PSYCHOLOGICAL SCIENCE

## Research Article

# ON THE IMMEDIACY OF PHONETIC PERCEPTION

Yi Xu,[1] Alvin M. Liberman,[2] and D.H. Whalen[2]

[1]*Northwestern University and* [2]*Haskins Laboratories*

**Abstract**—*If, as we believe, language is a specialization all the way down to its roots, then perception of its consonantal elements should be immediately phonetic, not, as in the conventional view, a secondary translation from percepts of an auditory sort. Supporting observations come from an experiment in which formant transitions that distinguish [da] and [ga] were presented as sinusoids and combined with a synthetic syllable made of resonances, thus causing the auditory system to treat these acoustically incoherent parts as different sources. Evidence for the source difference was varied by changing the intensity of the sinusoids relative to the remainder of the syllable. Over the greater part of a 60-dB range, listeners accurately identified the consonants, indicating that they had integrated the stimuli according to a coherence that existed only in the phonetic domain. At the lowest intensities, indeed, the consonants were accurately identified, even though the whistles—the normal responses to the sinusoids—were not. There followed, then, a range over which perception was duplex: Both consonants and whistles were accurately identified. At the highest intensities, phonetic integration failed, but accurate perception of the whistles was maintained. That the phonetic percept was present when its auditory counterpart was absent, and vice versa, is evidence that the phonetic percept is independent of its auditory counterpart and not a translation from it, as is the fact that the two percepts followed very different courses in response to the experimental variable.*

The speech system is normally called on to form into a coherent phonetic percept acoustic information that comes from separate sources. Perhaps the clearest examples are to be found in the perception of fricative-vowel syllables. In such syllables, the band-limited noise produced at the point of consonant constriction is perceptually integrated with the shifting resonances (formant transitions) that reflect the filtering of the periodic voice source as the vocal tract changes shape from consonant to vowel. Thus, both the initial noise and the periodic signal that follows it contribute critically to the fricative percept (Whalen, 1981, 1984). (Parsing in the reverse direction, listeners use information in the fricative noise to perceive the vowel; Whalen, 1989.) Listeners are, of course, aware that the noise source they associate with the fricative is different from the periodic source appropriate to the vowel, but not that the fricative itself is critically informed by both sources. Thus, awareness of the fricative appears to be an immediate and coherent percept, not a postperceptual decision arrived at by combining distinct representations of a noise and the periodic sound that follows.

A further implication of the fricative-vowel example is that the objects of phonetic perception are the phonetically significant gestures, and that those gestures, not the sounds that convey them, are the ultimate constituents of language (Fowler & Rosenblum, 1991; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). On that theory, one sees that listeners parse the signal correctly, assigning the acoustically heterogeneous and disjoint consequences of a phonetically coherent fricative gesture to a correspondingly coherent percept of a distinctly phonetic sort. One form of the theory holds that this is an example of the way perceptual systems normally represent the distal object or event—in this case, the gesture that shaped the sound (Fowler, 1991). We agree that the immediate percept is the gesture, but hold that recovering it from the sound is the business of a distinct phonetic module, a biological specialization that also manages the overlapping and merging of a gesture with those that adjoin it (Liberman & Mattingly, 1985). Both forms of the theory are in opposition to the more common view, which is that perception of phonetic structures is by way of representations of a generally auditory sort (Diehl & Kluender, 1989; Massaro & Oden, 1980; Stevens & Blumstein, 1981). On that view, the primary percepts are distinctly auditory, requiring, then, translation into phonetic units by some associative process before lending themselves to the other components of the language they serve. On our view, the primary percepts are distinctly phonetic *ab initio*, hence perfectly suited, without cognitive translation, to a correspondingly distinct linguistic function.

In the perception of a fricative (followed by a vowel), phonetic integration occurs across sound-generating sources that produce dissimilar acoustic signals. If auditory grouping is accomplished by the kinds of Gestalt principles proposed by Bregman (1990), such sounds should constitute two sources. Only when these two sounds are treated as the result of a unitary phonetic gesture can they be combined into one percept (Mattingly & Liberman, 1988; Remez, Rubin, Berns, Pardo, & Lang, 1994). But such is the independence and strength of the phonetic module that integration also occurs even when the dissimilarity is, by experimental means, made to be ecologically implausible. In those cases, however, there is a duplex percept: Listeners perceive a unitary phonetic segment that could only have been formed by integration across the dissimilar sound-producing sources, while at the same time perceiving the dissimilarity to which the auditory system would be expected to respond.

The particular form of duplex perception investigated in the study reported here follows our earlier work (Whalen & Liberman, 1987). We synthesized stop-vowel syllables ([da] and [ga]) in which the perceived difference between the stops depended entirely on the 50-ms transition of the third formant. Then we divided the syllable into two parts. One part was just the third-formant transition that critically distinguished the consonants; the other was the (constant) base with which the transitions were combined. In order that auditory scene analysis would represent the two parts as different sources, we synthesized the base out of normal formants (i.e., resonances), but used a frequency-varying sinusoid for the critical transition. In isolation, the base sounded like a syllable, but, lacking the critical third-formant transition, it was ambiguously [da] or [ga]. Each of the sinusoids sounded in isolation like a whistle, bearing no relation to the ambiguous syllable, or to speech of any kind.

At moderate intensity levels, the patterns of our previous experiment did typically produce a duplex percept: an identifiable [da] or

---

Address correspondence to Alvin M. Liberman, Haskins Laboratories, 270 Crown St., New Haven, CT 06511; e-mail: haskins@haskins.yale.edu.

[ga] (depending on the sinusoidal transition) and, simultaneously, a whistle just like the one produced by the sinusoid in isolation. Thus, as in all instances of duplex perception the same piece of sound, contained in the same context and perceived by the same brain, was represented as an integral part of a phonetic percept and, at the same time, as a percept of a purely auditory sort. In prior experiments, such duplexity has been taken as one piece of evidence for the existence of separate phonetic and auditory modules, each capable of producing its own primary percept (Mann & Liberman, 1983; Mattingly & Liberman, 1990). It has also been seen as demonstrating the ability of the phonetic system to override auditory scene analysis as it parses the signal into the separate sources that the processes of ordinary auditory perception normally accept.

The purpose of our previous experiment was not only to add another form of duplex perception to those already studied, but, more particularly, to assess the relative sensitivity of the two systems—phonetic and auditory—that produce the two sides of the percept. To that end, we had each subject modify the intensity of the sinusoid relative to the base, hence the evidence for separate sources, until the whistle was at its absolute threshold. Then, having set the intensity of the sinusoid 4 dB below that level, we determined that listeners could, nevertheless, still accurately discriminate the stops. Indeed, subjects were as accurate at 4 dB below the whistle's threshold as they were at 6 dB above it. This finding—that the consonant was perceived, though the whistle was not—was called *phonetic precedence*. But, by any name, it testifies once again to the existence of a distinct phonetic system, and also to the remarkable sensitivity it exhibits when called on to do that for which it is biologically specialized. More pointedly, it supports the view that the phonetic representation is not a translation from an auditory percept, but rather a primary percept in its own right, because it was strongly present even when, at low levels of the sinusoid, the auditory representation (the whistle) had yet to make its appearance.

The precedence of the phonetic system has been confirmed in several experiments that used an earlier and quite different version of the duplex phenomenon (Bentin & Mann, 1990; Nygaard & Eimas, 1990), and an effect that implies phonetic precedence has been found in the responses of 2- to 4-month-old infants (Eimas & Miller, 1992). The most recent replication and extension of the precedence effect (in adults) is an experiment by Vorperian, Ochs, and Grantham (1995). Using stimuli like those in our previous experiment (Whalen & Liberman, 1987), they progressively increased the intensity of the sinusoid and found a large phonetic precedence effect over a range of about 20 dB, followed then by a range over which perception was duplex; beyond a certain intensity, however, duplex perception ceased as identification of the stop consonants sank to chance, having given way to a dominance of [da] responses. But, except in the determination of the duplexity threshold, perception of the nonspeech whistle was not measured, so we cannot know how the phonetic and auditory representations varied in relation to each other.

In an attempt to account for the phenomenon of phonetic precedence, we (Whalen & Liberman, 1987) suggested that the phonetic system preempts the information, passing on to the auditory system whatever it does not need. The same suggestion has been made (Liberman & Mattingly, 1989; Mattingly & Liberman, 1988) to explain why duplex perception does not occur in the normal case. Although that explanation for precedence had initial appeal, it does not explain further results showing the weakening (Ciocca & Bregman, 1989; Whalen & Liberman, 1996) or disappearance (Vorperian et al., 1995) of phonetic perception in a variety of circumstances. An alternative, put forth by Fowler and Rosenblum (1991), is that the auditory and phonetic representations stand to each other in a reciprocal relation. In that case, there could be a single processor that sometimes divides its outcome between representations of two candidates for recognition as the distal event. Still another possibility, of course, is simply that the phonetic and auditory processes are not related at all. That possibility seems especially attractive from a biological point of view, for it guarantees that the exigencies of the one function do not impinge on the other; it follows that evolutionary adaptations to the unique requirements of the combinatorial strategy that underlies phonological communication did not so warp the auditory system as to distort perception of acoustic events that are no part of language (Liberman, 1996, chap. 1).

To provide data relevant to those alternative interpretations, and to learn more about the properties of the phonetic system, we undertook to see how variations in the evidence for distinct sources affect the growth or decline of their phonetic and auditory representations.

## METHOD

### Stimuli

The stimuli, very similar to those we used before (Whalen & Liberman, 1987), were synthetic approximations to the syllables [da] and [ga], prepared on a parallel resonance software synthesizer. As shown in Figure 1, the patterns for both syllables comprised 250-ms coterminous trajectories for three formants (F1, F2, F3). Each trajectory had a 50-ms transition, linear in the frequency domain, ending at the value for the steady-state vowel. The steady-state center frequencies were 765, 1230, and 2527 Hz for F1, F2, and F3, respectively, as are appropriate for the vowel [a]. The excitation frequency (F0) was 100 Hz for the first 80 ms, followed by a linear transition to 80 Hz at the end of the syllable. For the F1 transition, the onset value was 400 Hz, as is appropriate for a voiced stop before [a]. The F2 onset was at 1650 Hz, which was as close as we could come to a setting that would be neutral between [da] and [ga]. (The amplitude of F2 was lower than it would normally be, in order to reduce its influence on the perception of the place of production of the stop consonant.) The distinction between the two stops rested solely on the F3 transitions, with onsets of 2800 Hz for [da] and 2018 Hz for [ga].

Each syllable was divided into two parts: one, the *transition*, was just the movement of the third formant that critically distinguished [da] from [ga]; the other, the *base*, was all of the common remainder. The transition was a frequency-varying sinusoid, whereas the base consisted of the resonances that are characteristic of speech. Thus, the two parts were vastly different in fundamental frequency and in spectrum, hence calculated to be sufficiently incoherent from an auditory point of view to be perceived as different sources. Not surprisingly, the sinusoids sound like whistles and are easy to discriminate or identify in isolation. The resonances of the base produce a percept that sounds like a syllable; in pilot tests, most listeners judged it, with considerable uncertainty, to be more like [da] than [ga].

To vary the evidence for separate sources, we set the intensity of the sinusoids at each of 11 levels, starting near the absolute threshold of the excerpt as heard in isolation, and then increasing in steps of 6 dB to a level 60 dB higher. There were, then 22 stimuli in all, half

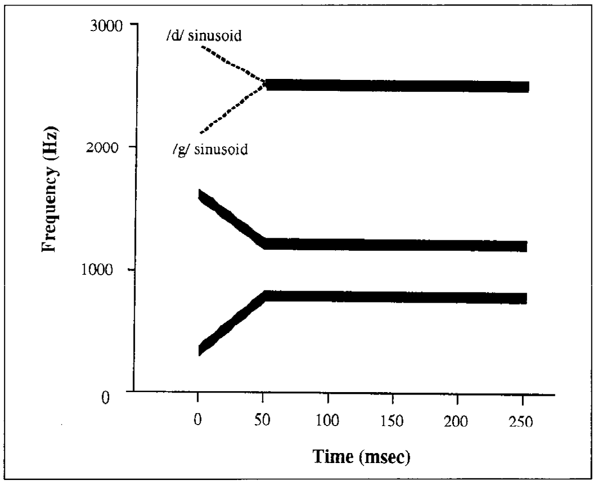

**Fig. 1.** Schematic spectrograms showing the acoustic composition of the stimuli. Dotted lines represent the sinusoidal transitions, and solid lines, the base.

with the sinusoid for [da], the other half with the sinusoid for [ga]. Ten copies of each were recorded on magnetic tape, the order of the syllables with the two types of excerpts having been randomized at each intensity level.

### Procedure

The stimuli were presented in one condition with instructions to listen for the speech and identify each token as [da] or [ga]; in a second condition, the instructions were to listen for the whistle and identify it as "high" or "low." Different randomizations of the stimuli were used in the two conditions. At each intensity level, the subject made 10 judgments of each of the two stimulus types. For all subjects, the order of presentation was from low to high intensity of the sinusoidal excerpt. Half of the subjects were given the speech condition first, and half, the whistle condition. Stimuli were presented binaurally through headphones in a sound-treated room.

There was a training session with sinusoids set at a level sufficient to produce a duplex percept. Each subject had 40 trials of speech identification and 40 trials of whistle identification. No subject was rejected on the basis of performance during the training.

### Subjects

The subjects were 12 undergraduates at the University of Connecticut. All were paid for their participation.

## RESULTS

In Figure 2, one can see that the identification functions for speech and whistle are radically different. (A two-factor repeated measure analysis of variance found a significant interaction between identification task and attenuation level, $F[10, 110] = 12.89, p < .001$; also, the order of judgments made no difference in the results.) As the attenuation of the sinusoid (relative to the base) decreases, the functions divide into three sections. Over the first section, from −60 dB though −48 dB attenuation, identification of the whistles was at chance, but consonant identification was significantly high. The results in this section confirm yet again the phonetic precedence described in the introduction: Phonetic integration occurred across the two types of sounds (the sinusoidal transition and the resonances of the base) that must be presumed to have come from distinct sources, though the auditory representations of the sinusoid (the whistles) were not yet in evidence. The second section of the function, from −42 dB through −18 dB, is characterized by duplex perception: The auditory and phonetic forms were both significantly represented. Through the first part of that range, the whistles became progressively more identifiable, reaching a peak at −24 dB, beyond which they declined somewhat, though they remained above chance to the end of the range at 0 dB. However, the phonetic integration that is responsible for the phonetic side of the percept began to weaken beyond −36 dB, declining progressively from that point until, at −12 dB, it failed utterly. At that level, the third section begins; there, the consonants were gone and only the whistles remained.

At the three greatest attenuations of the sinusoidal transitions, where identification of the whistle was at chance, listeners were unsure whether or not they could even hear it. Then, over the range of duplex perception, the whistle side of the percept grew progressively louder. That increase continued beyond the point where the phonetic side of the percept had declined to chance. Analogous statements cannot be made about the speech percept, because the consonants one hears do not register as loudnesses apart from the syllables in which they are contained; the difference listeners hear is reflected rather in the confidence they place in their identifications.

Having established that the two identification functions are significantly different, we should consider whether or not the relation between them is reciprocal. By inspection, it appears not to be. If we examine the direction of change in accuracy between adjacent members of the continuum, we find that in five cases, the functions go in opposite directions, but in five cases they go in the same direction. This can be seen statistically in the lack of a significant negative correlation between the percentage correct values for the two functions of Figure 2 at the 11 levels of whistle intensity ($r = -.39$, n.s.) and a lack of correlation between the differences of adjacent members of the continuum ($r = .00$, n.s.). (That these correlations are a reasonable measure of reciprocity is seen in the fact that the functions of

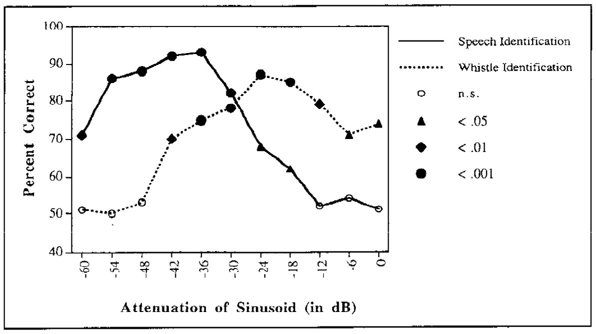

**Fig. 2.** Identification of the consonants and whistles. Levels of significance (difference from chance calculated via a $t$ test) are coded according to the key at the right of the figure.

Fig. 3, in which the reciprocity is not between speech and nonspeech but between the two alternative responses in speech, give correlations of −.78, $p < .01$, for both measures.) Thus, we would be hard put to claim that the functions were reciprocal.

In an earlier experiment with the same stimuli, we found in a pretest that subjects performed at chance levels when they tried to identify the sinusoidal transitions in isolation as [d] and [g] (Whalen & Liberman, 1987). That finding has been challenged by Bailey and Herrmann (1993), who claim that listeners can, in fact, perceive the sinusoids as the correct consonants. We note, therefore, that if that had been so in the present experiment, we should not have found the radically different response functions seen in Figure 2. Bailey and Herrmann also challenged the method of our 1987 study for obtaining thresholds for the whistle detection. However, for the purposes of demonstrating phonetic precedence, the exact threshold for the whistle is not critical. What is critical is that, as the present results and those of Vorperian et al. (1995) show, there is a range of intensities over which the speech is discriminable and the whistles are not.

The functions in Figure 3 show that the subjects were equally accurate in consonant identification for the patterns containing the [d] sinusoids and those containing the [g] sinusoids. Of particular interest are the results for the three highest intensities of the sinusoids, where we see that responses were equally divided between [d] and [g]. This finding means simply that when phonetic integration failed, listeners were left with the more or less ambiguous syllable that is produced by the base alone. Vorperian et al. (1995) found that [d] responses predominated at the highest intensities of the whistle. The relative amplitude levels used in the present study were similar to those of Vorperian et al., and the method was similar except for the (seldom-used) category of "ambiguous" that Vorperian et al. included. We do not have an explanation for this difference in results.

That the phonetic representation does not retain its strength with increasing evidence for different sources shows that the phonetic system is not, in the strict sense, preempting the information. Because the response functions are not reciprocal, we should not conclude that there is a single processor that can yield one or another of two distal events. Rather, it is plain that the two representations are simply independent of each other. Each is sometimes present when the other is absent, and they follow radically different courses in response to increasingly compelling evidence that the two parts of the acoustic signal come from different sources.

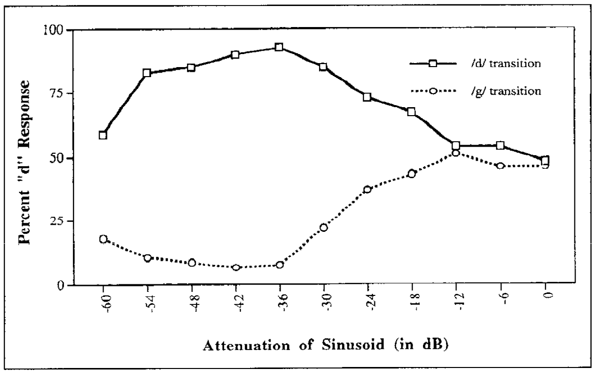

**Fig. 3.** Phonetic identification of syllables, with [d] and [g] transitions shown separately.

## DISCUSSION

The common view of speech perception as primarily auditory honors a tradition that associates primary perceptual representations with end organs. Because the phonetic system lacks an end organ of its own, its representations cannot, in that tradition, be primary; they must, rather, be secondary translations from auditory percepts, because it is only the auditory percepts that have first claim on the ear. In its most general form, such a view was enshrined almost 300 years ago by Berkeley (1709), who, in pursuit of his philosophical objectives, asked how people manage to perceive visual depth, given no end organ appropriate to that function. Berkeley's answer was that viewers converge their eyes to perceive objects at different depths, learning, then, to experience depth secondarily by association with the primary perception of strain that arises as the relevant muscles contract. But is that not strikingly similar to the contemporary view of speech, which is that phonetic structure is experienced, also secondarily, by association with the primary auditory percepts that are evoked when the sounds of speech strike the ear? In the case of depth, scholars have in modern times come to the better understanding that it is, in fact, a primary percept, the product of a system specifically adapted to processing binocular disparity so as to represent depth immediately.

The foregoing is not a wholly idle digression, because the results of the experiment reported here support the view that, notwithstanding the absence of an end organ of their own, phonetic percepts are, like phenomenal depth, the immediate products of a perceptual specialization. This conclusion follows from the very existence of duplex perception, but also from the fact, once again confirmed in the current study, that the phonetic percept is present under conditions in which its auditory counterpart is not; and now we also know, as we did not before, that the auditory and phonetic sides of the duplex percept follow radically different courses as they respond to increasing evidence for the separateness of the sources.

The comparison between phonetic perception and stereopsis is appropriate for yet another reason, which is that both can be seen as members of a class of biological specializations that produce heteromorphic percepts, that is, percepts that are incommensurate with the dimensions of the stimuli to which they respond (Liberman, 1996; Liberman & Mattingly, 1989; Mattingly & Liberman, 1988). Thus, in speech, in which the acoustic information for each phonetic element is overlapped with information for other elements and presented as continuously changing formant trajectories, listeners do not perceive the continuous variations in timbre that would seem the proper homomorphic response, but rather a string of discrete, invariant, and categorical segments. In similar fashion, the specialization for stereopsis does not represent width disparity homomorphically as double vision but rather heteromorphically as phenomenal depth.

Among the other properties common to such specializations, one that has relevance here is elasticity (Liberman, 1996; Whalen & Liberman, 1996). Thus, stereopsis produces its heteromorphic percept over a range of ecologically implausible disparities—that is, disparities greater than might ever be provided by the distance between the eyes—much as the phonetic module accommodates to ecologically implausible dissimilarity of sources. The analogy holds further, for as disparity is progressively increased, homomorphic double vision appears. Depth continues to be perceived, however, producing the visual counterpart of duplex perception (Richards, 1971). Then, beyond the limit of elasticity of the stereopsis module, depth perception weakens and finally fails utterly, leaving the perceptual field to double vision, much as phonetic perception was found to weaken and finally fail, as the evidence for a separate tone source became too great for the phonetic system to overcome.

Thus, the phonetic system is a specialization like stereopsis, or, indeed, like language itself. It is independent of the auditory specializations, so each is free to do that for which it was adapted. They share a common end organ, and all take part in the perception of speech: The phonetic specialization represents the commutable—hence, discrete, invariant, and categorical—segments that the combinatorial strategy of language requires, relying on the auditory specializations to represent those paraphonetic aspects of speech perception (e.g., pitch, loudness, and timbre) that characterize the response to all sounds.

**Acknowledgments**—This research was supported by National Institutes of Health Grant HD-01994 to Haskins Laboratories.

## REFERENCES

Bailey, P.J., & Herrmann, P. (1993). A reexamination of duplex perception evoked by intensity differences. *Perception & Psychophysics, 54,* 20–32.
Bentin, S., & Mann, V. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88,* 64–74.
Berkeley, G. (1709). *An essay towards a new theory of vision.* Dublin: Aaron Rhames.
Bregman, A.S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.
Ciocca, V., & Bregman, A.S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics, 46,* 39–48.
Diehl, R.L., & Kluender, K.R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.
Eimas, P.D., & Miller, J.L. (1992). Organization in the perception of speech by young infants. *Psychological Science, 3,* 340–345.
Fowler, C.A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America, 89,* 2910–2915.
Fowler, C.A., & Rosenblum, L.D. (1991). The perception of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (Eds.) *Modularity and the motor theory of speech perception* (pp. 33–59). Hillsdale, NJ: Erlbaum.
Liberman, A.M. (1996). *Speech: A special code.* Cambridge, MA: MIT Press.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.
Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.
Liberman, A.M., & Mattingly, I.G. (1989). A specialization for speech perception. *Science, 243,* 489–494.
Mann, V.A., & Liberman, A.M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14,* 211–235.
Massaro, D.W., & Oden, G.C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67,* 996–1013.
Mattingly, I.G., & Liberman, A.M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Auditory function* (pp. 775–793). New York: John Wiley & Sons.
Mattingly, I.G., & Liberman, A.M. (1990). Speech and other auditory modules. In G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501–519). New York: John Wiley & Sons.
Nygaard, L.C., & Eimas, P.D. (1990). A new version of duplex perception: Evidence for phonetic and nonphonetic fusion. *Journal of the Acoustical Society of America, 88,* 75–86.
Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., & Lang, J.M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.
Richards, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America, 61,* 410–414.
Stevens, K.N., & Blumstein, S.E. (1981). The search for invariant acoustic correlates of phonetic features. In P.D. Eimas & J.L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Erlbaum.
Vorperian, H.K., Ochs, M.T., & Grantham, D.W. (1995). Stimulus intensity and fundamental frequency effects on duplex perception. *Journal of the Acoustical Society of America, 98,* 735–744.
Whalen, D.H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. *Journal of the Acoustical Society of America, 69,* 275–282.
Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49–64.
Whalen, D.H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics, 46,* 284–292.
Whalen, D.H., & Liberman, A.M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169–171.
Whalen, D.H., & Liberman, A.M. (1996). Limits on phonetic integration in duplex perception. *Perception & Psychophysics, 58,* 857–870.

(Received 6/7/96; Accepted 9/18/96)