Journal of Experimental Psychology:
Human Perception and Performance
1992, Vol. 18, No. 3, 752–762

0096-1523/92/$3.00

Musical Duplex Perception: Perception of Figurally Good Chords With Subliminal Distinguishing Tones

Michael D. Hall and Richard E. Pastore
State University of New York at Binghamton

In a variant of duplex perception with speech, phoneme perception is maintained when distinguishing components are presented below intensities required for separate detection, forming the basis for the claim that a phonetic module takes precedence over nonspeech processing. This finding is replicated with music chords (C major and minor) created by mixing a piano fifth with a sinusoidal distinguishing tone (E or E♭). Individual threshold intensities for detecting E or E♭ in the context of the fixed piano tones are established. Chord discrimination thresholds defined by distinguishing tone intensity were determined. Experiment 2 verified masked detection thresholds and subliminal chord identification for experienced musicians. Accurate chord perception was maintained at distinguishing tone intensities nearly 20 dB below the threshold for separate detection. Speech and music findings are argued to demonstrate general perceptual principles.

Duplex perception (DP) is said to occur when one stimulus simultaneously contributes to two distinct percepts; this statement provides a broad definition of DP.[1] The DP phenomenon, first described for speech stimuli, has been the source of controversy about the nature of auditory speech processing, specifically whether there is a biologically distinct mechanism (a phonetic module) that processes only speech information (Liberman & Mattingly, 1989; Mattingly & Liberman, 1988). One recent variant of DP has been claimed to provide evidence that this phonetic module takes precedence over general auditory processing (Whalen & Liberman, 1987). We discuss the nature of this variant of DP, relating it to other DP demonstrations. We then present evidence replicating this variant with music stimuli. After discussing alternative explanations for this variant of DP, we discuss the general nature of DP phenomena and the contribution of DP research to understanding auditory perception.

Summary of Terms and Symbols

Demonstrations of DP encompass a variety of different procedures and underlying perceptual processes. To evaluate any demonstration of DP it is necessary to compare and contrast it with other examples of DP. The Appendix provides a summary of the symbols used to specify the physical stimuli and perception of the stimuli. Lowercase letters refer to physical stimuli, and uppercase letters refer to perception. *s* is a single, complex stimulus that can be partitioned into two fixed components, *b* and *c*. $s(i)$ represents one of two or more complex stimuli, each of which can be partitioned into two components, a common base, *b*, and a distinguishing component, $c(i)$. *S* indicates perception that is based on *s*, whereas $S(i)$ indicates perception that is based on the different versions of $s(i)$. In most demonstrations of DP, the different intact stimuli, $s(i)$, result in relatively discrete labeling categories, $S(i)$, on the basis of qualitative aspects of $c(i)$, thus providing an important objective standard for the correctness of responses that are based on the merging, grouping, or fusion of $c(i)$ with *b*.[2]

Original DP Phenomenon

Table 1 summarizes the physical stimuli, presentation conditions, and percepts for two important versions of DP: the original demonstration and the Whalen and Liberman variant of DP. Original speech DP is labeled DP(f) to reflect the critical importance of perceptual fusion of the base, *b*, and distinguishing component, $c(i)$. One typical example of DP(f) is based on third formant (F3) transitions defining a place continuum (Rand, 1974). In analyzing speech waveforms, initial stop consonants exhibit rapid spectral (frequency) transitions to vowel formants. It is possible to synthesize a place continuum that varies only in F3 transition onset frequency and that is readily and consistently partitioned by listeners into relatively discrete categories (e.g., /da/ and /ga/).

This work was supported by National Science Foundation Grant BNS8911456. Opinions, findings, conclusions, and recommendations are the authors' and do not necessarily reflect views of the National Science Foundation.

We gratefully acknowledge Xiao-Feng Li, Jody K. Layer, and Cynthia M. Connine for their comments and suggestions.

Correspondence concerning this article should be addressed to either Michael D. Hall or Richard E. Pastore, Department of Psychology, State University of New York, Binghamton, New York 13902-6000.

---

[1] Bregman (1990) used the abbreviation DPS for duplex perception of speech. Because duplex perception is not restricted to speech stimuli, we use the abbreviation DP.

[2] DP is claimed to reflect perception, not arbitrary changes in response tendency due to task or experimental demands. Therefore, phenomena must provide strong evidence for fusion (in DP[f]) or separation by stimulus quality (in DP[s]) to be regarded as valid demonstrations of DP. The use of an objective standard helps validate findings as representing actual perceptual differences rather than simple criterion differences.

Table 1
Summary of Original and Current Conditions Demonstrating Duplex Perception (DP)

	Physical stimuli		Perception		
DP type	Ear 1	Ear 2	Ear 1	Center	Ear 2
	Original DP: Rand (1974), Pastore, Schmuckler, Rosenblum, and Szczesiul (1983)				
DF(f)	b		B		
	$s(i)$		$S(i)$		
	b	$c(i)$	$S(i)$		$C(i)$
	DP variant: Whalen and Liberman (1987), current study				
DP(s)	b	b		B	
	$b + c'(i)$	$b + c'(i)$		$S(i), C'(i)$	
	$b + c'(i)\downarrow$	$b + c'(i)\downarrow$		$S(i)$	

In such a continuum, the invariant portion of a consonant–vowel (CV) syllable (the base), missing any F3 transition, is perceived as having a relatively ambiguous initial consonant. Isolated formant transitions are heard as nonspeech chirps, which without practice may be perceived relatively continuously (Pastore, Li, & Layer, 1990).

In speech DP(f), one of the F3 transitions, $c(i)$, and the base, b, are presented to separate ears. If perception corresponds to the physical stimuli, and if the ears represent separate input channels, subjects should perceive two events represented by the separate stimuli: the chirp, $C(i)$, in one ear, and the ambiguous base, B, in the other. While maintaining the separate perception of the chirp, however, the perceptual system combines (fuses) information across the two ears, resulting in perception of the intact CV syllable, $S(i)$, instead of the ambiguous base, attributing location to the ear that receives b (Liberman, Isenberg, & Rakerd, 1981; Rand, 1974). Extensive research on dichotic fusion clearly indicates that information across the two ears will be combined or fused at any of a number of levels and often will result in two simultaneous perceptions (Cutting, 1976).

Liberman and Mattingly (1985, 1989; Mattingly & Liberman, 1988) attributed DP(f) to the distinct operations of a specialized phonetic module and a separate module that performs auditory scene analysis. Auditory scene analysis, which is a construct in many theories of perception, is a hypothetical stage or process that uses the various elements of sensory input to develop representations of reality in terms of objects and events. Auditory scene analysis is conjectured to operate according to a number of different principles, including exclusive allocation. According to one version of the exclusive allocation principle, stimulus information is used only once by the scene analysis system (Bregman, 1987; Liberman & Mattingly, 1989). In DP(f), however, $c(i)$ is perceived as a simple nonspeech auditory event, $C(i)$, and it fuses with the base to produce perception of $S(i)$. This dual role of $c(i)$ is cited by Liberman and Mattingly (1985, 1989; Mattingly & Liberman, 1988) as a violation of the exclusive allocation principle and thus as evidence for the operation of separate, biologically distinct processing mechanisms or modules. One perception (the chirp) is thought to be based on acoustic information assigned to one percept by the operation of scene analysis. The other, more complex perception (the syllable) is presumably based on a biologically distinct phonetic module (Liberman & Mattingly, 1985, 1989). The phonetic module has its own distinctively phonetic primitives and is independent of other modules (e.g., pitch, loudness, and timbre), including the module for scene analysis (Liberman & Mattingly, 1989).

DP(s) Variant of DP

Whalen and Liberman (1987) produced a variant of speech DP (DP[s] in Table 1) by diotic presentation of components b and $c(i)$. b was a synthetic three-formant CV syllable without any F3 transition that was physically mixed with $c'(i)$, a sinusoidal version of an F3 transition. At least conceptually, the base b and original version of the distinguishing component $c(i)$ were partitioned from two versions of a complex stimulus $s(i)$. Because the stimuli, $s(i)$, were perceived discretely as $S(i)$, there is an objective standard for evaluating fusion. Subjects reported simultaneously hearing a complete syllable, $S(i)$, and a nonspeech chirp, $C'(i)$, thus demonstrating DP. We refer to this condition as DP(s) to reflect the importance of perceptual separation of a complex stimulus, $s'(i)$, into two distinct percepts, $S'(i)$ and $c'(i)$.

Replacement of the critical F3 transition, $c(i)$, with a sinusoidal analogue, $c'(i)$, should (a) maintain some level of perceptual correspondence between the original ($s[i] = b + c[i]$) and modified ($s'[i] = b + c'[i]$ stimulus and (b) be perceived as arising from other than the vocal tract that might have produced the base b (e.g., Remez, Rubin, Pisoni, & Carrell, 1981). The respective expectations represent applications of (a) the Gestalt principles of proximity, good continuation, and pragnanz and (b) segregation (or separate streaming) required by the dissimilar stimuli (violating the grouping principle of similarity).

In both DP(f) and DP(s), components (b and either $c[i]$ or $c'[i]$) are either actual segments or derived from segments of an originally intact unambiguous signal and are presented in a manner that preserves good temporal and frequency continuation. These properties should encourage perceptual integration, grouping, or streaming that results in perception of the original, intact stimulus. Likewise, the presentation of information either to separate ears or with distinctive attributes should encourage perceptual segregation. The critical difference between DP(f) and DP(s), respectively, is whether the perceptual system groups aspects of component information across location or assigns aspects of component information (all presented in the same location) to a separate event, $C'(i)$. Thus, from one nonmodular perspective, DP(f) and DP(s) are based primarily on different general principles of perceptual organization.[3] This distinction between DP(f) and DP(s) would be rendered irrelevant if the phonetic module determines perception of the intact stimulus, with unused stimulus information then made available to general auditory processes (including scene analysis) to enable perception of the separate, nonspeech event.

The major DP(s) finding from the Whalen and Liberman study concerns the difference in threshold intensity for each of these two types of perception. Whalen and Liberman first measured *duplexity threshold*, the intensity at which sinusoidal transitions, $c'(i)\downarrow$, were no longer heard (detected) as chirps in the context of the base. Subjects were still able to accurately label syllables as *da* or *ga* with the transitions at −4 dB in relation to their individual duplexity thresholds. Whalen and Liberman also reported the additional, anecdotal claim that they personally could label syllables with F3 transitions attenuated to approximately −20 dB duplexity threshold. (Secondary reports of this study often cite only the anecdotal 20-dB effect [e.g., Coren & Ward, 1989; Goldstein, 1989].) Because accurate speech perception remained at intensities below duplexity threshold, it was claimed that a specialized phonetic module takes precedence over other auditory processing. According to this claim, energy that could be interpreted as speech is first processed by the speech module, with only energy left over from speech processing then available to be treated as nonspeech.

Other Examples of DP

DP, now claimed to be a relatively common occurrence for speech (Mattingly & Liberman, 1988), also may be a relatively common auditory phenomenon. In a number of auditory phenomena, a stimulus simultaneously contributes to two distinct percepts, thus meeting the broad definition of DP. These phenomena represent very different paradigms with various probable causes for integration and segregation of information by auditory scene analysis. These various phenomena do not share important elements with the demonstrations of DP(f) and DP(s). Our goal in introducing a few of these phenomena as examples of DP is to demonstrate that DP may be a fairly common attribute of perception and may reflect many different types and levels of underlying processes. The alternative DP phenomena include the phonemic restoration effect and a study by Fowler and Rosenblum (1990) that used a door slam, as well as aspects of a tone interaction study by Wegel and Lane (1924; see specifically the summary by Fletcher, 1940/1972).

In the phonemic restoration effect (summarized in Table 2), noise is added where a segment of speech had been excised from a sentence. The noise is critical to two different simultaneous percepts, with subjects reporting the presence of both the noise and the missing (restored) speech segment (Samuel, 1981; Warren, 1970). Nonspeech restoration effects have been demonstrated for music (Dewitt & Samuel, 1990) and tone glides (Dannenbring, 1976).

Fowler and Rosenblum (1990) recently claimed both to demonstrate DP and to replicate DP(s) by using door-slamming sounds under diotic and dichotic conditions. Their study (summarized in Table 2) used aspects of one natural nonspeech stimulus, the sound of a metal door slam, s. With filters, the high-frequency portion of the stimulus, c, was separated from the remainder, b, of this single stimulus, s. The different conditions were defined in terms of an intensity continuum of c mixed with a fixed intensity of b, and the ipsilateral or contralateral location of c in relation to b. With training, subjects learned to partition the intensity continuum. In addition, when c was sufficiently amplified, auditory scene analysis seemed to accommodate the resulting difference in quality by the simultaneous perception of a separate event, producing a version of DP. Therefore, the quantitative change in c (the only major variable) results in three different response categories: S (metal door) + C (high-frequency stimulus) at high levels of $c\uparrow$, S at only moderate levels of c, and $S-$ (labeled *wooden door*) when $c\downarrow$ is sufficiently attenuated or absent. We suspect that a similar pattern of results could be found for many stimuli (including those used in any example of DP[f] and DP[s]) by manipulating the intensity of a portion of a stimulus (e.g., one version of $c[i]$).

Fowler and Rosenblum's loose pattern of results is quite similar to some of the Wegel and Lane results reported by Fletcher (1940/1972). Wegel and Lane fixed the intensity of a 1200-Hz tone and then mapped perceptual thresholds as a function of the intensity of a higher frequency tone (e.g., 1700 Hz). The quantitative change in c resulted in three response categories: S (low tone colored by nonlinear [e.g., difference] attributes) + C (high tone) at high levels of $c\uparrow$, S at only moderate levels of c, and $S-$ (pure low tone) when $c\downarrow$ is sufficiently attenuated or absent. In the middle perceptual category, in which c is subliminal to perception of C, c continued to contribute to the perception of S over a range of 8–10 dB.

These are but three of many examples of auditory phenomena that meet the broad definition of DP. On the basis of a traditional view of stages of perceptual processing, we believe that these demonstrations of DP can be discussed in terms of levels of processing. These levels could range from low-level sensory interaction (e.g., Wegel & Lane, 1924) to higher level cognitive processes (e.g., phonemic restoration). The Fowler and Rosenblum demonstration, DP(s), and DP(f) would fall at (probably different) levels between these extremes (see the *Results and Discussion* section). We therefore would not argue for a common basis for the various examples of DP we have cited; a less traditional theoretical perspective (e.g., one that rejects constructs of internal processing) might view some of these findings as equivalent.

In the Whalen and Liberman study and our study, one qualitative aspect of $c(i)$ distinguishes two perceptual cate-

---

[3] We have used the terms *fusion* and *separation* as labels for the two primary principles. Fusion typically has been used to refer to integration of information across separate location (e.g., Cutting, 1976), which is not found in DP(s). We acknowledge that both DP(f) and DP(s) involve types of perceptual grouping and that any type of grouping could be claimed to reflect some type of fusion.

Table 2
Summary of Additional Conditions That Have Been or Could Be Claimed to Demonstrate Duplex Perception

Physical stimuli		Perception		
Ear 1	Ear 2	Ear 1	Center	Ear 2
Phonemic restoration effect: Samuel (1981)				
b		B		
$b + n$		S, N		
Diotic: Fowler and Rosenblum (1990), Wegel and Lane (1924)				
b		$S-$		
$b + c\downarrow$	$b + c\downarrow$		$S-$	
s	s		S	
$b + c\uparrow$	$b + c\uparrow$		S, C	
Dichotic: Fowler and Rosenblum (1990)				
b	$c\downarrow$	$S-$		
b	c	$S-$		C
b	$c\uparrow$	$S-$		C

gories. A different qualitative aspect of c' (its sinusoidal nature) determines its separate perception. Under these conditions (and in contrast to the Fowler and Rosenblum study, which only varied quantity), both studies could evaluate the differential effect of manipulating quantity (intensity) on the perceptual consequences of the two different qualitative attributes.

Music DP

Nonspeech DP(f) has been demonstrated with music (Collins, 1985; Pastore, Schmuckler, Rosenblum, & Szczesiul, 1983). Music represents a class of nonspeech stimuli that shares some interesting parallels with speech. Most notably, both speech and musical sequences follow strict rules of organization and construction (e.g., syntax). Music stimuli are perceived in terms of discrete categories with highly familiar labels (Burns & Ward, 1978; Siegel & Siegel, 1977). Musical stimuli therefore readily lend themselves to investigations of possible analogues to speech phenomena that are based on differential classification of stimuli. Although the possibility cannot be dismissed that a distinct music module exists, the frequent finding of processing equivalence for speech and music would lessen the significance of postulating distinct modules. Minimally, music DP reveals that speech stimuli are not always required for DP, and thus DP cannot be used as unequivocal evidence for phonetic modularity.

In music, an isolated tone (E/E♭) differentiates a C major (C–E–G) and C minor (C–E♭–G) chord. When the distinguishing tone ($c[i]$ = E/E♭) and the remainder of the chord (b = C–G) are presented to separate ears, many (but not all) musicians perceive a tone in one ear and a chord (C major or C minor) in the other. The distinguishing tone (E or E♭) is used once at the tone level and once at the chord level, implying either that DP is a general property of perception or that there is a music module that in terms of DP is equivalent to the conjectured speech module (Pastore et al., 1983).

Gestalt principles of figural organization loosely define "good" figures as frequently occurring multicomponent stimuli that give rise to organized, unified perception.[4] "Strong" figures are those that resist analysis into component elements. "Good," "strong" figures are perceived more easily than "weak" figures (Wertheimer, 1958). When elements of spectral patterns (tones or formants) frequently occur in particular combinations, the perceived patterns are good, strong figures (i.e., chords or syllables). In addition, there is a strong body of evidence in the music literature establishing that simple major and minor chords (e.g., C–E–G and C–E♭–G) represent good, strong, perceptual figures (Dowling & Harwood, 1986). The Gestalt framework seems especially appealing because it provides a widely applicable, parsimonious explanation of both speech and nonspeech demonstrations of DP that does not necessarily require the conjectured operation of distinct processing modules.

Experiment 1

We sought to replicate DP(s) with musical stimuli. We predicted that subjects would accurately detect differences between two chords on the basis of changes in the tones that distinguish them (E/E♭ or E♭/E) even when those tones were below duplexity threshold. The replication of DP(s) with musical stimuli would demonstrate that speech processing is not unique in taking precedence over other processing. On the basis of a general perceptual framework, this finding could suggest a hierarchical processing in which figurally good objects (complete syllables or chords) are more likely to be perceived.

---

[4] Modularity supporters have attempted to dismiss music DP by arguing that it represents *triplex* perception in which base, tone, and chord are heard simultaneously and individually (Liberman & Mattingly, 1989). Triplex perception has yet to be demonstrated with any stimuli. Our laboratory has attempted to develop strong tests for triplex perception, but has found only results that suggest that if triplex perception actually does exist for musical stimuli, it must occur very infrequently. (Further description of this research is beyond the scope of this article.)

Method

Subjects. Twenty subjects participated, consisting of 18 State University of New York at Binghamton introductory psychology students, 1 graduate student, and Hall. All subjects had studied a musical instrument, so in theory all understood the distinction between major and minor chords. Data from 8 other potential subjects were discarded because they did not meet the a priori criterion of reliable differentiation between the distinguishing tones in the discrimination task and thus failed to exhibit a duplexity threshold.[5]

Stimuli. $s'(i)$ were 1,424-ms tones from C major and C minor chords in an equitempered Western scale digitized at a 10-kHz sample rate with 4-kHz low-pass filtering. b was the C–G fifth (266 Hz and 398 Hz) recorded from a synthesizer's digitally sampled piano sound and was presented at 68.8 dB(A) peak intensity. Like the transitions used by Whalen and Liberman, the distinguishing E (335 Hz) and E♭ (316 Hz) tones ($c'[i]$) were sinusoidal. These tones were synthesized by the computer, matched the attack and decay amplitude envelope of the base, and ranged in peak intensity from 32 to 62 dB(A) in 5-dB increments (produced by uniform attenuation of the complete tone). b and $c'(i)$ tones were always diotic (binaural).

Procedure. Our procedure differed slightly from that of Whalen and Liberman. We used the method of constant stimuli to establish duplexity threshold.[6] E or E♭ tones were randomly presented 10 times at each intensity in base context, with subjects indicating whether they heard the sinusoidal tone. Subjects were presented only one of the distinguishing tones for all of these trials, either E ($n = 11$) or E♭ ($n = 9$); the distributions of duplexity thresholds were highly similar for both E and E♭ distinguishing tones (see Figure 2).

A second block of 145 random trials was used to evaluate chord discrimination ability at each of the seven tone intensities. Because the ability to label chords as major or minor can be highly variable even in trained musicians (Collins, 1985), a same–different (AX) procedure was used. Although nonprofessional musicians may not be able to label major and minor chords reliably, they can easily discriminate differences between them. Therefore, the AX procedure eliminated the need for using professional musicians as subjects. Subjects determined whether the two chords on a trial were the same or different. (Experiment 2 used experienced musicians with chord labeling or identification tasks.)

Each trial was approximately 6 s in duration; two chords were presented, with the middle tones always equal in intensity. The middle tone for each chord on a trial was E or E♭, with an independent probability of .5. A 500-ms interstimulus interval (ISI) separated chords on a trial. A 2-s response interval followed the second chord on each trial. There were 20 trials for each tone intensity. As a partial check on response bias, five additional catch trials were included in which only the piano base was presented as both stimuli within a trial.

Results and Discussion

Duplexity threshold was the tone intensity yielding 69.2% ($d' = 1.0$) detection of E or E♭ presented in base context. Individual duplexity thresholds, estimated with a least squares linear regression analysis, ranged from 61.6 to 41.7 dB(A) peak amplitude.

For AX trials, we calculated d' as a measure of accuracy at each tone intensity (Pastore & Scheirer, 1974). Here, d' is based on the probabilities that the subject responded *different* to nonequivalent and equivalent chords. d' can range up to approximately 4.7. Again, we used linear regression to estimate the discrimination threshold ($d' = 1.0$) for each subject.

Figure 1 plots mean psychophysical functions for tone detection (filled triangles) and chord discrimination (open circles) across the seven intensities of the sinusoidal distinguishing tones. The average duplexity (detection) and chord discrimination thresholds were 51 and 42 dB(A), respectively, yielding an average chord discrimination threshold of −9 dB average duplexity threshold. The large difference between these average thresholds is statistically significant, $F(1, 19) = 26.262$, $p < .0001$. Therefore, at tone intensities significantly below those required to detect E/E♭ in base context, subjects still accurately discriminated changes in the chords solely on the basis of the E/E♭ tones.

Figure 2 plots chord discrimination threshold as a function of duplexity threshold for each subject. The dark diagonal line labeled *0* indicates identical thresholds (0 = dB difference). Data points falling below this equivalent threshold line indicate that the chord discrimination threshold is below duplexity threshold. The remaining parallel lines indicate, in 5-dB steps, the degree to which chord discrimination threshold falls below duplexity threshold.

Nineteen of the 20 subjects could discriminate chords ($d' > 1$) at tone intensities below their duplexity thresholds; 1 subject (upper point in Figure 2) failed to discriminate the chords at any intensity. Subsequent analyses are based on the 19 subjects who could perform the discrimination task. Chord discrimination thresholds ($d' = 1$) for the 4 poorest subjects were only approximately −2 dB duplexity threshold. The remaining 15 subjects, however, exhibited thresholds beyond the −4 dB relative intensity at which Whalen and Liberman tested their subjects. Three subjects even exhibited chord discrimination thresholds that were an average of −17.4 dB duplexity threshold. Performance of these subjects is comparable to Whalen and Liberman's anecdotal claim from their own listening that the isolated component may determine figurally good perception up to approximately −20 dB duplexity threshold.

---

[5] Subjects were obtained by self-selection through the use of brief descriptions of the experiment that included requirements for participation (e.g., musical experience). A priori criteria are needed to identify subjects who have ignored or have lax criteria for the participation requirements.

[6] Whalen and Liberman measured duplexity threshold with a method of adjustment, a procedure developed to measure a difference threshold that is seldom used to measure an absolute threshold (Kling & Riggs, 1972; Snodgrass, 1975); the method of constant stimuli is preferred for precise measurement of absolute or masked threshold (Coren & Ward, 1989). In the method of constant stimuli, the random ordering of stimuli and the addition of catch trials avoids many of the problems associated with sequential methods (e.g., limits, adjustment). We do not question the Whalen and Liberman finding that subliminal stimuli for detection can contribute to recognition of highly familiar complex stimuli. Our point, in response to concerns raised by reviewers, is that our data are based on the use of a more appropriate method and that our findings are no less valid, and probably more accurate, than the findings reported by Whalen and Liberman. The use of different methods in the two studies actually provides evidence for the validity of both findings.

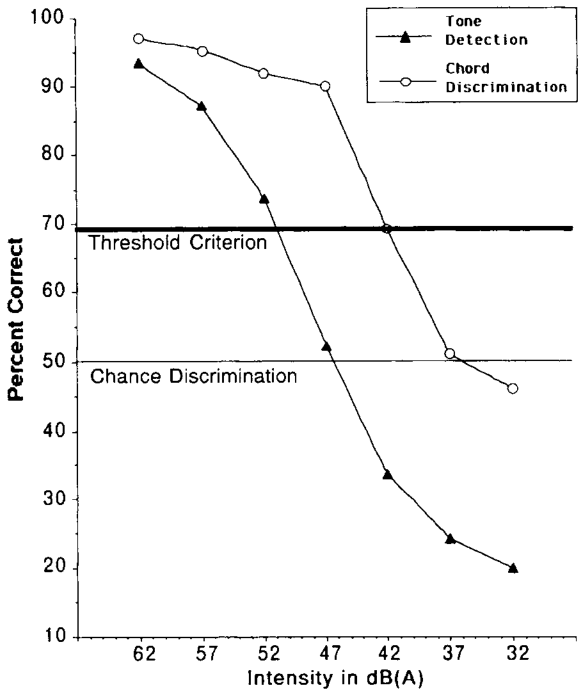

Figure 1. Mean psychometric functions for sinusoidal distinguishing tone detection and major-minor chord discrimination. (The bold horizontal line indicates threshold [69.2% correct] for both measures. The lighter horizontal line indicates chance discrimination performance.)

It seems, then, that major-minor chord perception is maintained at intensities well below that required to perceive sinusoidal distinguishing tones as separate events. Experiment 1 results also suggest that the subliminal, multicomponent, complex percept is maintained over a comparable range of intensities below duplexity threshold for both speech and muscial stimuli.

Two reviewers raised strong concerns about whether the differences in threshold measurements summarized in Figure 2 may be due to criterion effects rather than to actual differences in thresholds. Criterion concerns also apply to the Whalen and Liberman study and for different reasons to the Fowler and Rosenblum study. There really are two criterion hypotheses that could apply to the current study. A detection criterion hypothesis (to be addressed in Experiment 2) would suggest that duplexity threshold reflects a very strict criterion for detection. A discrimination criterion hypothesis would suggest that the discrimination threshold reflects a very lax criterion for discrimination. To account for the magnitude (9 dB), consistency (19 of 20 subjects), and statistical significance ($p < .0001$) of the difference in threshold, both criterion effects would have to be large and present in the form described previously.

The chord discrimination paradigm is a forced-choice task in which there is an external criterion for correctness whose validity is orthogonal to the intensity of the stimulus; the stimuli are physically either the same or different. In the absence of any criterion effect ($\beta = 1.0$), the psychometric functions should exhibit (a) perfect (100% correct) performance when the distinguishing tones are significantly above discrimination threshold, (b) chance (50% correct) performance when the distinguishing tones are significantly below threshold (or are absent, as on catch trials), and (c) a steep transition between these performance extremes as the stimulus magnitude crosses threshold.

A criterion bias for below-threshold stimuli will have equal and opposite effects on measured performance for same and different trials, thus maintaining chance performance. Criterion bias can only alter the middle and upper portions of the psychometric functions (at the extreme, allowing reductions from 100% performance). Therefore, any significant criterion effect can only flatten the psychometric function by depressing performance at the high performance end. A depressed psychometric function will exhibit a somewhat elevated discrimination threshold. Therefore, if subjects in the current study (or in the Whalen and Liberman study) exhibit significant discrimination criterion biases, the magnitude of differences between duplexity and discrimination threshold will have been underestimated. All subjects, however, exhibited the type of psychometric function described for minimal criterion effects. Furthermore, the pooled psychometric function also meets these general characteristics expected for minimal bias.

The current study also provided additional guards against the discrimination criterion hypothesis. Classic catch trials

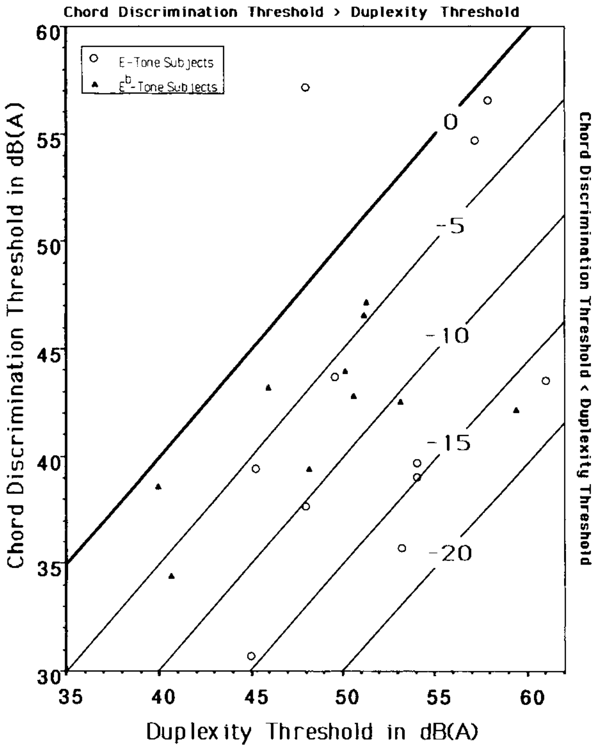

Figure 2. Chord discrimination threshold as a function of duplexity threshold for each of the 20 subjects. (The bold diagonal line, labeled 0, designates equal thresholds. The lighter lines designate the degree to which chord discrimination threshold is lower than duplexity threshold in 5-dB increments. Legend symbols indicate the stimulus used in measuring duplexity threshold.)

present subjects with a null event that should be equivalent to a stimulus that is significantly below threshold; such catch trials can therefore only detect one of two types of biases (Kling & Riggs, 1972). Individual performance on catch trials was nearly perfect, indicating almost total absence of that type of response bias.

We provided another guard against criterion effects by converting the response probabilities for each subject at each intensity into *d′*. We then used a statistical regression of *d′* against intensity to compute the individual discrimination thresholds. Note that if assumptions (equal-variance Gaussian distributions) of the *d′* model are reasonably valid or if the criteria for individual subjects is relatively stable, then *d′* estimates of performance and threshold should be relatively bias-free. Given the nature of the forced-choice task, it is not surprising that we obtained very similar threshold estimates with *d′* and *p*(*c*) psychometric functions.

Experiment 2

This experiment addresses the detection criterion hypothesis raised by reviewers, which challenges the validity of the results from Experiment 1. Direct measurement of detection threshold for the E or E♭ tone is complicated by the major premise in this study. The premise is that $c'(i)$ (the E or E♭ tone) continues to define the qualitative perceptual differences in fused $s'(i)$ stimuli (major or minor chord) at intensities below quantitative (intensity) threshold for detecting the independent presence of the sinusoidal tone.

Better psychophysical procedures are variations of the method of constant stimuli, which include built-in catch trials or null events (i.e., the presence of the piano base in the absence of the sinusoidal tone). For example, in the two-alternative temporal forced-choice (2IFC) procedure, the base would be presented twice on each trail, once with a sinusoidal tone. In this 2IFC procedure, the subject must identify which of the two presentations contained the tone in addition to the base. If the tone continued to fuse with and alter the perception of the base, then the change in perceptual quality from B (a fifth) to $S(i)$ (a major or minor chord) could be the basis for subjects correctly responding on each trial at intensity levels below that necessary to perceive the tone as a separate identity. In other words, the major premise for the current research also limits the psychophysical procedures that can be used to evaluate the duplexity threshold directly. Therefore, the measurement of duplexity threshold in Experiment 1 and in Whalen and Liberman (like the indications of perceptual categories in the Wegel and Lane and the Fowler and Rosenblum studies) relied on the ability of subjects to respond reliably to specific perceptual qualities. The detection criterion hypothesis challenges the validity of this hypothesis. Fortunately, we can provide indirect evidence for the validity of the duplexity threshold measures.

Duplexity threshold is a masked detection threshold with the base acting as a masking stimulus. The critical parameter in masked thresholds is the highest available signal-to-noise ratio (e/n_0) for masker energy in the frequency region (critical band) of the signal. Therefore, we can provide an independent estimate of the masked detection threshold for the tones in the presence of a narrow band noise (one critical band or fewer centered on the signal frequency) whose amplitude matches the peak amplitude of the piano base (approximating an equivalent e/n_0 condition). We can closely approximate n_0 for the narrow band noise with an expected error in n_0 (and therefore measured thresholds) of no more than about 1.5 dB, significantly less than the 9-dB average effect reported in Experiment 1. With this change in masking stimulus, we can use a 2IFC procedure that is bias-free (Green & Swets, 1966).

One reviewer raised a second concern about Experiment 1, arguing that the (subliminal) tones may not have fused with the base to produce qualitatively different chords. Instead, it was conjectured that the subjects could have used perceived differences in the tones to perform the AX task at intensities both above and below duplexity threshold. The possible validity of this hypothesis would be enhanced (or reduced) if the detection criterion hypothesis were demonstrated to be reasonable (or unlikely). In Experiment 1 we had used an AX procedure to avoid the need for highly trained musicians who could both discriminate and identify chords. This second hypothesis is evaluated both indirectly by evaluating the detection criterion hypothesis and directly by asking highly trained musicians to identify the base plus critical tones as major or minor chords. By randomly presenting either an E or an E♭ tone that randomly varies in intensity across trials, full psychometric functions can be generated and thresholds estimated.

Method

Subjects. Seven subjects participated in the masked threshold task. Four of these subjects were members of our laboratory staff who, while participating in a different music experiment (Cho, Hall, & Pastore, 1991), had demonstrated that they could accurately discriminate major and minor chords. The other 3 subjects had limited music experience at best and could not reliably discriminate major and and minor chords. Therefore, only the 4 practiced musicians participated in the chord discrimination task.

Stimuli. The stimuli and the tone intensities were identical to those used in Experiment 1. In the additional masking condition, Gaussian noise was passed through a band-pass filter (24 dB/octave skirts) with cutoff frequencies of 250 and 400 Hz, which roughly correspond to the frequencies of the C and G tones. The noise intensity matched the peak amplitude of the C and G tone pair, thus providing an approximately equal value for n_0.

Procedure. The 2IFC procedure was used in the masked threshold condition. Each trial consisted of the two stimulus intervals, a 500-ms ISI and a 2-s response interval. On each trial, the computer randomly determined both the specific tone intensity and which of the two intervals contained the tone. Subjects were presented each tone intensity 20 times. The identical sample of noise was presented in the two intervals on each trial, with the tone physically mixed with the masking noise by using a solid-state mixing amplifier. Subjects were required to indicate which interval contained the sinusoidal tone.

The chord-labeling (discrimination) task consisted of a single stimulus presentation plus a 2-s response interval. In this task, the computer randomly determined whether to present the E or E♭ tone on each trial and the intensity of that tone. Each tone was again mixed physically with the piano base. The subjects were required to label whether the full stimulus was perceived as a major or minor chord. In both tasks, subjects were required to respond on every trial (guessing if necessary) by using two response buttons. In a brief separate task, the subjects were asked to label the isolated tones (presented at maximum intensity) as major or minor.

Results and Discussion

The results of the two conditions are summarized in Figure 3. The two functions on the left of Figure 3 are the masked tone-detection psychometric functions for all 7 subjects and for the 4 subjects who participated in the chord-labeling condition. The two nearly identical functions indicate a masked threshold of 50 dB(A), which closely matches the duplexity threshold of 51 dB reported in Experiment 1. (In fact, the $\sqrt{2}$ factor required to equate *d'* for yes–no and 2IFC procedures accounts for the 1-dB difference in threshold.)

Three of the 4 musically trained subjects were able to reliably attribute isolated full-intensity E and E♭ tones to C major and C minor chords on the basis of their knowledge of the structure of these chords. The 4th subject was at chance in labeling the isolated tone.[7] The 4 obtained chord-labeling functions were independent of tone-labeling ability. Therefore, it is theoretically possible that 3 of these 4 subjects could use perception of the isolated tones as a basis for chord labeling at tone intensities above duplexity threshold. It is improbable, however, that these 3 subjects could continue to use this separate information at subliminal intensities.

We used the chord-labeling results to compute *d'* for each subject at each tone intensity. Individual psychometric functions for the 4 subjects were highly similar (and thus independent of the ability to label suprathreshold tones). The individual *d'* measures were averaged for each intensity and then converted back into percentage correct to produce the average psychometric function shown in Figure 3. This average function indicates a chord identification threshold of 32 dB (which is 10 dB lower than the average 42-dB threshold from Experiment 1; see Figure 1 for comparison). This significantly lower chord discrimination threshold simply reflects the use of highly practiced musicians. Therefore, in Experiment 2 chord discrimination threshold is 18 dB lower than the masked detection threshold, exceeding the largest difference found for any subject in Experiment 1.

Experiment 2 therefore provides evidence that the average duplexity threshold reported in Experiment 1 accurately reflects the expected masked detection threshold for the isolated tones. It also provides evidence that chord perception continues at distinguishing tone intensities significantly below the masked detection threshold. Finally, we found that highly trained musicians who can easily discriminate chords have chord discrimination thresholds that are at the extreme of the range of subliminal results reported for any subjects in Experiment 1 and the anecdotal results reported personally by Whalen and Liberman (1987). These findings therefore support the validity of the threshold measurements reported in Experiment 1 and the general pattern of results reported by Whalen and Liberman.

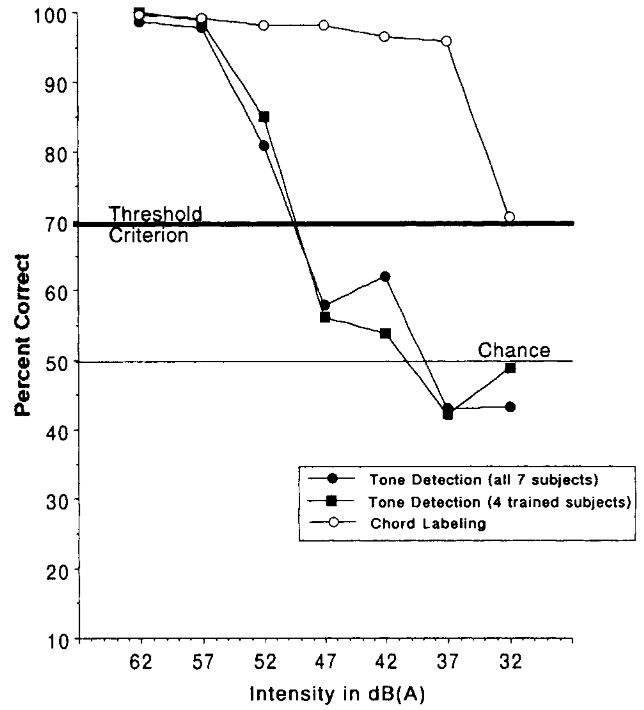

Figure 3. Mean performance on masked detection threshold task at each level of sinusoidal tone intensity, with similar functions for all 7 subjects and the 4 highly trained musicians.

General Discussion

Sinusoidal stimulus components presented at intensities significantly below that necessary for isolated perception can define the perception of ecologically valid stimuli that represent perceptual units in either speech or music. If the Whalen and Liberman findings are accepted as evidence for the existence of a phonetic module that takes precedence over nonspeech perception, then the current findings would require postulation of a music module that functions for DP in a manner and at a perceptual level equivalent to the phonetic module.

Alternatively, we believe that both findings indicate the existence of naturally occurring hierarchical processes of perception in which complex, organized, figurally good, strong stimulus configurations are perceived more readily, and with less specification of stimulus elements, than simpler, less organized stimulus configurations. This theoretical framework is a direct application of extensive research by Gestalt psychologists on visual perception (e.g., Wertheimer, 1958),

---

[7] Five of the 10 Whalen and Liberman subjects also could accurately (90%–100%) apply phonetic labels to the isolated, full-intensity sinusoidal chirps, and only 1 of their subjects was at chance in such labeling. Two of the remaining subjects appear to have reversed their use of labels, with a resulting average chirp-labeling performance of 65%. The consistent ability of some subjects in both studies to respond to the individual, full-intensity stimuli, $c'(i)$, in terms appropriate for the fused stimuli, $s(i)$, raises the possibility that these subjects might not necessarily have responded to the fusion of the stimuli when $c'(i)$ was above duplexity threshold, such as proposed by Nusbaum, Schwab, and Sawusch (1983; see also Repp, 1984). This hypothesis, however, cannot account for the suprathreshold fusion results for the other subjects in both studies, or the equivalence of results across all subjects at subliminal intensities of $c'(i)$.

with more recent application to audition (Bregman, 1990). Gestalt psychology is one of many frameworks that describes a perceptual tendency toward closure, resulting in the perception of good, strong, well-articulated figures. This alternative interpretation treats speech and music as following general principles of perception that give rise to more organized and complete perceptual structures. It does not address the possibility that the perceptual structure may be specific to a given language or music system.

There are several long-standing reasons to expect the major findings reported here for music stimuli and by Whalen and Liberman for speech stimuli. These reasons are really addressed by answering two major questions. First, why does the sinusoidal element combine with the ambiguous base to create a strong perception? Second, why does this combination take place at low intensities of the critical stimulus?

Gestalt psychology describes closure as the tendency to perceive closed (good and strong) figures (Wertheimer, 1958). Research on closure has demonstrated the tendency to perceive figures as being closed even when elements in the physical stimulus are incomplete, missing, or distorted (Bobbitt, 1942). In both the current and the Whalen and Liberman studies, the physical stimuli are based on complex stimuli, $s(i)$, which represent closed figures. This is indicated by the stimuli being perceived in relatively discrete categories, $S(i)$, which are highly familiar and frequently occurring for the subjects. Because the base component, b, has a considerably weaker perceptual organization than the different versions of the complete stimuli, the base is a perceptually open figure. The actual stimuli, $s'(i)$, have a distortion of the distinguishing element, with the distortion due to the absence of harmonic structure. The distinctive property of elements should tend toward separate grouping on the basis of lack of similarity; however, the temporal properties of the stimuli should lead toward grouping with the base because of good continuation (for the speech stimuli) or common fate (for the musical chords). On the basis of the notion of closure, subjects should use information from the distinguishing component to modify the open figure to achieve perception of a good, strong (closed) figure, which is a reasonable representation of the stimulus event. The dissimilarity of elements should also cause the scene analysis system to perceive a separate event. When the distinguishing component is attenuated to a degree at which the scene analysis system is no longer faced with explicit evidence for a separate event, there still should be ample information available to achieve closure. This account of the current findings does not require that we posit the existence of distinct modules, with one module taking precedence over other modules. It only requires that we apply long-standing principles of perception.

General Conceptualization of DP

Bregman (1987, 1990) developed a general theory of auditory perception that provides a reasonable alternative to the modular conceptualization. Prior to any scene analysis, auditory inputs are analyzed in basic terms of periodicity, as well as spatial and modulation properties at local regions of time and frequency. The primitive scene analysis provides links between these lower level analyses on the basis of indications of origin from the same acoustic source. The principle of exclusive allocation does not apply to either basic acoustic analyses or primitive scene analysis. More complex pattern-recognition processes then come into play. These processes are based on schemata that represent "particular knowledge of regularity in the sensory evidence" (Bregman, 1990, p. 637). When such regularity is detected in the available sensory information, that fact is added to the perceptual representation or description being developed. Exclusive allocation refers to the "criteria of consistency or noncontradiction" that constrain the building of perceptual attributes into descriptions, "not to the way in which sensory evidence is used to support the derivation of these attributes" (Bregman, 1990, p. 637).

> Speech (and music) probably are schema driven. However, schemas for speech recognition interact with primitive scene analysis in the same way that other schemas do. . . .When they make use of the information that they need from a mixture, they do not remove it from the array of information that other description-building processes can use. (Bregman, 1990, pp. 637, 638)

Most conditions that result in DP violate constraints in the natural environment and thus can exist only in the laboratory. It is under such unusual circumstances that "the usual heuristics for description formation can work at cross-purposes and odd assignments of properties to sources can occur" (Bregman, 1990, p. 635). These heuristics exhibit the types of properties that were described in one form decades ago by Gestalt psychologists and that in various forms are the basic major principles of nearly all theories of perception. It is our opinion that within this rational framework, DP represents a tool to investigate the nature of these heuristics. With this conceptualization of DP, Ciocca and Bregman (1989) recently demonstrated DP(f) by using second-formant transitions, $c(i)$, and a common base, b, from two-formant synthetic syllables, $s(i)$. They then added simple stimuli that because of their physical properties were expected to stream with the ipsilateral $c(i)$. When $C(i)$ was perceived as part of the ipsilateral stream, there was a consistent and significant reduction in the perceptual fusion of $c(i)$ with the contralateral b. This finding again argues against phonetic processes that are independent of and take precedence over general auditory processing.

Concluding Remarks

We have provided a basis for understanding when DP might occur, but we have not provided a model that explains DP. Stating that DP is the expected result of the application of perceptual principles (such as closure and the grouping or segregation of stimulus elements) is a type of black-box conceptualization. This conceptualization is not any less predictive of behavior than to attribute DP to the operation of independent specialized modules, with one module taking precedence over other modules. A number of perceptual theorists have provided more empirical bases for the principles of perception we have invoked (e.g., Bartley, 1980; Hochberg, 1978; Rock, 1983), but modern psychology has yet to develop an adequate model or theory of perception. Such an adequate model or theory would be required to fully explain DP. Our claim is simply that DP can be found in various forms throughout perception and probably represents the typical application of normal perceptual processes, rather than simply being evidence for a series of distinct open or closed modules for each different type of stimulus.

Because speech and music are distinct classes of stimuli, it is given that the auditory system treats some aspects of these stimuli differently. Specialization for speech, without necessarily implying closed modules, is demonstrated by different results for truly analogous speech and nonspeech stimuli. Processing differences among such stimuli certainly can be described in terms of types and levels of perceptual organization, and if sufficiently distinct might even justify the argument for closed modules. As with the current study, however, many aspects of speech and other auditory stimuli are processed singularly. Only by manipulating the same variables and carefully mapping the perceptual effects for analogous speech and nonspeech stimuli will we understand the nature of auditory perception. This approach eventually will specify the differences between classes of stimuli such as speech, music, and simpler types of sounds.

References

Bartley, S. H. (1980). *Introduction to perception.* New York: Harper & Row.

Bobbitt, J. M. (1942). An experimental study of the phenomenon of closure as a threshold function. *Journal of Experimental Psychology, 30,* 273–294.

Bregman, A. S. (1987). The meaning of duplex perception: Sounds and transparent objects. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 95–111). Dordrecht, The Netherlands: Martinus Nijhoff.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Burns, E. M., & Ward, W. D. (1978). Categorical perception of musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468.

Cho, J. L., Hall, M. D., & Pastore, R. E. (1991). Normalization processes in the human auditory system (Abstract). *Journal of the Acoustical Society of America, 89,* 1988.

Ciocca, V., & Bregman, A. S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics, 46,* 39–48.

Collins, S. C. (1985). Duplex perception with musical stimuli: A further investigation. *Perception & Psychophysics, 38,* 172–177.

Coren, S., & Ward, L. M. (1989). *Sensation and perception* (3rd ed.). San Diego, CA: Harcourt Brace Jovanovich.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83,* 114–140.

Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling transitions. *Canadian Journal of Psychology, 30,* 99–114.

Dewitt, L. A., & Samuel, A. G. (1990). The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General, 119,* 123–144.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* New York: Academic Press.

Fletcher, H. (1972). *Speech and hearing in communication.* New York: Krieger. (Original work published 1940)

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 742–754.

Goldstein, E. B. (1989). *Sensation and perception.* Belmont, CA: Wadsworth.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hochberg, J. E. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kling, J. W., & Riggs, L. A. (1972). *Experimental psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics, 30,* 133–143.

Liberman, A. M., & Mattingly, I. G. (1985). Motor theory of speech perception revisited. *Cognition, 21,* 1–36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489–494.

Mattingly, J. G., & Liberman, A. M. (1988). Speech and other auditory modules. In *Haskins Laboratories status report on speech research* (SR-93/94, pp. 67–84). New Haven, CT: Haskins Laboratories.

Nusbaum, H., Schwab, E., & Sawusch, J. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics, 33,* 323–332.

Pastore, R. E., Li, X.-F., & Layer, J. K. (1989). Categorical perception of nonspeech chirps and bleats. *Perception & Psychophysics, 48,* 151–156.

Pastore, R. E., & Scheirer, C. J. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin, 81,* 945–958.

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics, 33,* 469–474.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55,* 678–680.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–950.

Repp, B. H. (1984). Against a role of "chirp" identification in duplex perception. *Perception & Psychophysics, 35,* 89–93.

Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics, 21,* 399–407.

Snodgrass, J. G. (1975). Psychophysics. In B. Sharf (Ed.), *Experimental sensory psychology* (pp. 17–67). Glenview, IL: Scott, Foresman.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–393.

Wegel, R. L., & Lane, C. E. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physics Review, 23,* 266–285.

Wertheimer, M. (1958). Principles of perceptual organization. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 115–135). New York: Van Nostrand.

Whalen, D., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169–171.

(*Appendix follows on next page*)

Summary of Symbols Used to Designate Stimulus Conditions and Resulting Perceptions in Various DP Demonstrations

Physical stimuli	Perception
$s = b + c$	
$s(i) = b + c(i)$	
s = physical stimulus	S = based on s
$s(i)$ = ith version of s	$S(i)$ = based on $s(i)$
s' = spectral modification	$S-$ = weak version of s
b = base component from s	B = based on b
c = fixed critical component in s	C = based on c
$c(i)$ = critical variable component in s	$C(i)$ = based on $c(i)$
$c'(i)$ = sinusoidal version of $c(i)$	$C'(i)$ = based on $c'(i)$
$c\uparrow$ = amplified version of c	
$c\downarrow$ = attenuated version of c	
$c\downarrow(i)$ = attenuated version of $c(i)$	
n = noise stimulus	N = based on n
e = average signal power per unit bandwidth	
n_0 = average noise power per unit bandwidth	
e/n_0 = signal-to-noise ratio	

Received February 13, 1991
Revision received September 9, 1991
Accepted September 11, 1991 ■