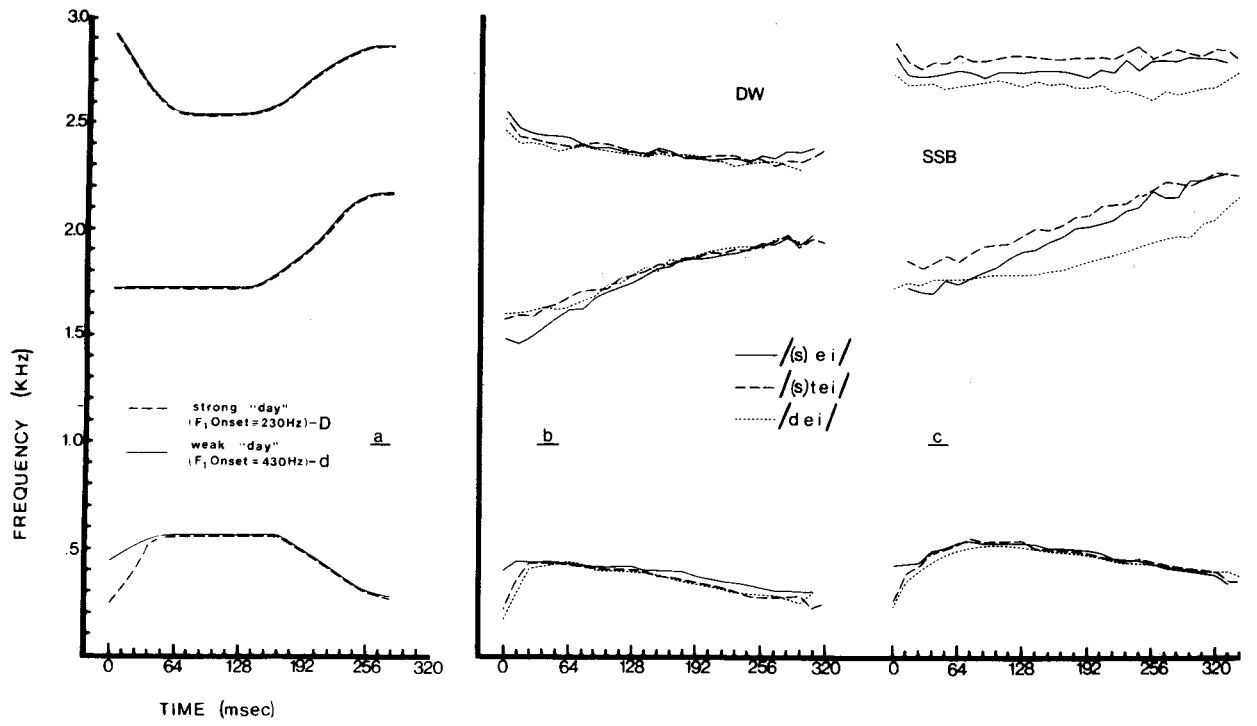

Figure 1. Acoustic measurements of F1, F2, and F3 for the following stimuli, from spectral sections of consecutive 12.8-msec windows: (a) synthetic weak "day" and strong "day," (b) averages of five tokens each of "say," "stay," and "day" by male talker D.W. (Fort Worth, Texas), and (c) by male talker S.S.B. (Brooklyn, New York—variations in S.S.B.'s F2 and F3 frequencies reflect slight vowel color variations among these words in his dialect).

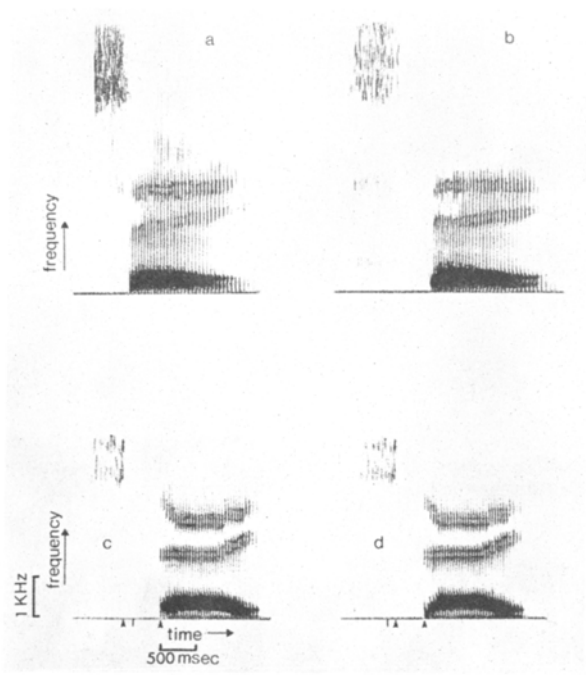

Figure 2. Spectrograms of natural, male utterances of (a) "say," and (b) "stay," by talker S.S.B., and synthesized vocalic tokens of "stay" made from (c) weak "day" and (d) strong "day." The synthetic tokens are preceded by an natural 120-msec /s/, and although they had different silent gap durations, a trading relation showed them to be identified equally consistently as "stay."

left-hand panel of Figure 1 illustrate. The formant amplitudes and overall amplitude envelopes of the stimuli were identical, as were the time-varying frequency characteristics of F2 and F3, and also of F1 beyond the initial 40-msec transition differences (details in Appendix A). Acoustic analyses on five tokens each of "say," "stay," and "day" uttered by two males (center and right-hand panels of Figure 1) showed that the most pronounced spectral difference between the vocalic parts of "say" and "stay" was a lower F1 onset frequency for "stay." The vocalic portion of "stay" and "day" involve nearly identical articulatory gestures[2]; as would be expected, they were virtually identical in formant onset characteristics. Figure 2 shows spectrograms of a natural "say" and "stay," and of "stay" tokens made from the two synthetic syllables by preceding each with a natural /s/ and a silent (closure) interval.

To determine how well the F1 onset frequencies of the two isolated synthetic syllables would support perception of an alveolar stop, 12 additional synthetic syllables were generated for an "ay"-"day" continuum. F1 onset was varied between 160 Hz and the 611-Hz steady state, in 33-Hz steps. A randomized forced-choice identification test (10 judgments/token) with 18 naive listeners (9 from Yale, 9 from University of Massachusetts/Amherst) revealed a fairly sharp category boundary. The test syllable with the 430-Hz F1 onset was perceived as nearly equivocal between "ay" and "day," and will hereafter be called *weak "day,"* abbreviated d. The stimulus with the 230-Hz F1 onset was identified 100% of the time as "day," and will be called *strong "day,"* abbreviated D.

The D and d syllables were each used to generate a "say"-"stay" continuum that incorporated a natural 120-msec /s/ from a male "say" utterance. The /s/ and the synthetic syllable were separated by silent gaps ranging between 0 and 136 msec, in 8-msec increments, resulting in two "say"-"stay" continua with 18 members each (from s[0]d to s[136]d, and from s[0]D to s[136]D).

Procedure

For the forced-choice identification test, a randomized sequence of 360 single-item trials was generated, with 2.5-msec interstimulus intervals (ISIs). The sequence contained 10 repetitions of all items from the two "say"-"stay" continua, and was presented in sound-attenuated test rooms at a comfortable listening level (approximately 75 dB). The subjects identified each stimulus in writing as "say" or "stay."

The subjects took a 15-min break following the identification test, and then completed a three-way oddity discrimination test that included the following comparison types: "one cue," "two cooperating cues," and "two conflicting cues." In the 18 possible "one-cue" comparisons, the three stimuli on each trial had identical gap durations (comparisons covered the entire 0-to-136-msec range), and the "odd" stimulus differed from the other two in its F1 onset frequency (D vs. d). For both two-cue comparison types, a 24-msec gap difference was chosen to compensate phonetically for the F1 difference.[3] In the 15 possible "cooperating cues" comparisons, D stimuli always had a 24-msec longer silent gap (both cues biased toward "stay") than did d stimuli, so that the phonetic complementarity of the cues enhanced the between-category differences; comparisons ranged from s[0]d-s[24]D to s[112]d-s[136]D. For the 15 possible "conflicting cues" comparisons, d stimuli (spectral bias toward "say") had a 24-msec longer gap (temporal bias toward "stay") than did D stimuli, so that phonetic *cancellation* between the cues minimized any between-category differences; these comparisons ranged from s[0]D-s[24]d to s[112]D-s[136]d. Phonetically based discrimination should be facilitated for **"cooperating cues" comparisons that straddled the category boundary, because between-category differences were enhanced for those** comparisons, relative to "one-cue" comparisons that straddled the boundary. Discrimination of the "conflicting cues" comparisons should be lowest, because between-category differences were minimized in all comparisons of that type—they never straddled the boundary.

The discrimination test was a randomized sequence of all stimulus comparisons for all three comparison types, and included six presentations[4] of each of the 48 possible stimulus comparisons (total items = 288). Within each trial, ISIs were 1 sec, and inter-trial intervals (ITIs) were 3 sec.

Results and Discussion

Identification Test

The results for the forced-choice identification test are shown in Figure 3. The mean category boundary

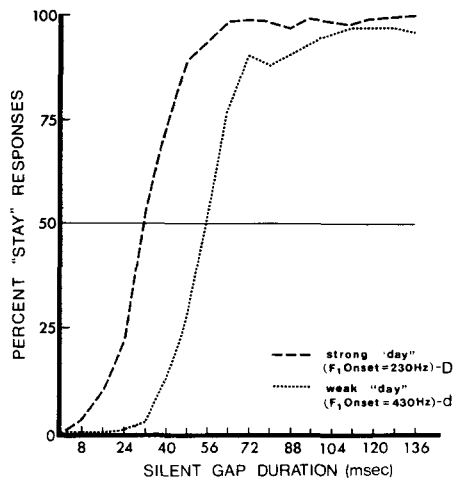

Figure 3. Identification functions for the strong "day" and weak "day" continua in Experiment 1.

(50% "stay" responses) for the strong "day" (D) function fell at 32.4 msec (range = 11.4-52.0 msec), and that for the weak "day" (d) function at 57.1 msec (range = 40.0-94.0 msec); the boundary difference was significant ($t = 7.23$, $p < .001$). The average trading relation between the two continua was thus 24.6 msec (range = 9.3-54.0 msec). To be perceived as "stay," the d stimuli required approximately 24 msec more silence between the /s/ and the vocalic syllable than did the D stimuli.

Three-Way Oddity Test

The results for the three-way oddity test are shown in Figure 4. Obtained functions for the three comparison types are represented in the left panel. The right panel represents the corresponding *predicted* functions derived from the identification data (formula in Appendix B), which indicate the limits of the effect of perceived category differences upon discrimination performance. The *obtained* data were submitted to an analysis of variance (ANOVA) crossing 3 comparison types with 15 stimulus pairs, which included only the range of overlap between one-cue and two-cue comparison types (mean gap durations per comparison of 12-128 msec). The Comparison Types effect [$F(2,28) = 34.14$, $p \ll .001$] supported the perceptual equivalence prediction that the order of performance levels would be: "cooperating cues" > "one cue" > "conflicting cues" (see Table 1 for Tukey pairwise contrasts). The "phonetic" argument also predicted improved discrimination performance on comparisons that straddled category boundaries, especially if category differences were enhanced. In line with this prediction, performance near the boundary was higher than within-category performance; that is, there were boundary-related peaks in performance [Stimulus Pairs: $F(14,196) = 7.01$, $p \ll .001$]. In addition, the Comparison Types by Stimulus Pairs interaction [$F(28,392) = 3.31$, $p \ll .001$, broken down by simple effects tests] indicated that the magnitude of peak vs. trough level differences followed the order: "cooperating cues" [$F(14,588) = 27.13$, $p \ll .001$] > "one cue" [$F(14,588) = 2.39$, $p < .005$] > "conflicting cues" [$F(14,588) = 1.75$, $p = .05$].[5] The "phonetic" predictions were clearly supported.

Analyses were also conducted on the predicted data, obtained vs. predicted comparisons, and individual performance patterns (see details in Appendix C). As Figure 4 shows, the predicted discrimination pattern was essentially the same as the obtained pattern. Obtained performance levels were slightly higher than predicted, but only for stimulus comparisons that were removed from the between-category performance peaks by 16-24 msec or more (i.e., those that did *not* straddle the boundary). Moreover, the residual performance levels (obtained minus predicted level—the performance that was unexplainable by phonetic *category* differences) still followed the "phonetic" order: "cooperating cues" > "one cue" > "conflicting cues." Residual discrimination patterns will be discussed later in the paper, since they are best understood relative to the obtained-predicted differences found in Experiments 2 and 3.

Conclusions

The results of Experiment 1 clearly indicated a trading relation and perceptual equivalence between

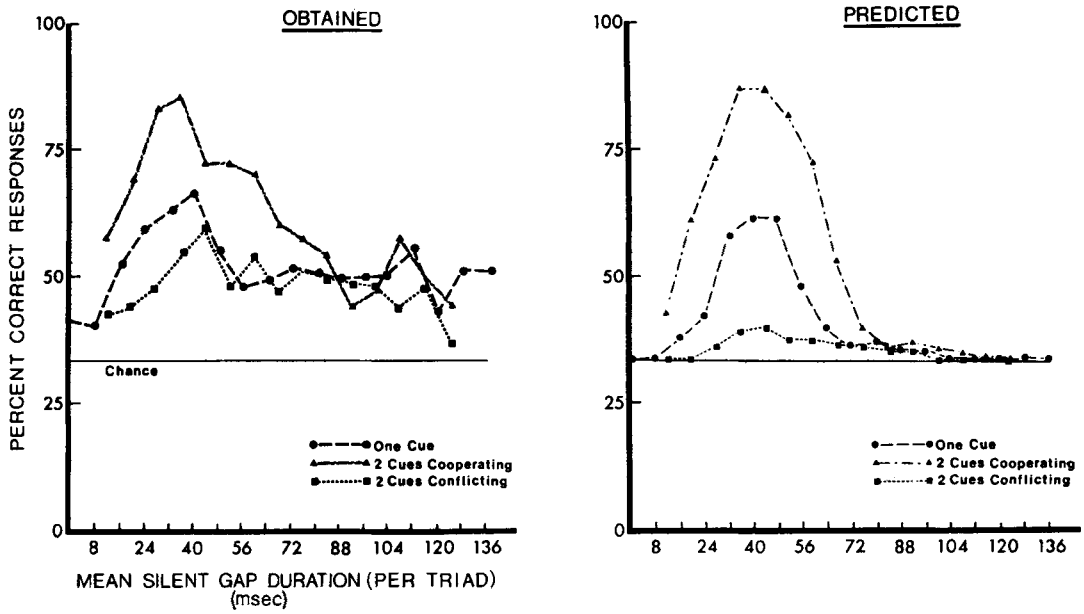

Figure 4. Obtained (left-hand panel) and predicted (right-hand panel) results for the three-way oddity test in Experiment 1.

Table 1
ω Values of Tukey Tests on Significant Comparison Type (CT) Effects From ANOVAs for Experiments 1, 2, and 3

	One Cue vs. Cooperating Cues	Cooperating Cues vs. Conflicting Cues	Conflicting Cues vs. One Cue
	Experiment 1 (Represented as Mean Percentages Correct)		
Obtained Data	8.72††	13.16††	4.44*
Predicted Data	9.57†	18.14††	8.58†
Obtained "Peak-Range" Data	17.22††	25.97††	8.75*
Predicted "Peak-Range" Data	25.91††	43.22††	17.31†
	Experiment 2 (Represented as Mean True d′ Values)		
Obtained Data	1.43††	2.04††	.61*
Predicted Data	1.85††	2.13††	.29
Observed "Peak-Range" Data	1.68††	2.88††	1.20**
Predicted "Peak-Range" Data	2.31††	3.32††	1.01**
	Experiment 3 (Represented as Mean Number Correct, Out of Six)		
"Say"-"Stay" Listeners	.72†	1.08††	.36*
"Temporal" Listeners (Peak)	.76*	.91*	.15

*$p < .05$. **$p < .01$. †$p < .005$. ††$p < .001$.

gap duration and F1 onset frequency as cues for the "say"-"stay" distinction. The Fitch et al. (1980) findings were thus replicated for a different phonetic category contrast. Silent gap duration and F1 onset frequency appeared to have converged on a single dimension in phonetic perception.

EXPERIMENT 2

Although Experiment 1 suggested that the silence and F1 spectral cues for the "say"-"stay" contrast are perceptually equivalent in quality, the oddity task's heavy demands on auditory short-term memory[6] may have biased the subjects to *recode* the rapidly fading sensory information into phonetic category information. Phonetic categorizations are believed to be better retained in memory than are raw acoustic properties, especially in the case of consonants (cf. Crowder, 1971, 1973; Crowder & Morton, 1969; Darwin & Baddeley, 1974; Fujisaki & Kawashima, 1969; Pisoni, 1975; Pisoni & Tash, 1974; Repp, Healy, & Crowder, 1979; Pisoni, Note 2). Thus, the oddity task may not be a sensitive test for qualitative equivalence of cues *at the sensory level*. Experiment 2 used a 2IAX ("same"-"different") discrimination procedure with short ISIs to induce performance that would better reflect perceptual sensitivity to the physical properties of the stimuli. In addition, the identification test was run after the discrimination task, to control against the possibility that obtaining forced-choice identifications before discrimination judgments might have introduced an experimental bias toward phonetic categorization, and thus against comparison of physical properties.

We used the 2IAX procedure to assess perceptual sensitivity, based on several considerations.[7] While standard signal detection theory (SDT—MacMillan, Kaplan, & Creelman, 1977) does not allow estimation of perceptual sensitivity from oddity data, it *does* permit estimation of perceptual sensitivity from ABX, 4IAX, and 2IAX data, through the use of the d′ sensitivity index. According to SDT predictions (MacMillan et al., 1977), d′ values (hence sensitivity) should be lowest for the 2IAX procedure. This is because that procedure biases observers to give "same" responses for physically different stimulus pairs that are difficult to discriminate, which artificially deflates standard d′ values [computed as z(Hits) − z(False Alarms); Kaplan et al., 1978]. However, the data on actual sensitivities of the paradigms are equivocal.[8] Moreover, a relatively new formula for computing bias-corrected, *true* d′ values from 2IAX data yields sensitivity values at least as high for 2IAX data as for AXB and 4IAX (Kaplan, MacMillan, & Creelman, 1978).[9] Finally, memory demands for making "same"-"different" judgments on two "say"-"stay" stimuli per trial would seem lower than for ABX or 4IAX judgments.

Method

Subjects

A new group of 14 subjects participated in this experiment. Seven were Yale undergraduates; the other seven were undergraduates at the University of Massachusetts/Amherst. All reported having normal hearing in both ears.

Stimuli

The stimuli from Experiment 1 were used again. This time, however, stimuli containing gaps over 96 msec were eliminated from the test, since they had been identified as "stay" nearly 100% of the time in Experiment 1. The truncated "say"-"stay" continua contained 14 stimuli each.

Procedure

The subjects first completed a three-way 2IAX test, which employed 300-msec ISIs and 2.5-sec ITIs. Stimulus pairs for the three types of test comparisons ("one cue," "cooperating cues," and "conflicting cues") were chosen from the truncated "say"-"stay" continua by the same means as described in Experiment 1, with 24-msec silence again used as the temporal compensation value in both two-cue comparison types. In addition, a fourth set of

"physically same" catch-trial comparisons was included to provide false-alarm-rate data. There were 11 possible mean gap values each for "one cue" (s[8]d-s[8]D to s[88]d-s[88]D) and "physically same" comparisons (s[8]d-s[8]d and s[8]D-s[8]D, to s[88]d-s[88]d and s[88]D-s[88]D), and 10 possible pairs each for "conflicting cues" (from s[0]D-s[24]d to s[72]D-s[96]d) and "cooperating cues" comparisons (from s[0]d-s[24]D to s[72]d-s[96]D).

Six judgments were obtained for each of the 42 possible stimulus contrasts (total items = 252), randomized across pairings and comparison types. Instructions attempted to focus attention on the differences in acoustic properties of the stimuli, rather than on phonetic categories. The subjects were told that most of the "different" pairs would be tokens of the same word, so they should listen closely for slight *sound* differences between members of same-word pairs.

A randomized 280-trial forced-choice identification test (2.5-msec ISIs), containing 10 repetitions of all stimuli in the truncated continua, was administered after the 2IAX test.

Results and Discussion

Identification Test

The identification results (lower panel, Figure 5) replicated the trading relation found in Experiment 1, this time with a boundary difference of 18.5 msec (range = 5.1-37.3 msec). The somewhat smaller trading relation may have resulted from truncating the continua (stimulus range effect), but was nonetheless significant ($t = 8.88$, $p < .001$). The D (strong "day") category boundary fell at 25.3 msec (range = 10.7-51.2 msec), and the d (weak "day") boundary at 43.8 msec (range = 33.6-60.0 msec).

Three-Way 2IAX Test

The three-way 2IAX results (upper panel, Figure 5) showed the "phonetic" order of "different" response levels for the three types of test comparisons ("cooperating cues" > "one cue" > "conflicting cues"), as did the three-way oddity results of Experiment 1. The small peak in percentage of "different" responses for "physically same" and catch trials near the category boundary resulted from the ambiguous identifications of the boundary stimuli.

The true d′ values offer a more accurate measure of differential *perceptual sensitivities* than do the raw data; Figure 6 shows the obtained (left-hand panel) and predicted sensitivity functions (right-hand panel) represented by these true d′ values (formulas in Appendix D). An ANOVA was run on the obtained true d′ data, spanning the overlap among comparison types (12-88-msec mean gaps), for the 3 comparison types by 10 stimulus pairs. The pattern of results (see Table 2, and the Tukey pairwise contrasts in Table 1) essentially replicated the Experiment 1 findings. The results of analyses on the predicted data, obtained vs. predicted comparisons, and individual performance patterns also replicated the previous findings (see Appendix E for details). Once again, detailed discussion of obtained vs. predicted differences will be deferred until later in the paper.

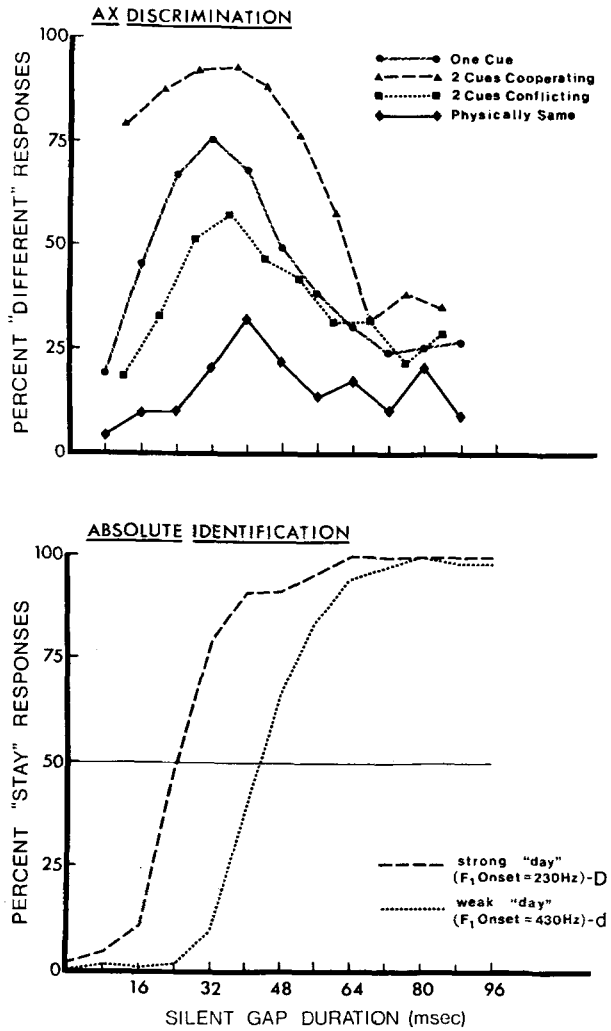

Figure 5. Obtained functions for the three-way 2IAX discrimination test (upper panel) and the forced-choice identification test (lower panel), for Experiment 2.

In contrast with the oddity test of Experiment 1, the 2IAX test produced a small, but significant, performance-level peak (higher d′ values) near the category boundary for "conflicting cues" comparisons. However, this "conflicting cues" peak reflects the fact that the trading relation found in Experiment 2 was only 18.5 msec, which did not match the predetermined 24-msec gap compensation value used. Since gap duration and F1 onset differences in the "conflicting cues" comparisons did not precisely cancel one another, there was a small but predictable enhancement of sensitivity near the boundary.

Conclusions

All of the major findings from Experiment 1 were replicated—a trading relation was found between the

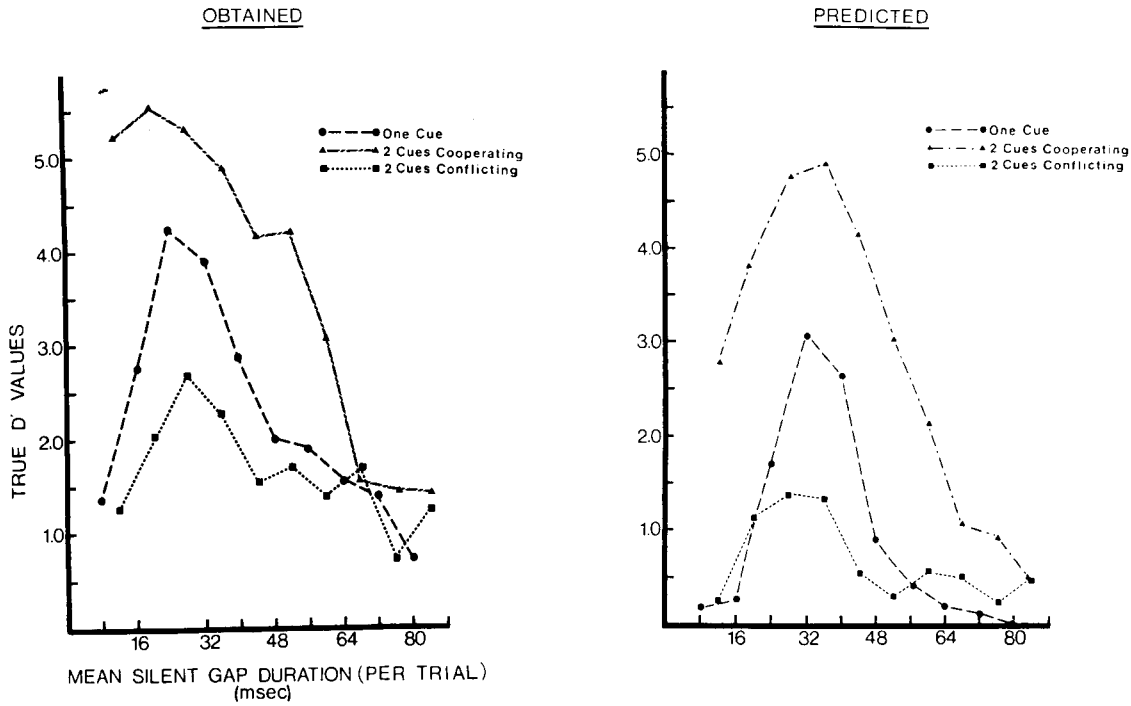

Figure 6. Obtained (left-hand panel) and predicted (right-hand panel) functions for true d' values on the three-way 2IAX test in Experiment 2.

Table 2
Results From the ANOVAs Performed on the True d' Values Computed for the Three-Way 2IAX Data From Experiment 2

ANOVA	Effect	df	F	p
Obtained Data (3CT* by 10SP**)	CT	2, 26	41.78	⩽.001
	SP	9,117	17.08	⩽.001
	CT by SP	18,234	6.48	<.001
CT by SP Simple Effects				
"Cooperating Cues"	SP	9,351	18.43	⩽.001
"One Cue"	SP	9,351	8.47	<.001
"Conflicting Cues"	SP	9,351	1.97	<.05
Predicted Data (3CT by 10SP)	CT	2, 26	109.99	⩽.001
	SP	9,117	17.88	⩽.001
	CT by SP	18,234	9.42	<.001
CT by SP Simple Effects				
"Cooperating Cues"	SP	9,351	26.91	⩽.001
"One Cue"	SP	9,351	13.39	<.001
"Conflicting Cues"	SP	9,351	1.83	>.05†
"Cooperating Cues" (2F†† by 10 SP)	F	1, 13	17.26	⩽.001
	SP	9,117	37.96	⩽.001
	F by SP	9,117	2.09	<.05
"One Cue" (2F by 10SP)	F	1, 13	35.23	⩽.001
	SP	9,117	13.02	⩽.001
	F by SP	9,117	1.90	>.05‡
"Conflicting Cues" (2F by 10SP)	F	1, 13	21.11	⩽.001
	SP	9,117	4.58	<.001
	F by SP	9,117	.49	n.s.
Obtained "Peak-Range" Data‡‡ (3CT)	CT	2, 26	31.66	⩽.001
Predicted "Peak-Range" Data (3CT)	CT	2, 26	66.23	⩽.001

**Comparison types ("cooperating cues," "one cue," "conflicting cues").* ***Stimulus pairs.* *†Marginal (.05 cut-off = 1.88).* *††Functions (predicted vs. observed).* *‡Marginal (.05 cut-off = 1.96).* *‡‡Mean value for gap durations between 20 and 48 msec (average per AX pair) in each type of test comparison.*

F1 onset and silence cues, and the pattern of 2IAX sensitivities fit the "phonetic" predictions. The two acoustic cues for the "say"-"stay" contrast appear to be perceptually equivalent, even under conditions designed to reduce demands on auditory short-term memory and to reduce experimentally induced biases to categorize stimuli before discriminating them. That is, the "phonetic" pattern of three-way discrimination performance seems not to depend on the employment of a task that places heavy demands on memory.

EXPERIMENT 3

The question that now arises, however, is: What is the origin of the equivalence in perceptual sensitivity to the temporal and spectral cues for the "say"-"stay" contrast? At least two possibilities present themselves: (1) the "phonetic" alternative, that the equivalence derives specifically from perception of phonetic information (recall that even in Experiment 2, the subjects perceived the stimuli as "say" and "stay") —that is, it occurs "only for sounds . . . being processed as speech" (Fitch et al., 1980, p. 344); or (2) the "psychoacoustic" alternative, that the pattern derives from general (not speech-specific) properties of auditory perception. Although we know of no research on psychoacoustic integration of acoustic cues like those we used in the "say"-"stay" research, the "psychoacoustic" alternative gains converging support from: (a) known tradeoffs in nonspeech perception (e.g., the time-intensity trade in auditory localization: Green, 1976); and (b) speech-relevant discontinuities in perception of changes along a single acoustic dimension in nonspeech stimuli (e.g., categorical perception for rise-time and onset-time nonspeech contrasts: Cutting & Rosner, 1974, 1976; Cutting, Rosner, & Foard, 1976; Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977; Pisoni, 1977).

Therefore, it would be important to determine whether there was some psychoacoustic interaction between the "say"-"stay" cues. For example, it could be that longer gaps were needed for the d stimuli than for the D stimuli to be heard as "stay," because the /s/ offset was closer in frequency to the 430-Hz F1 onset than to the 230-Hz F1 onset (by 3-4 critical bands). This possibility seems unlikely, though, because it contradicts findings that gap sensitivity is *inversely* related to the amount of frequency difference between the acoustic components surrounding the gap (Divenyi, 1979; Divenyi & Danner, 1977; Divenyi & Sachs, 1978). Hence, the cue integration we found in Experiments 1 and 2, and that found by Fitch et al. (1980), may indeed be unique to the perception of phonetic information. Nonetheless, the possibility remained open that the inverse relationship between temporal and spectral sensitivity *might* be reversed, for purely (and as yet unknown) psychoacoustic reasons, under certain stimulus and task conditions like those used in the "say"-"stay" tests. A third experiment was run to determine whether the "phonetic" or the "psychoacoustic" alternative would better explain trading relations and perceptual equivalence between phonetic cues.

This test required nonspeech control stimuli that maintained the crucial temporal and spectral properties of the "say"-"stay" stimuli, since the potential psychoacoustic effects might be dependent upon that particular array of physical properties. On the other hand, however, the stimuli had to be dissimilar enough from the "say"-"stay" stimuli that most naive listeners would fail to hear them as speech. We used "sine-wave analogues" of the "say"-"stay" continua for this experiment because they fit both criteria—they were essentially identical to the synthetic speech stimuli, except that their "formants" had bandwidths of 1 Hz. Another recent study employed sine-wave analogues of speech continua, and reported that most naive listeners heard the sine-wave stimuli as nonspeech sounds (e.g., beeps, chimes, slide-guitar notes, electronic tones). Only a few listeners spontaneously perceived them as distorted ("chime-like") speech. Identification tests revealed distinct differences in category boundaries, dependent on whether the sine-wave stimuli were perceived as speech or as nonspeech (Dorman, 1979; Bailey, Summerfield, & Dorman, Note 3). Because sine-wave analogues can be perceived either as speech or as nonspeech, which can affect categorization performance, we grouped our subjects according to their posttest reports of what the stimuli sounded like. The "phonetic" alternative predicted that the trading relation and perceptual equivalence between cues would occur only for subjects who heard the sine-wave analogues as "say" and "stay." In contradistinction, the "psychoacoustic" alternative predicted that those perceptual patterns would occur even for subjects who perceived the sine-wave stimuli as nonspeech.

Method

Subjects

Twenty-two naive listeners completed this experiment. Fifteen were Yale undergraduates and seven were enrolled at the University of Massachusetts/Amherst. All reported normal hearing in both ears.

Stimuli

Sine-wave analogues of the weak "day" (d) and strong "day" (D) speech syllables were made by synthesizing three simultaneous sine waves, using a software program developed for the PDP-11/45 at Haskins Laboratories.[10] In each of the sine-wave "day" analogues, the time-varying amplitude and center frequency characteristics of each of the synthetic speech formants was imitated by a frequency- and amplitude-modulated sine wave (see Figure 7). The sine-wave analogue of D will be termed "strong SW1 transition" (SW), and the analogue of d will be termed "weak SW1 transition" (sw).[11]

The nonspeech analogue of /s/ also had to differ enough from natural /s/ to be heard as a nonspeech sound by most listeners, and yet be similar enough to /s/ that it *could* be heard as speech.

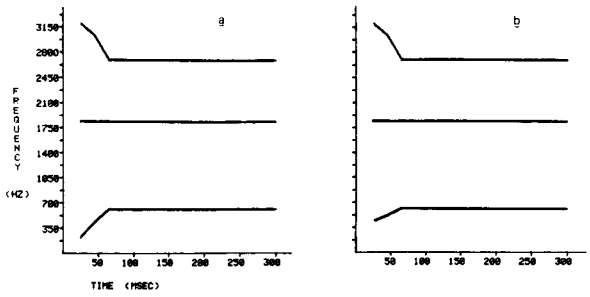

Figure 7. Schematic time-spectrum representations of the three-sine-wave analogues for the synthetic speech syllables used in Experiments 1 and 2: (a) weak "day" analogue: (b) strong "day" analogue.

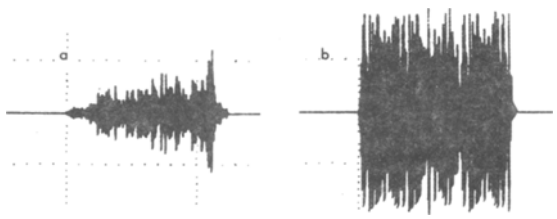

Figure 8. CRT display of digitized waveforms for (a) natural male /s/ (from "stay") used in the "say"-"stay" continua for Experiments 1 and 2, and (b) amplitude-reshaped "hiss" used in the sine-wave continua for Experiment 3.

We used a "hiss" created by changing the rise time and overall amplitude (but not the fall time) of the natural /s/ used in the synthetic "say"-"stay" continua (see Figure 8). The hiss (abbreviated h) had the same frequency and offset properties as the /s/ of Experiments 1 and 2, and thus met the requirements just outlined.

The SW and sw continua were constructed by inserting varying gap durations (in 8-msec steps, from 0 to 96 msec, as in Experiment 2) between h and the sine-wave analogue. Each continuum contained 14 stimuli (h[0]SW to h[96]SW, and h[0]sw to h[96]sw). which most of our naive listeners heard as bizarre electronic sounds or distorted nonspeech sounds, such as beeps, water drips, etc. Only about one-quarter of the subjects perceived them as "chime-like" utterances of "say" and "stay."

Procedure

An AXB identification procedure was used (cf. Bailey et al., 1978; Dorman, 1979), since the labels required for a standard forced-choice identification test might have encouraged subjects to perceive the stimuli as speech. On each trial, the second stimulus (X) had to be identified as being more similar either to the first stimulus (A) or the last stimulus (B). Categories A and B were fixed across trials; A was a sine-wave analogue to natural "say" (h[0]sw, the analogue for s[0]d), and B was the closest analogue to natural "stay" (h[96]SW, the analogue for s[96]D). Each of the 28 possible AXB trials was presented 10 times in a randomized test sequence.

After the AXB test and a subsequent 15-min break, the subjects took a three-way oddity test, which was designed exactly as in Experiment 1. It included "cooperating cues" (10 contrasts: h[0]sw-h[24]SW to h[72]sw-h[96]SW), "one cue" (13 contrasts: h[0]sw-h[0]SW to h[96]sw-h[96]SW), and "conflicting cues" comparisons (10 contrasts: h[0]SW-h[24]sw to h[72]SW-h[96]sw). Six judgments were obtained for each of the 33 possible stimulus contrasts (total = 198), and the test sequence was randomized across all comparison types.[12]

Group assignments. Sixteen subjects (all seven University of Massachusetts subjects and nine of the Yale subjects) were told before testing that the stimuli were computer sounds with two components—a "hiss" followed by a "chime-like" sound, with varying gap lengths between the components. The remaining six Yale subjects were told that the stimuli were distortions of "say" and "stay," and that they should listen for those words as they completed their tasks. We hoped this would induce a "speech perceptual set," and thereby allow us to assess the contribution of speech processing/perception to performance.

Subjects answered a posttest questionnaire on what the stimuli sounded like, and which stimulus properties they had attended to. The subjects were divided into five groups, according to their questionnaire responses. One subgroup included four subjects who claimed to have been guessing or changing their perceptual strategies from trial to trial; their performance was near chance, and appeared haphazard. Another group of three subjects perceived speech contrasts other than "say"-"stay" (i.e., "sleh"-"sreh," the French "un"-"rien," and two Greek words), including one subject who had been instructed to listen for "say" and "stay." The perceptual patterns for these first two groups will not be discussed further. Only the remaining three subgroups will be discussed in more detail. They were:

(1) Five subjects who heard "say" and "stay," either by instruction (three subjects) or spontaneously (two subjects), for even a portion of the test session. Three claimed to have occasionally listened for tone differences (the two "spontaneous" subjects) or for different water drips (one "instructed" subject, who used this strategy throughout most of the three-way oddity test), because at times they "lost touch with" the words. Therefore, this grouping provides a conservative test of the "phonetic" alternative.

(2) Five subjects who focused primarily on nonspeech *temporal* contrasts related to changes in gap duration, including differences in the gaps (two subjects) or spaces (one subject) between the hiss and sine waves, and overall length differences (two subjects).

(3) Five subjects who perceived nonspeech contrasts related to the SW vs. sw *spectral* difference. These subjects generally ignored the hiss and listened for contrasts between two different kinds of water drips (two subjects), two different pitches (one subject), the presence or absence of a ringing quality (one subject), or the presence or absence of electronic "waw" (one subject).

Groups 2 and 3 each included one of the subjects instructed to listen for "say" and "stay"; each reported that they could not hear the words, and had instead perceived a nonspeech contrast.[13] The three groups of subjects will be referred to as: (1) "say"-"stay" listeners, (2) "temporal" listeners, and (3) "spectral" listeners.

Results and Discussion

AXB Identification Test

Group comparisons. The AXB identification data for the three groups are shown in the upper panels of Figures 9-11. To determine whether the AXB differences among the three groups were statistically significant, a two-way ANOVA was performed for 3 groups by 2 continua. The data used in this test were "percent Category B responses" because the majority of AXB functions for the "spectral" group had no 50% crossovers within the range tested. The significant Groups by Continua interaction [$F(2,12) = 14.77$, $p < .01$] indicates that categorizations of the SW and sw continua contrasted substantially among the three groups of listeners. Only the "say"-"stay" group showed the sort of trading relation found in Experiments 1 and 2.

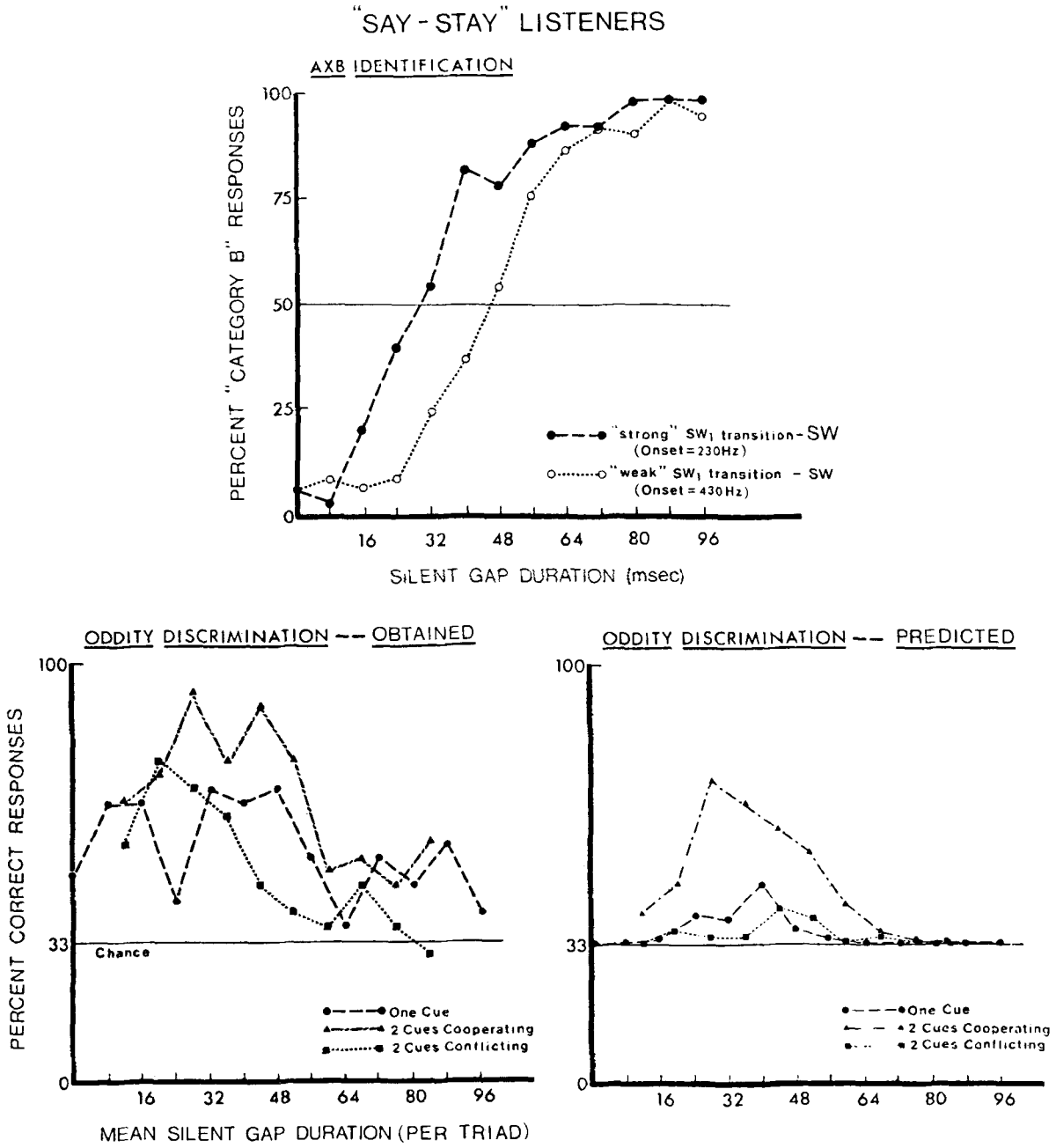

Figure 9. Sine-wave AXB identification functions (upper panel), and obtained (lower left) and predicted (lower right) functions for the three-way oddity test, "say"-"stay" listeneners. Experiment 3.

"Say"-"stay" listeners. The "say"-"stay" listeners showed a 17.8-msec trading relation (range = 7.6-29.6) between gap duration and the SW-sw spectral contrast (upper panel Figure 9), which was a significant boundary difference by a one-way ANOVA [$F(1,4) = 28.8$, $p < .01$]. The crossover value for the sw continuum was 45.8 msec (range = 40.0-52.0 msec); for the SW continuum, it was 27.9 msec (range = 22.4-38.4 msec). The trading relation magnitude and category boundaries were nearly the same as in Experiment 2, which included the same range of gap durations.

"Temporal" listeners. In contrast with the "say"-"stay" group, the "temporal" listeners (upper panel, Figure 10) failed to use the SW-sw contrasts consistently in their categorizations. They categorized stimuli according to duration changes showing a SW boundary at 34.3 msec (range = 26.0-44.8) and a sw boundary at 40.7 msec (range = 32.0-48.0). The 6.4-msec boundary difference (range = −2.65 to +8.0) was not significant.

"Spectral" listeners. In contrast with the other two sine-wave groups, the "spectral" listeners consistently

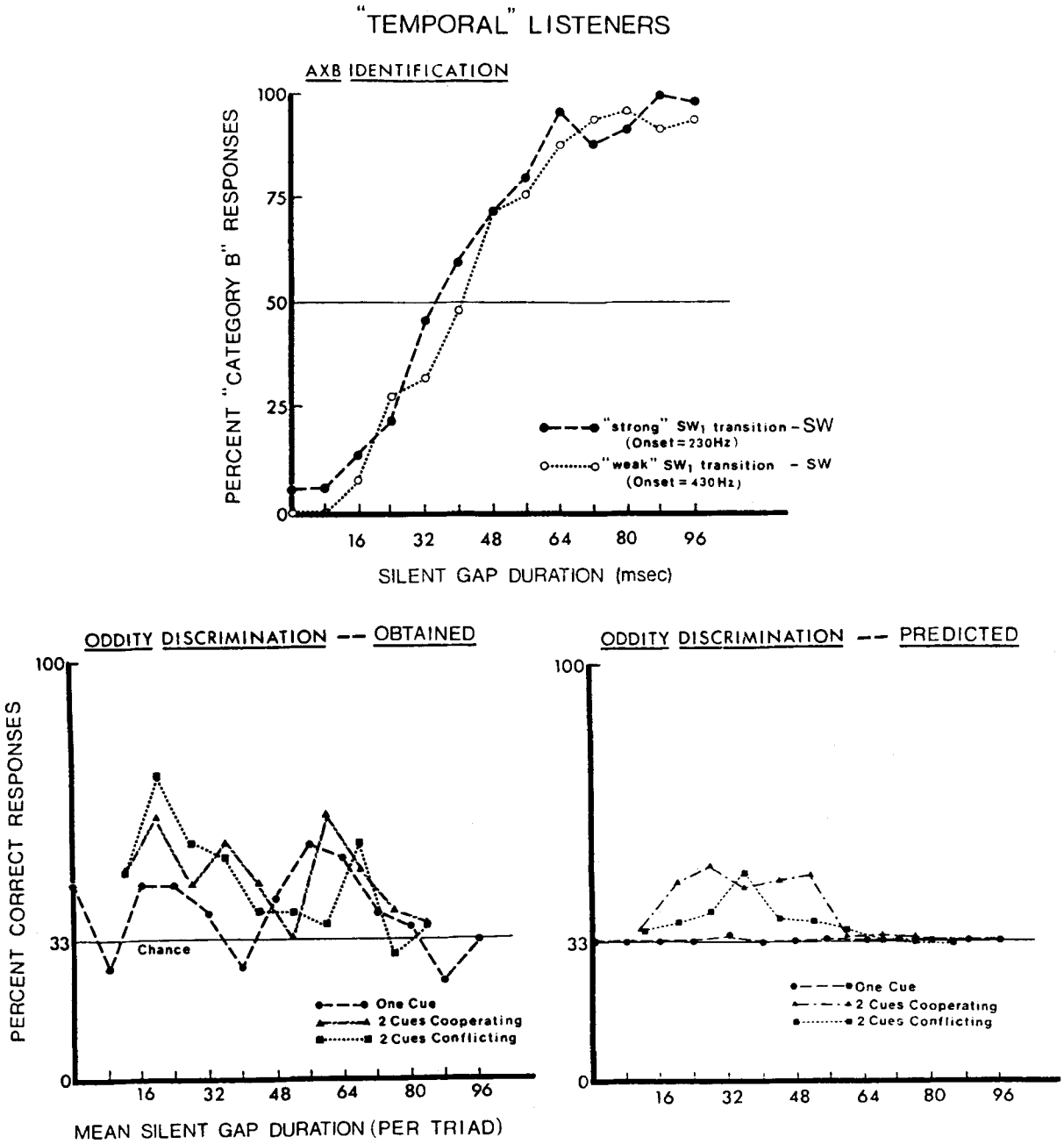

Figure 10. Sine-wave AXB identification functions (upper panel), and obtained (lower left) and predicted (lower right) functions for the three-way oddity test, "temporal" listeners, Experiment 3.

categorized stimuli by SW-sw differences (upper panel, Figure 11), according to their simple effects test for the Groups by Continua interaction [$F(1,12) = 60.1$, $p < .001$]. None of the SW stimuli were identified with A more often than chance, nor were the sw stimuli identified with B more often than chance, except for one token categorized as B 65% of the time. Thus, lengthening the gaps did not *completely* compensate for the SW-sw difference. The asymptote of the sw function at long gaps makes it unlikely that extending the range of the gap durations would have resulted in a complete trading relation for this group.

Three-Way Oddity Test

Group comparisons. There were also group differences in the pattern of three-way oddity discrimination (Figures 9-11) for both the obtained (lower left panels) and the predicted data (lower right panels). To compare discrimination performance among the three groups, mean "peak range" (16-48-msec gaps)

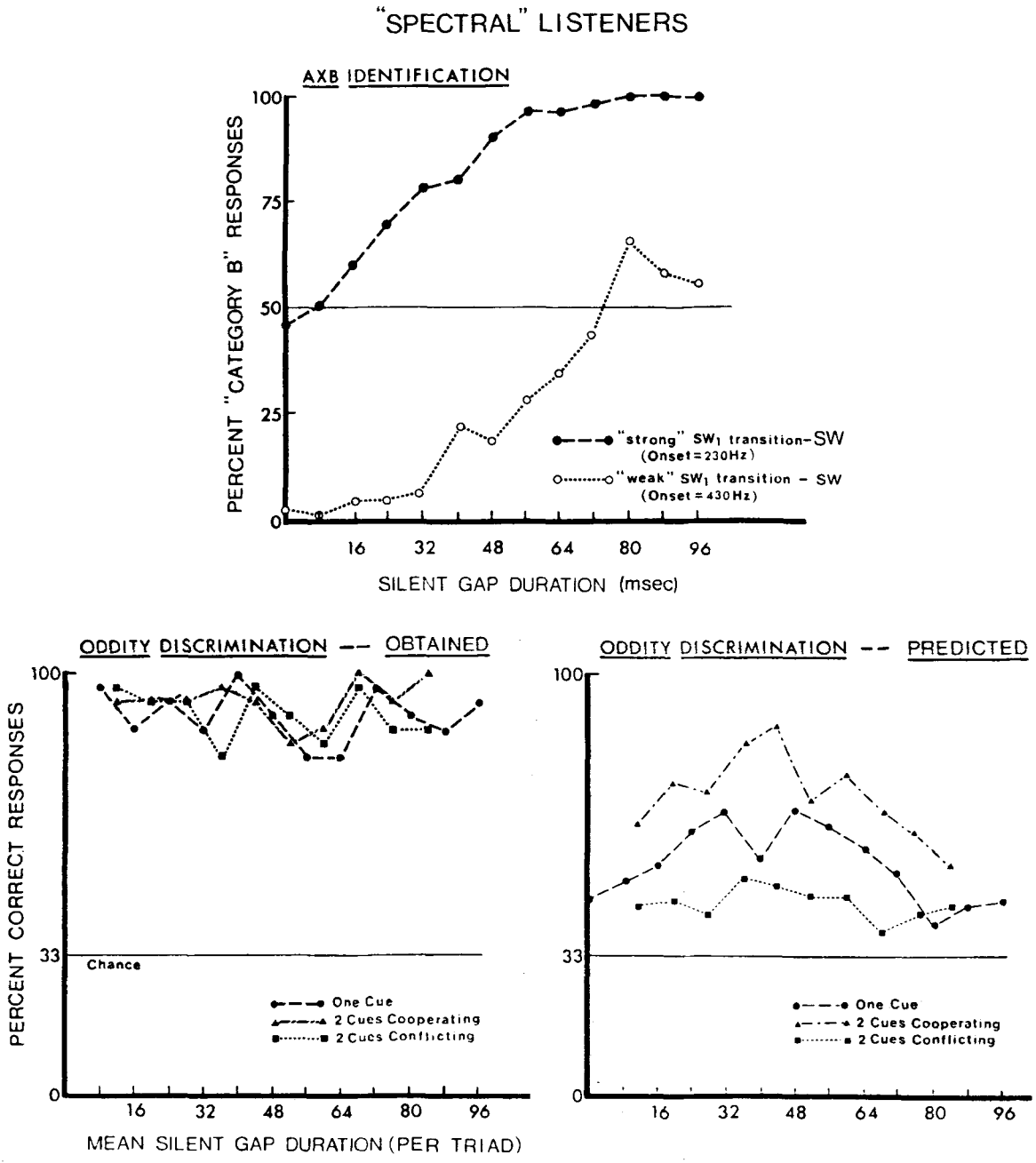

Figure 11. Sine-wave AXB identification functions (upper panel), and obtained (lower left) and predicted (lower right) functions for the three-way oddity test, "spectral" listeners, Experiment 3.

performances were calculated for the three comparison types. The range was extended beyond the 20-48-msec "peak range" used in Experiments 1 and 2 in order to include all the peaks of the "say"-"stay" group (lower left, Figure 9) and the peaks of the "temporal" group at 20 msec (lower left, Figure 10). An ANOVA was performed on these data for 3 groups by 3 comparison types. The order of "peak-range" performance levels differed significantly among the three groups [Groups by Comparison Types: $F(4,24) = 3.49$, $p < .025$].

"Say"-"stay" listeners. Only the "say"-"stay" group (lower left, Figure 9) showed the "phonetic" pattern found in the previous two experiments [simple effects test for the Groups by Comparison Types interaction: $F(2,24) = 9.4$, $p < .01$]. An ANOVA was run on their obtained data for 3 comparison types by 10 stimulus pairs (mean gap durations of between 12 and 88 msec). The Comparison Types effect [$F(2,8) = 33.35$, $p < .001$] supported the performance order: "cooperating cues" > "one cue" > "conflicting cues" (see Tukey pairwise contrasts, Table 1).[14] Peak-level per-

formance was significantly higher than trough-level performance across the three comparison types, according to the Stimulus Pairs effect [$F(9,36) = 4.36$, $p < .001$] (see Appendix F for individual patterns).

"Temporal" listeners. The obtained discrimination pattern for this group (lower left, Figure 10) suggested that the "cooperating cues"-"conflicting cues" distinction was moot for them; what mattered were noticeable temporal differences between stimuli. The Stimulus Pairs effect for this group [$F(9,36) = 3.08$, $p < .01$] indicates that there were two performance peaks; the level of the 20-msec peak appears higher than the one around 60-68 msec. The order of performance in the 20-msec "peak-range" (simple effects test for the Groups by Comparison Types interaction $F(2,24) = 4.86$, $p < .025$] was "conflicting cues" $\geq$ "cooperating cues" > "one cue" (see Tukey pairwise contrasts, Table 1). This pattern fits the "auditory" prediction that performance for two-cue comparisons would be better than for "one cue."

"Spectral" listeners. The "spectral" listeners, unlike either of the other two sine-wave groups, discriminated the SW-sw contrast nearly perfectly across the board (lower left, Figure 11). Performance was equally high across all three comparison types and at all gap durations.

Obtained vs. Predicted Differences

Because all three groups of listeners in Experiment 3 showed higher obtained than predicted performance, as was found in the previous two experiments, residual performance patterns (obtained minus predicted performance levels) among the three experiments will be discussed here. The Experiment 3 groups differed in their residual performance patterns (see details of analyses in Appendix F). Residual performance for the "temporal" listeners followed the order: "conflicting cues" > "cooperating cues" > "one cue."[15] The "spectral" listeners, in contrast, showed the residual performance pattern of "conflicting cues" > "one cue" > "cooperating cues," which also deviates from the "phonetic" pattern of Experiments 1 and 2. Only the "say"-"stay" listeners replicated the "phonetic" pattern of residual performance levels found in the previous experiments: "cooperating cues" > "one cue" > "conflicting cues." Thus, residual performance patterns were consistent across the "say"-"stay" listener groups from all three experiments. The "say"-"stay" residual performance pattern was distinctly different from the "temporal" and "spectral" patterns in Experiment 3.

These findings suggest that the "say"-"stay" residual performance was due to perception of the subcategorical differences as phonetic, rather than as purely auditory (nonphonetic), distinctions. Had the origin of the residual discriminability been purely auditory, the pattern should have followed the "auditory" prediction, or at least should have followed the "temporal" or "spectral" patterns. Although the residual discrimination performance of the "say"-"stay" listeners cannot be explained by *between*-category phonetic contrasts, it can be explained by *within*-category distinctions *that are nonetheless phonetic* (i.e., relevant to allophonic or articulatory variations). We note here that, for the "say"-"stay" listeners in all three experiments, the position of the obtained peaks was shifted toward the D (or SW) boundary. This shift, coupled with the consistent "day" categorization of D (whose F1 onset was like natural "day" and "stay") and the equivocal categorization of d (whose F1 onset was like natural "say"—see Stimuli, Experiment 1), suggests that residual discrimination was probably based on a distinction such as "clear /t/ closure" vs. "inexact (or weak) /t/ closure." Gap duration and F1 onset frequency differences apparently provided equivalent information about within-category, as well as between-category, phonetic distinctions.

Conclusions

The two most important points to be made about Experiment 3 are: (1) The identification and discrimination patterns of the "temporal" and "spectral" listeners differed substantially from the "say"-"stay" results of Experiments 1 and 2; and (2) the sine-wave "say"-"stay" results were essentially identical to the results of the two earlier experiments with synthetic speech. Only the subjects who perceived "say"-"stay" showed a trading relation and perceptual equivalence between the two acoustic cues. Thus, the trading relation and "phonetic" discrimination pattern appear to occur specifically with perception of *speech* contrasts. They are not attributable to general psychoacoustic sensitivities or interactions, since they did not appear in the two nonspeech groups. The nonspeech listeners focused on only one acoustic dimension (for the most part), and failed to integrate the two into a unitary percept. When the stimuli were perceived as speech, however, gap duration and spectral information were perceptually integrated in a manner that took account of their common origin in speech. Thus, the two cues are integral in speech perception, but separable in auditory perception (cf. Garner & Morton, 1969).

The residual performance patterns of the "temporal" and "spectral" groups were basically consistent with "auditory" predictions, whereas the residual performance pattern of the "say"-"stay" group was consistent with "phonetic" predictions (as were the residual patterns found in Experiments 1 and 2). The discrimination performance on "say"-"stay" contrasts that cannot be explained by phonetic *category* differences may nonetheless results from per-

ception of subcategorical stimulus differences as providing phonetic, rather than purely auditory, information.

SUMMARY AND GENERAL CONCLUSIONS

The three experiments present five major findings that bear on the integration of diverse acoustic properties in speech perception. First, there is a trading relation between the two primary acoustic consequences of the articulatory distinction between "say" and "stay." If unequivocal spectral information about the occurrence of a medial /t/ is provided, listeners hear "stay" when the duration of a silent gap between /s/ and the vocalic syllable minimally specifies a stop closure. However, when spectral information provides only equivocal information about an alveolar stop, listeners need stronger evidence for stop closure from another acoustic cue (e.g., longer closure gap) in order to perceive "stay."

Second, the two cues for the speech distinction, although from different *acoustic* dimensions, are perceptually equivalent. They converge upon a unitary phonetic dimension and provide qualitatively equivalent information about contrastive speech events. Stimulus tokens within a single phonetic category are quite difficult to distinguish perceptually, even though distinct along two different acoustic dimensions ("conflicting cues" comparisons). However, discrimination is comparatively easy when the parameter values for the same two acoustic dimensions are such that the stimuli being discriminated are in different phonetic categories ("cooperating cues").

Third, the qualitative equivalence of the two cues within a single phonetic dimension reflects equivalence in sensitivity to those properties of the speech stimuli. When subjects listen to "say"-"stay" stimuli, the "phonetic" discrimination pattern emerges even under conditions designed to reduce memory demands and eliminate an experimentally induced "set" to categorize stimuli before discriminating them.

Fourth, trading relations and perceptual equivalence between cues derive specifically from the integrated perception of multiple acoustic properties as *phonetic information*, and not from psychoacoustic factors. Those perceptual patterns do not occur when listeners perceive the acoustic variations as nonspeech contrasts. Experiment 3 implies that the perceptual integration of diverse acoustic information is determined by what the listener perceives the stimuli to be, much more than it is by raw stimulus characteristics and/or their interactions with basic properties of the auditory system. Several other recent speech perception findings provide converging support for the notion that performance patterns are determined more by the type of information focused upon than they are by the absolute physical properties of the stimuli. Changes in identification functions occur not only for sine-wave speech continua, dependent on whether the stimuli are heard as speech or nonspeech (Dorman, 1979; Bailey et al., Note 3), but also for speech continua, dependent on the specific phonetic contrast subjects listen for (Carden, Levitt, Jusczyk, & Walley, 1981). Moreover, discrimination performance for speech continua differs substantially, depending on whether listeners are focusing on phonetic category information or ignoring phonetic categories to focus on purely acoustic properties (Repp, Note 4).

Fifth, the *pattern* of residual performance on "say"-"stay" discriminations suggests that even within a phonetic category, acoustic variations in the two cues are treated perceptually as if they provide phonetic, not simply auditory, information. To our knowledge, this is the first time it has been possible to distinguish empirically between the contributions of auditory and phonetic perception to speech discrimination performance levels that cannot be explained by phonetic *category* differences. For this reason, and also because the differences among the comparison types were small (see Appendices C, E, and F), the effect needs replication. We suggest that the residual discriminability of speech contrasts should most likely reflect phonetic, not auditory, perceptual contributions whenever the acoustic characteristics of the stimuli fall within the range of natural speech variation, and whenever listeners perceive the stimulus properties as information about speech contrasts.

The pattern of perceptual integration of the two cues by the three groups of "say"-"stay" listeners paralleled the pattern of acoustic and articulatory qualities found in natural "say" and "stay" utterances (cf. Experiment 1). That is, "stay" differs from "say" in that only the former word involves a *complete* linguoalveolar closure, which results in a longer closure silence and lower F1 onset frequency than found in the latter word. Perceptual integration of the silence and F1-onset cues indicated that listeners had acted as though both cues provided comparable information about whether a complete linguoalveolar closure had occurred. This pattern of perception-production similarities leads us to agree with the conclusion of Fitch et al. (1980) that trading relations and perceptual equivalence indicate that phonetic perception takes account of the common articulatory origin of diverse cues for a given speech contrast. The perception-production commonalities implied by our results and those of Fitch et al. may suggest that, when listeners attend to the phonetic properties of speech stimuli, they are perceiving articulatory information provided by the acoustic waveform. An excellent discussion of this possibility can be found in Summerfield (1978). Further corroboration for this hypothesis comes from research on the parallel effects of phonetic context on perception and produc-

tion. A variety of context effects indicate that phonetic perception takes account of articulatory consequences —e.g., context-dependent shifts in patterns of perception for consonant contrasts parallel the effects of context on the corresponding articulatory gestures (e.g., Mann, 1980; Mann & Repp, 1980, 1981; Miller & Liberman, 1979).

The "say"-"stay" results reported in this paper appear robust, and reflect the perceptual integrity of the multiple acoustic consequences of articulatory gestures *as phonetic information*. But the possibility that the trading relation/perceptual equivalence pattern may be unique to speech needs further investigation. To learn whether that perceptual pattern is uniquely human, the responses of animals to speech and nonspeech contrasts conveyed by multiple physical cues (even nonauditory) might be studied (cf. Kuhl, 1978; Liberman & Pisoni, 1977; but also compare Kuhl & Miller, 1975, 1978; Morse & Snowdon, 1975; Waters & Wilson, 1976). For example, a recent report of discrimination among natural leaf categories (oak vs. nonoak) by pigeons (Cerella, 1979) suggested that the animals may have treated several dimensions of contrast among leaf outlines as equivalent (e.g., smooth vs. serrated edge, shallow vs. deep notches between lobes, etc.). Further research would be necessary, however, to determine the completeness of the pigeon's perceptual integration of leaf-outline dimensions—that is, to determine whether they would show trading relations and perceptual equivalence among the diverse features.

Also, to assess whether perceptual equivalence is uniquely characteristic of *speech* perception, or whether it may be a more general quality in perception of complex acoustic information for naturally occurring contrastive events, the perceptual integration of multiple cues might be explored for familiar *non*speech events that are rich in dynamic acoustic information. For example, there are probably spectral as well as temporal contrasts between the acoustic products of plucking vs. bowing actions on a violin string (Schelleng, 1973), or between the acoustic consequences of hard vs. soft attack in the playing of piano notes (Weyer, 1976, 1976/1977). Contrastive nonspeech properties such as these might also be perceptually integrated, but it is not clear a priori whether such integration would imply qualitative equivalence among the diverse acoustic cues. Answers to questions about multiple acoustic properties of natural nonspeech events, and about perceptual integration of those (possible) properties, still await empirical exploration.

The strength of the current findings implies that perceptual equivalence among multiple cues for a given phonemic contrast is a key aspect of adult speech perception. Developmental research on trading relations and perceptual equivalence in speech perception may aid in understanding the interplay of maturation, perceptual experience, and articulatory competence in the ontogeny of the general ability to perceive phonetically relevant characteristics of human speech (again, see discussions by Kuhl, 1978; Liberman & Pisoni, 1977). Such research would help in appraising whether certain acoustic contrasts elicit innate or biologically determined perceptual responses, while others gain an effect on perception primarily through receptive and productive language experience. For example, the voiced-voiceless distinction can be cued for adults by contrasts in either VOT or F1 onset frequency (e.g., Lisker, 1975). However, though young children and even very young infants respond to VOT categories much like adults (cf. Jusczyk, 1981; Kuhl, 1978), children do not respond strongly to F1 onset distinctions until they are around 5 years old (Simon & Fourcin, 1978), at which time they may show phonetic trading relations that are smaller in magnitude than the corresponding adult trading relations (Robson, Morrongiello, Best, & Clifton, Note 5). These facts may suggest that the development of trading relations in phonetic perception is dependent on fairly extensive language experience, and would begin to show up only during the preschool-kindergarten years. On the other hand, a recent study of 6-month-olds' perception of the /slIt/-/splIt/ contrast, cued either by the artificial introduction of silence alone or by natural silence plus /p/ bursts and transitions, suggests that even young infants might show some evidence of a phonetic trading relation (Morse, Eilers, & Gavin, Note 6).

REFERENCE NOTES

1. Lisker, L. *Rapid vs. rabid: A catelogue of acoustic features that may cue the distinction* (Status Report on Speech Research, SR-54, 127-132). New Haven, Conn: Haskins Laboratories, 1978.
2. Pisoni, D. B. *On the nature of categorical perception of speech sounds* (Status Report on Speech Research, SR-27, 209-210). New Haven, Conn: Haskins Laboratories, 1971.
3. Bailey, P. J., Summerfield, A. Q., & Dorman, M. F. *On the identification of sinewave analogues of certain nonspeech sounds* (Status Report on Speech Research, SR-51/52, 1-25). New Haven, Conn: Haskins Laboratories, 1977.
4. Repp, B. H. *Two strategies in fricative discrimination*. Manuscript submitted for publication, 1980.
5. Robson, R., Morrongiello, B., Best, C. T., & Clifton, R. K. *Trading relations in the perception of speech by five-year-olds and adults*. Paper presented at the meeting of the Eastern Psychological Association, Hartford, Connecticut, April 1980.
6. Morse, P. A., Eilers, R., & Gavin, W. *Exploring the perception of the "sound of silence" in early infancy*. Paper presented at the second meeting of the International Conference on Infant Studies, New Haven, Connecticut, April 1980.
7. Pastore, R. E. Personal communication, 1980.
8. Remez, R. E. Personal communication, 1979.
9. Repp, B. H. Personal communication, 1979.
10. Summerfield, A. Q. Personal communication, 1979.

REFERENCES

Abramson, A. S., & Lisker, L. Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Commins

(Ed.), *Proceedings of the 5th Congress of International Acoustics.* Liège, Belgium: Thone, 1965.

BAILEY, P. J., & SUMMERFIELD, A. Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance,* 1980, **6**, 536-563.

CARDEN, G., LEVITT, A. G., JUSCZYK, P. W., & WALLEY, A. Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics,* 1981, **29**, 26-36.

CERELLA, J. Visual classes and natural categories in the pigeon. *Journal of Experimental Psychology: Human Perception and Performance,* 1979, **5**, 68-77.

CREELMAN, C. D., & MACMILLAN, N. A. Auditory phase and frequency discrimination: A comparison of nine methods. *Journal of Experimental Psychology: Human Perception and Performance,* 1979, **5**, 146-156.

CROWDER, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior,* 1971, **10**, 587-596.

CROWDER, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology,* 1973, **1**, 14-24.

CROWDER, R. G., & MORTON, J. Pre-categorical acoustic storage (PAS). *Perception & Psychophysics,* 1969, **5**, 365-373.

CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics,* 1974, **16**, 564-570.

CUTTING, J. E., & ROSNER, B. S. Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics,* 1976, **20**, 87-88.

CUTTING, J. E., ROSNER, B. S., & FOARD, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology,* 1976, **28**, 361-378.

DARWIN, C. J., & BADDELEY, A. D. Acoustic memory and the perception of speech. *Cognitive Psychology,* 1974, **6**, 41-60.

DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America,* 1955, **27**, 769-773.

DIVENYI, P. L. Some psychoacoustic factors in phonetic analysis. *Proceedings of the 9th International Congress of Phonetic Sciences.* Copenhagen: Stougaard Jensen, 1979.

DIVENYI, P. L., & DANNER, W. F. Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics,* 1977, **21**, 125-142.

DIVENYI, P. L., & SACHS, R. M. Discrimination of time intervals bounded by tone bursts. *Perception & Psychophysics,* 1978, **24**, 429-436.

DORMAN, M. F. On the identification of sineway analogues of CV syllables. *Proceedings of the 9th International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: Stougaard Jensen, 1979.

DORMAN, M. F., RAPHAEL, L. J., & ISENBERG, D. Acoustic cues for the fricative/affricate contrast in word-final position. *Journal of Phonetics,* 1980, **8**, 397-405.

DORMAN, M. F., RAPHAEL, L. J., & LIBERMAN, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America,* 1979, **65**, 1518-1532.

DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics,* 1977, **22**, 109-122.

FANT, C. G. M. Descriptive analysis of the acoustic aspects of speech. *Logos,* 1962, **5**, 3-17.

FITCH, H. L., HALWES, T. G., ERICKSON, D. M., & LIBERMAN, A. M. Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics,* 1980, **27**, 343-350.

FUJISAKI, H., & KAWASHIMA, T. On the modes and mechanisms of speech perception. *Annual Report of the Audio Engineering Society,* 1969, **28**, 67-73.

GARNER, W. R., & MORTON, J. Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin,* 1969, **72**, 233-259.

GREEN, D. M. *An introduction to hearing.* New York: Wiley, 1976.

HAGGARD, M. P., AMBLER, S., & CALLOW, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America,* 1970, **47**, 613-617.

HARRIS, K. S., HOFFMAN, H. S., LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. *Journal of the Acoustical Society of America,* 1958, **30**, 122-126.

HOFFMAN, H. S. Study of some cues in the perception of the voiced stop consonants. *Journal of the Acoustical Society of America,* 1958, **30**, 1035-1041.

JUSCZYK, P. W. Infant speech perception: A critical appraisal. In P. D. Eimas & J. A. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, N.J: Erlbaum, 1981.

KAPLAN, H. L., MACMILLAN, N. A., & CREELMAN, C. D. Methods and designs: Tables of d′ for variable-standard discrimination paradigms. *Behavior Research Methods & Instrumentation,* 1978, **10**, 796-813.

KUHL, P. K. Predispositions for perception of speech-sound categories: A species-specific phenomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities—Early behavioral assessment.* Baltimore, Md: University Park Press, 1978.

KUHL, P. K., & MILLER, J. D. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science,* 1975, **190**, 69-72.

KUHL, P. K., & MILLER, J. D. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America,* 1978, **63**, 905-917.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review,* 1967, **74**, 430-460.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology,* 1961, **61**, 379-388.

LIBERMAN, A. M., & PISONI, D. B. Evidence for a special speech-perceiving subsystem in the human. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals.* Berlin: Dahlem Konferenzen, 1977.

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In H. L. Teuber (Ed.), *Handbook of sensory physiology* (Vol. 8) *Perception.* Berlin: Springer-Verlag, 1978.

LISKER, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America,* 1975, **57**, 1547-1551.

LISKER, L., LIBERMAN, A. M., ERICKSON, D. M., DECHOVITZ, D., & MANDLER, R. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech,* 1977, **20**, 209-216.

MACKAIN, K. S., BEST, C. T., & STRANGE, W. Native language effects on liquid perception. *Journal of the Acoustical Society of America,* 1980, **67** (Suppl.), S27. (Abstract K12)

MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, C. D. The psychophysics of categorical perception. *Psychological Review,* 1977, **84**, 452-471.

MANN, V. A. Influence of preceding liquids on stop-consonant perception. *Journal of the Acoustical Society of America,* 1980, **67** (Suppl.). S99, (Abstract QQ1)

MANN, V. A., & REPP, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics,* 1980, **28**, 213-228.

MANN, V. A., & REPP, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America,* 1981, **69**, 548-558.

MCGOVERN, K., & STRANGE, W. The perception of /r/ and /l/

in syllable-initial and syllable-final position. *Perception & Psychophysics*, 1977, **21**, 162-170.

MILLER, J. A., & LIBERMAN, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 1979, **25**, 457-465.

MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-417.

MORSE, P. A., & SNOWDON, C. T. An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, 1975, **17**, 9-16.

PASTORE, R. E., AHROON, W. A., BAFFUTO, K. J., FRIEDMAN, C., PULEO, J. S., & FINK, E. A. Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **4**, 686-696.

PASTORE, R. E., FRIEDMAN, C. J., & BAFFUTO, K. J. A comparative evaluation of the AX and two ABX procedures. *Journal of the Acoustical Society of America*, 1976, **60** (Suppl.), S120. (Abstract)

PISONI, D. B. Auditory short term memory and vowel perception. *Memory & Cognition*, 1975, **3**, 7-18.

PISONI, D. B. Identification and discrimination of the relative onset time of two-component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, **61**, 1352-1361.

PISONI, D. B., & LAZARUS, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, **55**, 328-333.

PISONI, D. B., & TASH, J. Reaction times to comparisons within and across phonetic boundaries. *Perception & Psychophysics*, 1974, **15**, 285-290.

REPP, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 1979, **22**, 173-189.

REPP, B. H., HEALY, A. F., & CROWDER, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of the Acoustical Society of America*, 1979, **5**, 129-145.

REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. Perceptual integration of temporal cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 621-637.

REPP, B. H., & MANN, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 1980, **67** (Suppl.), S100. (Abstract QQ8)

SCHELLENG, J. C. The bowed string and the player. *Journal of the Acoustical Society of America*, 1973, **53**, 26-41.

SIMON, C., & FOURCIN, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, **63**, 925-935.

STEVENS, K. N. The role of rapid spectrum changes in the production and perception of speech. In L. L. Hammerich & R. Jakobson (Eds.), *Form and substance: Festschrift for Eli Fischer-Jørgensen*. Copenhagen: Akademisk Forlag, 1971.

STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In L. Pinson & D. Denes (Eds.), *Human communication: A unified view*. Cambridge, Mass: M.I.T. Press, 1974.

STUDDERT-KENNEDY, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

SUMMERFIELD, A. Q. Perceptual learning and phonetic perception. *Interrelations of the communicative senses. Proceedings of the NSF Conference at Asilomar*. Washington, D.C: NSF Publication, 1978.

SUMMERFIELD, A. Q., & HAGGARD, M. P. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, **62**, 435-448.

WATERS, R. S., & WILSON, W. A. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, 1976, **19**, 285-289.

WEYER, R. D. Time-frequency-structures in the attack transients of piano and harpsichord sounds—1. *Acustica*, 1976, **35**, 233-252.

WEYER, R. D. Time-varying amplitude-frequency-structures in the attack transients of piano and harpsichord sounds—II. *Acustica*, 1976-77, **36**, 241-252.

NOTES

1. The use of these words might have introduced a lexical bias (frequency counts for "say" and "stay" are not equal), although the effect of that bias on performance could not have interfered with our test results, since it should have involved both "say"-"stay" continua equally. Moreover, although neither /sɛ/ nor /stɛ/ is a word, and neither is phonologically permissible in American English because of the final lax vowel, our pilot work with that contrast showed a trading relation between silence and F1 onset.

2. The /t/ in "stay" is unaspirated ([t] as opposed to [t^h]). When an unaspirated /t/ follows an /s/ in American English, as in "stay," it is identified in context as "t" ("stay"). However, if the /s/ is removed from "stay" (leaving a vocalic syllable similar to our synthetic stimuli with low F1 onset frequency), the isolated unaspirated [tei] is identified as "day" rather than "tay" because word-initial voiceless stops in spoken American English are nearly always aspirated (i.e., [t^hei]).

3. Pilot testing had indicated that subjects would need an additional 24 msec of silent gap (approximately) to begin hearing "stay" for the d continuum, relative to the D continuum.

4. Although 12-18 judgments per comparison are typically obtained in discrimination tests, we used only six judgments per comparison in order to compare identification and discrimination results gathered within a single moderate-length test session, hoping thereby to minimize changes in response criteria, attention level, etc. In addition, since we were most interested in the subjects' "natural" or "normal" perception of the stimuli, we kept presentations per stimulus at the smallest number likely to yield reliable response functions. These concerns also related to our use of 10, rather than the usual 20, judgments per token in the identification test. This situation thus provides a conservative test of our hypotheses, because the potential for response variability was higher than usual.

5. The Stimulus Pairs simple effect for the "conflicting cues" comparisons did not support a performance peak near the category boundary: instead, it indicated lower discrimination performance for the extreme endpoints from the two continua (s[0]D vs. s[24]d, s[8]D vs. s[32]d, and s[112]D vs. s[136]d) than for all other comparisons.

6. Although the oddity task is commonly assumed to have higher memory demands than other discrimination procedures used in auditory research, and hence to yield the lowest performance level, no direct comparisons of oddity vs. other tasks are known (Pastore, Note 7; Remez, Note 8; Repp, Note 9), with one recent exception. MacKain, Best, and Strange (1980) found slightly higher above-chance performance for an AXB than an oddity task, using a synthetic /rak/-/lak/ continuum. Their finding corroborates the common intuitions about the relative difficulty (memory demands) of the oddity task.

7. We refer here to variable-standard rather than fixed-standard discrimination designs. Although fixed-standard designs yield higher performance than variable-standard designs, according both to theoretical models (SDT analysis: MacMillan et al., 1977) and to empirical work with nonspeech stimuli (Creelman & MacMillan, 1979), those task differences may be very small for discrimination of speech (Repp, Note 4). Furthermore, we were limited to a variable-standard design because the three-way discrimination design dictated that the magnitude of within-pair

differences be fixed and that stimulus comparisons should cover the range of the two continua.

8. Higher d′ values have been obtained through AXB than 4IAX tests for frequency (Creelman & MacMillan, 1979) and intensity discriminations (Pastore, Friedman, & Baffuto, 1976), but 4IAX tasks have obtained higher d′ values than ABX for phase (Creelman & MacMillan, 1979) and speech discriminations (e.g., Pisoni, 1975; Pisoni & Lazarus, 1974; Pisoni, Note 2). These discrepancies may indicate that two 4IAX observer strategies are logically possible, one of which is more sensitive; which strategy listeners will adopt appears to be unpredictable (Creelman & MacMillan, 1979; MacMillan et al., 1977).

9. True d′ values from 2IAX data are equal to or slightly higher than ABX or 4IAX d′ values for frequency discriminations, and may be even more improved for phase discriminations, relative to the latter two paradigms. Compare variable-standard designs in Figure 4 with the interrelation of Figure 1 and Table 2 (Creelman & MacMillan, 1979, pp. 151-152), which show the same pattern of ABX-4IAX performance level differences found in speech discrimination tests (e.g., Pisoni, 1975, Note 2).

10. Thanks are extended to Philip Rubin for his helpful modifications of the sine-wave synthesis program originally written by Rod McGuire at Haskins Laboratories.

11. The two sine-wave syllables lacked the final diphthongization from /e/ to /i/ because the additional frequency modulation seemed to make their speech-like qualities too obvious, and were physically more analogous to /dɛ/ than to /dei/. However, the offset of the three sine waves was such that they sounded like "day" rather than "deh" to most listeners who heard them as speech.

12. It was especially important in Experiment 3 to keep testing time and stimulus repetition low in order to minimize the possibility that nonspeech listeners might spontaneously begin to hear the stimuli as speech after prolonged exposure, and consequently shift their perceptual behavior.

13. As these descriptions of the subgroups suggest, there was great variation in individual perceptions of these stimuli, and attempts to impose perceptual characterizations did not work consistently. Both the individual variation and the inconsistent response to "perceptual instruction" seem to be characteristic of tests with sine-wave speech analogues, inasmuch as they have been noted before (e.g., Bailey et al., Note 3; Summerfield, Note 10).

14. The functions appeared a bit rough because of the small number of subjects, the difficulty that three of the subjects had in keeping "tuned" to "say" and "stay" throughout the test, and especially because of the subject who reported listening for different water drips through most of the three-way oddity test; all of these factors make interpretation of any visual irregularities difficult. For example, there is an apparent bimodality in the "cooperating cues" peak, but the "dip" between the two highest points represents a total drop of only 4-5 correct responses. Furthermore, the "double-peak" pattern was shown by only two subjects, each of whom showed a "dip" of two responses in magnitude. One of these subjects was a "spontaneous" "say"-"stay" listener, and had the noisiest data of this group.

15. The "temporal" group's obtained boundary-related peak (20 msec) was shifted from the predicted peak (36 msec), and their function showed a second unexpected obtained peak (60-68 msec), suggesting that the oddity and AXB tests may have tapped different perceptual processes—they may have been using three categories, rather than two. The shallow slopes of the AXB functions suggest that the two prototype categories (A and B) might not have been the most appropriate for these listeners. The 20-msec peak hints that one perceived contrast may have been "contiguity between hiss and sine wave" vs. "delay between hiss and sine wave," which would be consistent with the general psychoacoustic boundary at 20 msec for detection of temporal differences between components of two-part signals (e.g., Miller et al., 1976; Pastore et al., 1977; Pisoni, 1977). The later peak (60-68 msec) may distinguish "a two-component signal" vs. "two separate signals." Although the pattern of performance around this second peak (56-88 msec) indicated that in that range the "temporal" listeners must have responded to SW-sw differences, the drop in performance between 72 and 96 msec makes it unlikely that they had merely focused on the spectral contrasts at longer gap durations. It is more likely that they were still attending to temporal information, and that sensitivity to gap changes in that range may have been differentially affected by the SW vs. sw onset spectra (and/or that discrimination as a function of gap duration is nonmonotonic).

APPENDIX A

In both stimuli, there was a 25-msec rise time and a 50-msec fall time; the parameter values for the amplitudes of F2 and F3 throughout the stimuli were 4/5 that of F1. The F0 contour began at 120 Hz and remained at that frequency for 240 msec, after which it fell linearly to 90 Hz during the final 50 msec. F3 began with a linear 40-msec transition from 3,196 to 2,694 Hz (somewhat exaggerated with respect to natural stimuli), and remained at the latter frequency for 130 msec, after which it rose linearly to 3,029 Hz during the ensuing 70 msec. F2 remained at 1,840 Hz for the initial 150 msec, and rose linearly to 2,298 Hz during the following 90 msec. F2 and F3 remained at their last-named frequencies during the final 50 msec. F1 reached 611 Hz at the end of the initial 40-msec transitions, remained at that frequency for 110 msec, then fell linearly to 304 Hz during the final 90 msec.

APPENDIX B

We used the following formula to predict each subject's performance for each of the 44 stimulus comparisons tested (from McGovern & Strange, 1977):

$$\text{Pcorr} = [1 + 2(\text{Pa} - \text{Pb})2]/3,$$

where Pcorr is predicted probability of correct responses for a given stimulus comparison, Pa is obtained proportion of "stay" responses to stimulus a, and Pb is the observed proportion of "stay" responses to stimulus b in the comparison. Chance level responding was 33%.

APPENDIX C

Predicted Data

To determine whether the obtained pattern derived from the phonetic category judgments, an ANOVA was performed on the predicted data for 3 comparison types by 15 stimulus pairs. Figure 4 shows that the results from this ANOVA were essentially the same as for the obtained data [Stimulus Pairs, $F(14,196) = 10.72$, $p \ll .001$; Comparison Types, $F(2,28) = 29.60$, $p \ll .001$—see Table 1 for Tukey pairwise contrasts]. The Comparison Types by Stimulus Pairs interaction [$F(28,392) = 33.13$, $p \ll .001$] was also highly similar to the obtained results [simple effects tests: "cooperating cues," $F(14,588) = 17.76$, $p \ll .001$; "one cue," $F(14,588) = 4.36$, $p < .001$; "conflicting cues," $F(14,588) = .22$, n.s.].

Obtained vs. Predicted Differences

Although the obtained and predicted data showed similar *patterns*, Figure 4 suggests that obtained discrimination

was slightly better than predicted, particularly beyond the immediate neighborhood of the performance "peaks" that had been expected for clear between-category comparisons. This observation was supported by ANOVAs for each of the three comparison types, which crossed the 2 functions (obtained vs. predicted) with 15 stimulus pairs (12-128-msec range). The three significant functions effects showed that obtained performance was better than predicted for "cooperating cues" [$F(1,14) = 8.82$, $p < .025$], "one cue" $F(1,14) = 12.22$, $p < .005$], and "conflicting cues" comparisons [$F(1,14) = 19.39$, $p < .001$]. Significant Functions by Stimulus Pairs interactions for "one cue" [$F(17,238) = 2.54$, $p < .001$] and "cooperating cues" comparisons [$F(17,238) = 4.56$, $p \ll .001$] revealed that obtained performance was higher than predicted *only* for comparisons distant from the between-category performance peak by 16-24 msec or more; performance on between-category comparisons was fully determined by the identification functions. These interactions also indicated that the obtained peaks were shifted toward the D boundary (highest performance levels in the "cooperating cues" and "one cue" comparisons at approximately 36-40 msec), relative to the predicted peak positions, which were at the average of the crossover boundary values for the two continua (highest performance levels at approximately 44-48 msec).

Because differences between obtained and predicted performance were not found for between-category comparisons, the pattern of *within*-category differences was examined. For this purpose, within-category refers to all comparisons whose predicted level was 39% or less ($\leqslant$ 6% above chance), since in each of those comparisons the stimuli being compared never differed by more than 10% in their category assignments in the identification test. This "within-category" range included "conflicting cues" comparisons with mean gap durations of between 12 and 44 msec and between 60 and 128 msec, "one cue" comparisons with gaps of between 0 and 16 msec and between 56 and 136 msec, and "cooperating cues" comparisons with mean gap durations of between 76 and 128 msec. The pattern of mean above-chance discrimination for these within-category comparisons was "cooperating cues" (mean correct = 49.2%) > "one cue" (48.2%) > "conflicting cues" (47.0%). If we consider only the within-category obtained performance that was unaccounted for by phonetic category identifications (residual performance = obtained mean minus predicted mean), the pattern remains: "cooperating cues" (residual performance = 14.0%) > "one cue" (13.0%) > "conflicting cues" (11.7%).

Individual Patterns

Consistency of the "phonetic" perceptual equivalence pattern in individuals was also examined, for stimulus comparisons with average gap durations near the two identification boundaries (between 20 and 48 msec), henceforth termed the "peak range." This analysis was based on the mean percent correct obtained responses for "one cue" comparisons between s[24]D-s[24]d and s[48]D-s[48]d, "cooperating cues" comparisons between s[32]D-s[8]d and s[56]D-s[32]d, and "conflicting cues" comparisons between s[32]d-s[8]D and s[56]d-s[32]D. Ten of the 15 subjects showed the order "cooperating cues" > "one cue" > "conflicting cues," a proportion (67%) significantly better than chance by binomial test ($p < .002$). For the other five subjects, "one cue" performance was equal to or lower than "conflicting cues" performance. However, the key criterion for "phonetic" perceptual equivalence is that "peak-range" performance be better for "cooperating cues" than for "conflicting cues," since both involve contrast on two acoustic dimensions. This latter criterion was met by all 15 subjects, a proportion that is far beyond chance expectation by binomial test, $p \ll .001$. Two one-way ANOVAs performed on these obtained and predicted "peak-range" means, for the two comparison types, supported the "phonetic" predictions even more strongly than did the overall group ANOVAs. For the obtained "peak-range" data, the comparison types effect was significant [$F(2,28) = 36.59$, $p \ll .001$], as it was for the predicted "peak-range" data [$F(2,28) = 81.51$, $p < .001$] (see Table 1 for Tukey pairwise contrasts).

APPENDIX D

To determine true d′ values for obtained and predicted data in the 2IAX test, it was necessary to calculate P[H] (probability of a hit) and P[FA] (probability of a false alarm). These were derived by the following formulas:

$$P[H] = P(\text{"S"}/S)$$

for the obtained hit rate, where P is probability or proportion, "S" is "same" responses, and /S is "given a physically identical AX pair";

$$P[H] = P(\text{"stay"}/A)^2 + P(\text{"stay"}/X)^2$$

for the predicted hit rate, where P("stay"/A) is proportion of "stay" responses to stimulus A in the identification test, and likewise for stimulus X in P("stay"/X);

$$P[FA] = P(\text{"D"}/S)$$

for the obtained false alarm rate, where "D" is "different" responses; and

$$P[FA] = P(\text{"stay"}/A) \cdot P(\text{"say"}/A) + P(\text{"stay"}/X) \cdot P(\text{"say"}/X)$$

for predicted false alarm rate. These values of P[H] and P[FA] were used to look up true d′ values in the table provided by Kaplan et al. (1978). Because this table (and the computational formula used to derive the tabled values) does not allow for P[H] and P[FA] of either 0.00 or 1.00, we substituted 0.01 and 0.99, respectively, for occurrences of those values in our data. Thus, the true d′ values were artificially constrained (although to a small extent) at the high and low extremes (d′max = 6.93; d′min = 0.00).

APPENDIX E

Predicted data

An ANOVA was also performed on the predicted data, for the 3 comparison types by 10 stimulus pairs (12-88-msec mean gaps). The results are listed in Table 2, with the Tukey pairwise contrasts presented in Table 1. As Figure 6 indicates, the pattern of predicted results was quite similar in form to the pattern of obtained results.

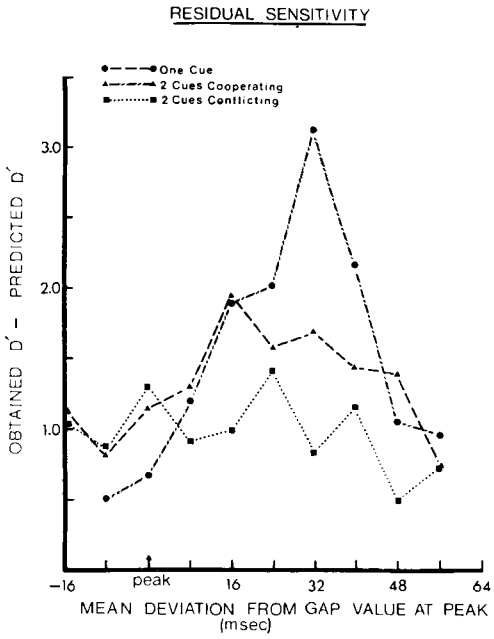

Figure E1. Residual sensitivity (obtained d' minus predicted d') found in the three-way 2IAX test, Experiment 2.

Obtained vs. Predicted Differences

In order to compare obtained and predicted true d' values, ANOVAs were conducted for each of the three comparison types, for the 2 functions (obtained vs. predicted) by 10 stimulus pairs (12-88-msec mean gaps). The pattern of differences between obtained and predicted functions was virtually the same as in Experiment 1 (see Table 2, "Cooperating cues," "One cue," and "Conflicting cues"), demonstrating a higher sensitivity for obtained than for predicted data, particularly for comparisons beyond the immediate range of the between-category performance peaks. Mean residual sensitivities were calculated (obtained true d' minus predicted true d') for the three comparison types, across all comparisons in which the pair members had deviated from one another by less than 10% in proportion of "stay" identifications (roughly "within-category" pairs, as defined in Experiment 1); included were "conflicting cues" comparisons with mean gap durations (per pair) of 12 msec and between 44 and 84 msec, "one cue" comparisons with gaps of 8 msec and between 48 and 88 msec, and "cooperating cues" comparisons with mean gaps of between 60 and 84 msec. These obtained-predicted difference values followed the order "cooperating cues" (mean residual sensitivity = 1.49) ≥ "one cue" (1.46) > "conflicting cues" (1.37). However, because of the shift in positions of the obtained peaks relative to the predicted peaks, we recalculated residual sensitivity values (obtained-predicted differences) according to deviations in gap duration from the value at the peak. The order of these residual sensitivities (Figure E1) was "cooperating cues" (1.58) > "one cue" (1.38) > "conflicting cues" (.98).

Individual Patterns

The individual patterns of mean "peak-range" sensitivities were virtually identical to those found in Experiment 1. Nine of the 14 subjects ($p < .001$ by binomial test) showed the expected order: "cooperating cues" > "one cue" > "conflicting cues." However, the key criterion ("cooperating cues" > "conflicting cues") was met by all 14 subjects ($p < .0002$). As in Experiment 1, the ANOVAs performed on the "peak-range" data (20-48-msec mean gaps—cf. Appendix C, Experiment 1) for both the obtained and the predicted true d' values upheld the individual analyses (see Table 2).

APPENDIX F

Individual Analysis: "Say"-"Stay" Listeners

Four of the five listeners in this group met the crucial criterion for "phonetic" perceptual equivalence ("cooperating cues" > "conflicting cues"; $p < .04$ by binomial test). The subject who failed had listened for "different water drips" during most of the discrimination test; her data show much the same pattern as the "spectral" group (described below). All four listeners who had more consistently perceived "say" and "stay" met the criterion ($p < .01$).

Obtained-Predicted Differences

Group comparisons. All three groups of listeners showed better obtained than predicted discrimination (compare lower left panels and the corresponding lower right panels, Figures 9-11). However, the pattern of obtained-predicted differences varied among the groups. To directly compare the group patterns, mean percentages of performance unexplained by AXB categorizations (residual performance = obtained mean minus predicted mean) were calculated for the entire range, within each comparison type.

"Temporal" listeners. The residual performance levels for the "temporal" group were small, and followed a different pattern from that observed in the two previous "say"-"stay" experiments: "conflicting cues" (mean residual performance = 8.4%) > "cooperating cues" (7.0%) > "one cue" (6.3%).

"Spectral" listeners. Residual performance for this group was very high, unaffected by between-category vs. within-category considerations, and showed the order: "conflicting cues" (mean residual performance = 43.9%) > "one cue" (35.2%) > "cooperating cues" (22.3%). This residual performance pattern was constrained by near-ceiling obtained performance for all three comparison types (artificially reducing the residual performance level, especially for "cooperating cues").

"Say"-"Stay" listeners. The residual performance pattern for the sine-wave "say"-"stay" group was like the patterns found in the two previous "say"-"stay" experiments: "cooperating cues" (mean residual performance = 19.5%) ≥ "one cue" (19.4%) > "conflicting cues" (14.6%). Also, the positions of the obtained peaks were slightly shifted toward the SW category boundary, relative to the positions of the predicted peaks, just as the obtained peaks in the first two experiments had shifted toward the D boundary.

(Received for publication October 15, 1980; accepted November 4, 1980.)