Perception & Psychophysics
1996, 58 (6), 857–870

Limits on phonetic integration in duplex perception

D. H. WHALEN and ALVIN M. LIBERMAN
Haskins Laboratories, New Haven, Connecticut

The telling fact about duplex perception is that listeners integrate into a unitary phonetic percept signals that are coherent from a phonetic point of view, even though the signals are, on purely auditory grounds, separate sources. Here we explore the limits on the integration of a sinusoidal consonant cue (the $F3$ transition for [da] vs. [ga]) with the resonances of the remainder of the syllable. Perceiving duplexly, listeners hear the whistle of the sinusoid, but also the [da] and [ga] for which the sinusoid provides the critical information. In the first experiment, phonetic integration was significantly reduced, but not to zero, by a precursor that extended the transition cue forward in time so that it started 50 msec before the cue. The effect was the same above and below the duplexity threshold (the intensity of sinusoid in the combined pattern at which the whistle was just barely audible). In the second experiment, integration was reduced once again by the precursor, and also, but only below the duplexity threshold, by harmonics of the cues that were simultaneous with it. The third experiment showed that the simultaneous harmonics reduced phonetic integration only by serving as distractors while also permitting the conclusion that the precursor produced its effects by making the cue part of a coherent and competing auditory pattern, and so "capturing" it. The fourth experiment supported this interpretation by showing that for some subjects the amount of capture was reduced when the capturing tone was itself captured by being made part of a tonal complex. The results support the assumption that the independent phonetic system will integrate across disparate sources according to the cohesive power of that system as measured against the evidence for separate sources.

According to a conventional theory of speech, there is, at the level of perception, no phonetic mode, hence no difference in primary perceptual representation between phonetic and auditory processes (Crowder & Morton, 1969; Diehl & Kluender, 1989; Kuhl, 1981; Miller, 1977; Schouten & Hessen, 1993; Stevens & House, 1972). On this view, the speech signal engages the ordinary mechanisms of the auditory modality, evoking a representation that is formed of the usual auditory primitives. It is, then, only at a second, cognitive stage that this purely auditory representation is marked as phonetic and so made available to the language system. This translation from auditory to phonetic is achieved by attaching the auditory representation to a phonetic name, fitting it to a phonetic prototype, or associating it with the distinctive features that linguists have taken as the true primitives of the phonologic system. On any view, however, the assignment of speech signals to sources and locations, which is our particular concern, must occur at the level of the primary representation, not at some subsequent cognitive stage. Such assignment must, therefore, be carried out by the auditory processes of scene analysis, the primary processes that assign sounds to sources according to the common auditory criteria of comodulation, common fate, interaural differences of time and intensity, and so forth (Bregman, 1990). The second-order, cognitive processes that lead, on the conventional view, to a phonetic conclusion must necessarily follow the dictates of scene analysis. Thus, information that is perceived as arriving simultaneously from two sources should not be combinable into a single phonetic unit.

Our less conventional view is that the primary perceptual representations in speech are not auditory, but phonetic (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Whalen & Liberman, 1987). These representations are evoked immediately by a specialization for language—a phonetic module—that responds to information about phonetically significant gestures of the vocal tract, which it represents to consciousness as primitives that are, ab initio, distinctly phonetic; unlike the auditory percepts of the conventional view, they do not need to be given communicative significance by cog-

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. We thank Diana Matson, Mark Pitt, and Yi Xu for assistance with the experiments. Portions of Table 1 reprinted with permission from D. H. Whalen and Alvin M. Liberman, "Speech Perception Takes Precedence Over Nonspeech Perception," *Science*, 237, 169–171. Copyright 1987, American Association for the Advancement of Science. Portions of this work were presented at the 121st meeting of the Acoustical Society of America, Baltimore, May 1991. We thank Andrea G. Levitt, Yi Xu, Carol A. Fowler, Ignatius Mattingly, Michael Studdert-Kennedy, Bruno H. Repp, Chris Darwin, and an anonymous reviewer for helpful comments. We thank our colleagues at Haskins Laboratories for their willingness to serve as subjects in Experiment 4. Correspondence should be directed to D. H. Whalen, Haskins Laboratories, 270 Crown St., New Haven, CT 06511-6695 (e-mail: whalen@haskins.yale.edu).

nitive translation into units of a phonetic sort. Thus, perception of phonetic structure takes place in a distinct phonetic mode, unmediated by auditory representations and independent of the processes that evoke them. In that case, the phonetic module might well have its own, specifically phonetic criteria for determining what counts as one representation and what counts as two—that is, it might be able to form a phonetic representation on the basis of phonetic coherence alone, without regard for the auditory considerations that scene analysis normally takes into account. In fact, there is evidence from several kinds of experiments that the phonetic system can do just that.

One relevant class of experiments employs sine-wave speech, so called (Remez & Rubin, 1984, 1990; Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981). In these experiments, the formants of a speech signal are replaced by frequency-varying sinusoids that follow the formant centers, so there is no commonality that might provide auditory coherence, hence no basis on which listeners might assign the sinusoids to a common source; they should be heard, rather, as three distinct, continuously varying pitches. However, the sinusoids preserve a significant amount of purely phonetic coherence, because they provide information about the phonetic structure of the signal, which we take to be the articulator trajectories. On our view, that is the information the phonetic module is specialized to use, so it should engage the module, even though that requires integration across disparate sources. Given sinusoids that follow the centers of the formants, listeners do, in fact, perceive phonetic structure. But they also perceive, at the same time, the continuously changing and seemingly disparate dissonances that would be expected as the normal responses of auditory scene analysis. Apparently, the sinusoids simultaneously evoke representations in the auditory and phonetic systems, producing a phenomenon identical in principle to one that has been called "duplex perception" and extensively investigated with very different kinds of stimulus arrangements.

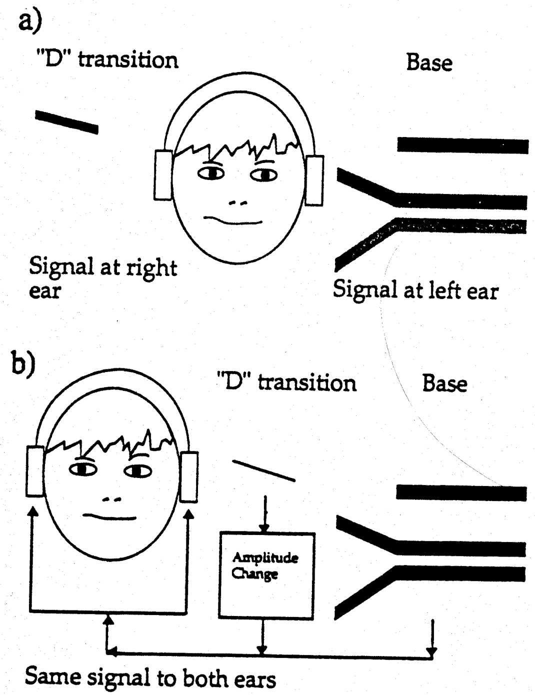

Figure 1. (a) A typical duplex experiment paradigm. The subject hears the syllable without the $F3$ transition (the "base") in one ear, while the $F3$ transition alone (the "excerpt") is presented to the other. (b) Duplex paradigm of the current experiments. The amplitude of the transition is manipulated independently of the base's, and then the two are combined and presented to both ears.

To produce the kind of duplex perception that has been more commonly studied, the experimenter divides a synthetic syllable pattern into two parts (Bentin & Mann, 1990; Ciocca & Bregman, 1989; Mann & Liberman, 1983; Nygaard & Eimas, 1990; Rand, 1974; Repp, Milburn, & Ashkenas, 1983). One part, which we will call the excerpt, is some portion (say, the 50-msec third-formant transition) that can determine, critically, which of two syllables (say, [da] or [ga]) a listener hears (see Figure 1a). When sounded by itself, this excerpt is not heard as speech. That is, it does not engage the phonetic module; being responded to by the standard auditory modules, it is heard, rather, as a smoothly changing timbre or chirp, exactly as everything we know about auditory perception would lead us to expect. The remainder of the pattern, which we will call the base, has fixed first and second formants appropriate to an initial stop, exactly as the full pattern has, but it lacks the critical third-formant cue that would, in this context, produce [da] in the one case, [ga] in the other: the consequence is perception of a syllable with an ambiguous initial consonant, or as one or the other of [d] or [g]. Next, the experimenter introduces a discrepancy or discontinuity between the excerpt and the base by presenting the excerpt over headphones at one ear, the base at the other (it can be done as well with loudspeakers at either side of the head), thus creating a sudden (and ecologically impossible) shift in the location of the third formant when the 50-msec third-formant transition has run its course. We will refer to this as the dichotic kind of duplex perception. Not surprisingly, such presentation of excerpt and base causes the listener to hear two sounds, one at each ear. At the ear that got the excerpt, the auditory modules assign their primitives to the source that is physically located there. The consequence is that listeners hear the same nonspeech chirps they had heard when the excerpts were presented in isolation. At the other ear, however, the listeners do not hear the ambiguous base that was, in fact, the only stimulus at that location, but rather an unambiguous [da] or [ga], a syllable that could only have been formed by integrating information from the base with information from the excerpt that is, at the same time, being perceived at a different location as a nonspeech chirp. Thus, the same piece of sound (the excerpt) is perceived simultaneously by the same listener at two places and in two completely different ways. Apparently, the phonetic system ignored or overrode the processes that represented two sources at two locations, responding, instead, to an exclusively phonetic coherence that existed across them.

That the phonetic part of the duplex percept is a genuinely primary representation, and not the result of a cognitive translation, follows from four critical observations: (1) Perception is demonstrably duplex, not triplex. That is, listeners hear only the unambiguous syllable and the chirp; they do not—indeed, cannot—also hear the ambiguous base (Repp et al., 1983). (2) Listeners perceive the integrated [da]/[ga] syllable and the chirp simultaneously, not one or the other, as in the familiar cases of reversible visual percepts. (3) Perceiving both sides of the duplex representation is mandatory; listeners cannot but perceive both and at the same time. And (4) discrimination functions for the two sides of the duplex percept are radically different in shape (Mann & Liberman, 1983). This shows that listeners cannot penetrate the integrated phonetic percept so as to divide it into syllabic base and chirp (as they could if the percept were a cognitive combination of the two), nor can they perceive (and discriminate) the chirps as [da] and [ga]; for if they could do either, they would, on each trial, have responded to the more readily discriminable form, and thus produced a single discrimination function. In fact, the phonetic side of the duplex percept proved more discriminable than the auditory side over part of the stimulus range and less discriminable over another part.

There are claims in the literature that duplex perception can be obtained with a variety of nonspeech acoustic patterns (Bregman, 1987; Fowler & Rosenblum, 1991; Hall & Pastore, 1992). However, in only one case (Fowler & Rosenblum, 1990) were any of the aforementioned tests applied; having tested for manditoriness of the duplex percept, those researchers found that the speech case passed while the nonspeech analogs did not, indicating that the apparently duplex result was simply a cognitively mediated combination of each of the two inputs, not a true perceptual integration. We hasten to say, however, that we do not therefore think that duplex perception is unique to speech. On the contrary, we should expect that it might appear wherever distinctly different modules can be made to respond to the same dimension of the stimulus, as in the response of the several components of the visual system to binocular disparity. There, disparity beyond a certain point produces the perception of depth by the specialization for stereopsis, but also, and at the same time, the double images that are evoked by other independently acting parts of the visual system (Richards, 1971).

In the experiments on duplex perception of speech, it is as if the phonetic and auditory modules were somehow sharing the information in the stimulus. Looking, for example, at the duplex percepts obtained when the critical $F3$ transition and base are presented as separate sources, we should conclude that they represent a resolution in which the information is somehow divided between two independent systems: The phonetic module uses the transition to represent a coherent phonetic percept that could only have resulted from an integration across the two sources, while, at the same time, the auditory system represents that same transition as a source distinct from the base and, in proper auditory fashion, as a nonspeech chirp. This demonstrates, at the least, that the phonetic system can ignore the results of the scene analysis module and must, therefore, be capable of independence from it, as the auditory system is not. By the same token, it demonstrates, as Mattingly and Liberman have suggested, that the phonetic module is elastic in that it can, within limits, respond in a phonetically appropriate way even when the relevant information is in an ecologically impossible form. But, surely, there are limits to that elasticity. Thus, there must be a point at which scene analysis defeats the cohesive power of the phonetic module. For the general case, then, we should suppose that the information in the excerpt will be divided between phonetic and auditory modules, depending on something like the weight of evidence for separate sources.

In an earlier experiment (Whalen & Liberman, 1987), which is the basis for the one to be reported here, we undertook to find how the division was, in fact, affected by variations in the evidence for separate sources. First, we contrived a case of binaural duplex perception, as contrasted with the dichotic variety described earlier. For that purpose, we made the base of normal resonances and the excerpt ($F3$ transitions for [d] and [g]) of frequency-varying sinusoids (Figure 1b). (In isolation, the frequency-varying sinusoids sound like whistles, and could not be identified by our subjects as "d" or "g.") Thus, the base and excerpt differed in everything that might produce auditory coherence. They were grossly different in their fundamental frequencies and in their spectral structures (the resonances changed position on the spectrum by moving through a harmonic series, the excerpts by frequency change). The perceptual consequence was that over a rather wide range of intensities of excerpt relative to base, listeners gave evidence of a duplex percept—that is, they correctly discriminated both the stop-vowel syllables and the whistles, which they reported hearing simultaneously. Thus, in this binaural arrangement, as in those done dichotically, there was integration of phonetic information across disparate sources, which is to say that here, too, the phonetic system ignored or overrode the results of scene analysis. But the main purpose of the experiment was to observe the effects of varying the strength of the evidence that there were, in fact, two separate sources. For that purpose, we varied the intensity of the sinusoidal transition relative to the base. The result was that there was, for the listeners, an intensity of sinusoid (the "duplexity threshold") below which they discriminated [d] and [g] accurately, but not the whistles, and above which they discriminated both. (But see Bailey & Herrmann, 1993.) This result has been replicated with formant transitions over a wider range of intensities by Vorperian, Ochs, and Grantham (1995). At very high intensities, these authors found that the speech percept deteriorated. Additionally, changing the fundamental frequency of those transitions reduced their phonetic effectiveness, again showing that a variety of acoustic changes can stress the phonetic system beyond tolerable bounds.

Results just like ours have also been obtained in duplex perception of the more commonly used dichotic form. Thus, Bentin and Mann (1990) found that, at very low intensities of the $F3$ transition, listeners heard [da]/[ga] at one ear, but could not discriminate the chirps at the other; at somewhat higher intensities of the transition, they heard duplexly—that is, [da] or [ga] at one ear, discriminable chirps at the other. In an experiment that tested for duplex perception in infants 2–4 months of age, Eimas and Miller (1992) determined that their infants could not discriminate excerpts of very low intensity when they were presented in isolation at one ear, but discriminated very well indeed when the base was presented simultaneously at the other. The strong implication was that, like adults, the infants were using the low-intensity excerpts to form the discriminable phonetic side of a duplex percept at one ear while at the same time failing to discriminate the chirps. More evidence for the same kind of effect comes from an unpublished study by Bentin and Repp (1986). There it was found that the absolute threshold for discriminating the excerpts in isolation was exactly equal to the absolute threshold for discriminating the [da]/[ga] contrast on the phonetic side of the duplex percept; however, the threshold for discriminating the excerpts as chirps in the same duplex percept was significantly higher. Thus, it appears that the division of information in the sinusoid depends on the cohesive power of the phonetic module in relation to the evidence for separate sources; and further, that the phonetic system takes "precedence" in that the transition cue is successfully used for phonetic purposes at levels of intensity not high enough for it to be represented by the auditory system as a separate source.

Returning now to limits on the elasticity of the phonetic system, we observe that there surely are stimulus arrangements for the duplex paradigm in which phonetic integration is defeated because the excerpt ($F3$ transition) is "captured" as part of an alternative, nonphonetic pattern. Thus, Ciocca and Bregman (1989) succeeded in reducing the extent to which the excerpt collaborated in a duplex percept by making it part of an aurally coherent stream, consisting of repetitions of the excerpt before and after the duplex syllable. However, their manipulations succeeded only in reducing the phonetic effect; there remained, in all of their conditions, a substantial amount of duplex perception—that is, integration of excerpt and base across the two sources.

Similar success in defeating the phonetic module was achieved by Darwin and Sutherland (1984). Having preceded a synthetic vowel by a single sinusoid that matched the frequency of one of the harmonics of the vowel, they found that the harmonic no longer contributed to the perceived color of the vowel. They were able to obtain this effect with a preceding sinusoid as short as 32 msec in duration. Interestingly, the capture effect of the preceding sinusoid was greatly reduced by being paired with one of its own harmonics, provided the harmonic ended when the vowel began. The investigators did not provide direct evidence that the "captured" harmonic fused with—that is, became part of—the preceding sinusoid, only that its contribution to the phonetic percept was nullified. A similar effect of capturing a tone was found by Ciocca and Darwin (1993) for pitch judgments. More recent experiments by Darwin (1995) found that the amount of the sinusoid that is captured depends on the intensity of the precursor sinusoid. For a fairly broad range of intensities of precursor sinusoids, the more intense versions reduced the phonetic effect of the captured sinusoid more than the less intense versions. For extremely short (56 msec) vowels, having the sinusoid continue after the vowel had an additional capturing effect. For longer vowels, following tones were ineffective. This indicates that the mechanism responsible for the capture is not an all-or-nothing process in which any evidence of a competing signal will capture all of the speech signal that is related to it; rather, the degree of capture depends on the strength of the competing signal.

The experiments we report here are designed according to the following considerations: (1) The experiment by Darwin and Sutherland (1984), in which a simple preceding stimulus caused capture of a stimulus element from the phonetic system, was carried out on a vowel, the perception of which was not duplex and therefore did not require phonetic integration across disparate sources. Is a correspondingly simple stimulus arrangement equally effective in duplex perception of a stop consonant? (2) Reductions in the phonetic integration observed in the aforementioned experiment, and also in the experiment by Ciocca and Bregman (1989), were obtained with stimuli that preceded the captured element in time. Is it temporal priority that is crucial, or can the capturing stimuli be simultaneous with the element that is captured? (3) Does the reduction or elimination of phonetic integration result from capture by the nonspeech patterns or from the distraction they provide? (4) Are the effects of the potentially capturing patterns the same above and below the duplexity threshold?

EXPERIMENT 1

The first experiment is an extension of our earlier study (Whalen & Liberman, 1987), described in the introduction, which demonstrated phonetic integration across sources in binaurally presented stop-vowel patterns. We here report a condition—run at the same time as that study—that was designed to determine whether it was possible to defeat such integration by beginning the critical stop-consonant cue before the onset of the syllable. Our aim was to see if the earlier-starting, "precursor" signal would "capture," and thus reduce, the consonantal cue. We tested these effects above and below the duplexity threshold, because a longer, more coherent pattern (above the threshold) could be expected to provide stronger competition for the phonetic system. That is, if the transition is audible on its own, it should be better able to merge completely with the precursor and so fail to inform the phonetic judgment.

Method

Stimuli and Equipment. Each stimulus consisted of two components, a fixed "base" composed of resonances (see the rightmost part of Figure 1b) and a frequency-varying sinusoid making up the transition of the third formant (the "sinusoid"; see the middle part of Figure 1b). Because it lacked the critical *F*3 transition, the base by itself was ambiguous between [da] and [ga]. The sinusoids, by themselves, were heard as whistles that bore no relation to speech and could not be correctly associated with either consonant. As in the earlier experiment, the acoustic (and auditory) difference between the resonances of the base and the sinusoids of the transitions was intended to provide clear evidence for distinct sources, and thus to test for phonetic integration across them. In the earlier experiment, we had found that, when combined at low intensities of the sinusoid, the stop-vowel syllable determined by the sinusoid was heard. At higher intensities of the sinusoid, listeners correctly perceived both the stop-vowel syllable and the whistle. That is, over a range of intensities of the sinusoid, listeners perceived duplexly.

To see if phonetic integration could be defeated, we created a second set of stimuli, identical to those just described, except that the sinusoidal *F*3s were extended forward in time by 50 msec, the duration of the transitions themselves (see Figure 2b). This duration was longer than the 32 msec that Darwin and Sutherland (1984) found sufficient to "capture" a vowel's harmonic. We will call the first set of stimuli the "syllable-only" set, and the ones with the extended *F*3 the "precursor-tone" set.

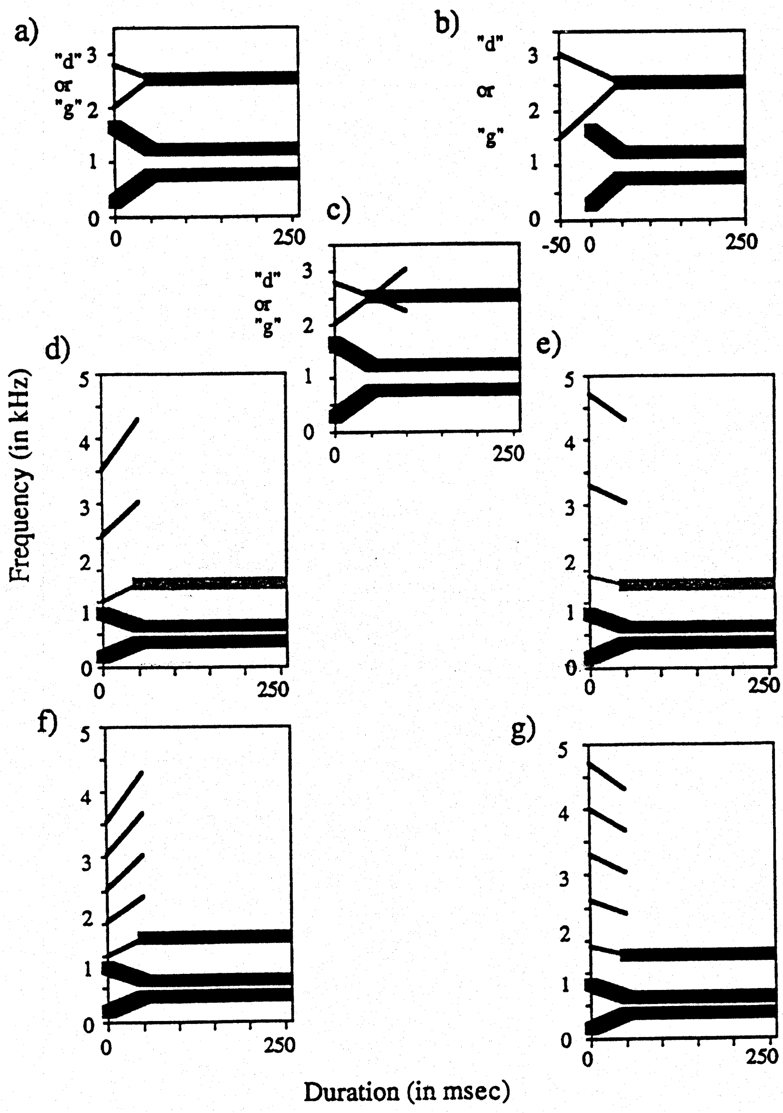

Figure 2. Schematic spectrograms of the stimuli for Experiment 2 (note that some of these are also used in other experiments). (a) The syllable-only stimuli. (b) Stimuli with preceding sinusoid. (c) Stimuli with continuation sinusoid. In these three cases, both transitions are shown together, though only one would be presented on any one trial. (d) Three-harmonic stimulus for [ga]. (e) Three-harmonic stimulus for [da]. (f) Five-harmonic stimulus for [ga]. (g) Five-harmonic stimulus for [da].

The base was created on a software parallel-resonance synthesizer and sounded like typical synthetic speech in which the cues have been pared to a minimum, especially those composed of fixed vowel formants and linear consonant transitions. It contained full first and second formants, together with a third formant from which the initial transition had been removed. $F1$ and $F2$ had onsets of 279 and 1650 Hz, respectively, and moved linearly to their steady states of 765 and 1230 Hz in the course of 50 msec. The $F3$ began 50 msec after $F1$ and $F2$, and was at a constant 2527 Hz. $F0$ was 100 Hz for the first 100 msec, then fell linearly to a final value of 80 Hz. The full patterns were 250 msec in duration.

The sinusoids were created with a software synthesizer (SWS, written by Philip E. Rubin at Haskins Laboratories) designed to allow sinusoidal tones to vary in frequency and amplitude, with values updated every sample. The [da] $F3$ began at 2800 Hz and dropped linearly to 2527 Hz (the $F3$ steady state) in the course of 50 msec. The [ga] $F3$ began at 2018 Hz and rose to the same 2527 Hz value, also in 50 msec. The precursor tones were 100 msec in duration and began at 3073 and 1509 Hz for [d] and [g], respectively, both ending at 2527 Hz. All tones had the same (arbitrary) input amplitude, but they were allowed to pass through the deemphasis filter of the PCM system (Whalen, Wiley, Rubin, & Cooper, 1990), which reduced the beginning of the [d] transition by approximately 2 dB more than the beginning of the [g] transition.

On each trial, the base and one of the sinusoids (either syllable only or precursor, with transitions appropriate for either [d] or [g]) were output through synchronized PCM channels (Whalen et al., 1990), the sinusoid was attenuated via an analog potentiometer, and the two signals were combined. This single signal was then presented binaurally (with equal intensities at the two ears) over TDH-49 headphones for identification of the consonant as "d" or "g."

Procedure. A screening test was employed to ensure that subjects were able to perceive the stylized synthetic syllables as intended. For that purpose, 10 repetitions of each of the full-formant versions of the syllables ("da" and "ga") were presented in randomized order. Failure to correctly identify at least 85% of the tokens resulted in a subject's exclusion from the rest of the study.

For those who passed the screening, we then determined the level at which each of the sinusoids was just barely audible as a whistle in the presence of the base. To this end, the subjects adjusted the intensity of the sinusoid via an analog potentiometer during successive presentations, 2.5 sec apart, until they judged they could just hear the whistle along with the syllable. The subjects were allowed to go above and below the threshold during this determination. The threshold was measured three times for each transition, and the mean was taken as the "duplexity threshold." The [d] transition became audible at lower intensities (−6.4 dB relative to the intensity of the $F3$, averaged across subjects) than the [g] transition (0.0 dB).

To test the effect of the precursors above and below the duplexity threshold, we took account of the different thresholds for the [d] and [g] sinusoids by selecting values for the above- and below-duplexity conditions such that both whistles would be heard in the former and neither in the latter. Accordingly, the attenuation level was set at 4 dB below the [d] sinusoid threshold for the below-duplexity condition and 6 dB above the [g] sinusoid threshold for the above-duplexity condition. This attenuation was applied to the transition and, equally, to the precursor tone when present. Thus, the precursors were less intense in the below-duplexity condition than in the above, but they matched the transition in intensity. Precursors were, by the evidence of our own listening, audible in both conditions. The below-duplexity condition was run first, and the above-duplexity condition, second.

In the original conditions reported in Whalen and Liberman (1987), we measured identifiability of the stop-vowel syllables above and below the duplexity threshold. That was in what we have here called the "syllable-only" condition. The conditions newly reported here (which were run in the same session as the previous one) assessed the identifiability in the "precursor" condition. In this condition, as in the earlier one, there were 20 presentations of each stimulus ([d] and [g]) in each stimulus set (syllable only or precursor tone) in each condition (below and above duplexity). And also, as in the earlier experiment, the interstimulus interval was 2.5 sec, with 6 sec after every 10 items, corresponding to the end of a line on the answer sheet. So, in the results to be shown in Table 1, the syllable-only conditions were reported in the earlier paper; the precursor conditions are reported here for the first time.

Table 1
Percent Correct on the Phonetic Identification Task for Two Stimulus Sets in Two Conditions, Experiment 1

	Condition			
	Below-Duplexity Threshold		Above-Duplexity Threshold	
Subject	Syllable Only	Precursor	Syllable Only	Precursor
1	100.0*	60.0	100.0*	—
2	100.0*	52.5	97.5*	47.5
3	100.0*	100.0*	100.0*	92.5*
4	100.0*	60.0	100.0*	72.5*
5	97.5*	100.0*	100.0*	52.5
6	92.5*	57.5	85.0*	75.0*
7	82.5*	65.0*	97.5*	52.5
8	52.5	55.0	100.0*	85.0*
9	100.0*	72.5*	97.5*	50.0
10	100.0*	60.0	100.0*	50.0
M	92.5*	68.3*	97.8*	64.2*
SEM	±4.8	±5.6	±1.5	±5.7
t	—	3.90	—	5.47
p		<.01		<.001

Note—The *t* test compares the syllable-only values with the precursor values. For the below-duplexity condition, there are 9 *df*, while the above-duplexity condition has 8. *Percentages that are individually better than chance.

Subjects. The subjects were the same 11 undergraduate students at Yale University used in the earlier-reported study. All said they had no hearing problems. One failed the screening test and was excluded. Due to circumstances unrelated to the experiment, 1 subject was unable to serve in the above-duplexity condition for the precursor stimuli. His results in the other conditions will be reported.

Results

As shown in Table 1, there was significant phonetic integration across the two sources—sinusoidal transition and the resonances of the base—above and below the duplexity threshold for both the syllable-only and the precursor stimuli. As previously mentioned, the syllable-only results were reported in Whalen and Liberman (1987). What is new here are the data for the precursor stimuli, and they show that the precursors did significantly reduce the amount of phonetic integration, but not completely; even with the precursors, there was some phonetic integration across the sources. As for the results above and below the duplexity threshold, the amount of reduction in phonetic integration was about the same.

Discussion

Our results with consonants, binaural presentation, and a simple precursor closely parallel those obtained by

Ciocca and Bregman (1989) with consonants, dichotic presentation, and a complex precursor. Both sets of results differ, though only in magnitude, from those obtained by Darwin and Sutherland (1984) with vowels, binaural presentation, and a simple precursor: In the vowel study, the phonetic effect of the target stimulus was nullified, while in the consonant studies, the phonetic effect was significantly reduced, but not to zero. There are several possible reasons for the difference, although the current state of our knowledge may not allow us to choose which is correct.

We should first consider not only the difference between the vowels of the one study and the consonants of the other two, but also the fact that the vowels were steady state. Such vowels are not representative of the vowels that occur in speech, and it is possible that, being only marginally phonetic, they can easily be perceived in an auditory, as opposed to a phonetic, way. Surely, that assumption is consistent with the fact that perception of steady-state vowels tends very little toward the categorical (Liberman et al., 1967; Repp, 1984; Schouten & Hessen, 1993), and there is little, if any, of the right-ear advantage that characterizes stop consonants and implies left-hemisphere processing (Shankweiler & Studdert-Kennedy, 1967). One might expect, then, that the purely auditory precursor would "blend" more readily with the correspondingly auditory harmonic of the vowel, and so more readily capture it.

Interpretation of the difference between the vowel results and those obtained with the consonants must also take into account the presence, in the one case, and the absence, in the other, of phonetic integration across disparate sources. In the vowel study, the captured cue was one harmonic of the harmonically coherent series that formed the original vowel. Thus, all components of the stimulus were harmonics of a common fundamental, hence a single source from a purely auditory (as well as phonetic) point of view; phonetic perception was not required, as it was in the consonant studies, to ignore an auditory disparity. We hardly know what to make of this difference, except to note about our consonant study that the precursor was, from an acoustic (and auditory) point of view, the same source as the critical sinusoidal cue it was to capture, hence distinct from the resonances that had to be integrated with the sinusoid if a proper phonetic percept was to be formed. However, such a difference would lead us to expect, by comparison with the vowel study, a greater degree of capture, rather than the lesser degree we found.

Finally, we must consider a difference between the two kinds of experiment in the durations of the several parts of the stimulus. In the case of the vowels, it had been found that short precursors were effective with short vowels, while, with longer vowels, longer precursors were needed (Hukin & Darwin, 1995). In our experiment, the critical sinusoidal cue was 50 msec in a total syllable of 250 msec. However, the vocalic part of the syllable was steady state and the same for both consonants, so it made no contribution to the consonant distinction. It might be more appropriate, therefore, to consider that the precursor of 50 msec was exactly as long as the sinusoidal cue it was supposed to capture. In the vowel experiment by Darwin and Sutherland (1984), a short precursor (32 msec) was just as effective as a long one (240 msec) in capturing the vowel's harmonic (which was 56 msec in duration). But here again one might have expected the difference in the experiments to have produced a larger, not a smaller, effect in our experiment with the consonants.

All effects in our experiment were similar whether the sinusoids were presented above the duplexity threshold or below it. This seems, perhaps, a little strange, given that the precursor was an average of 16.4 dB more intense in the one condition than in the other. Darwin (1995) had found that more intense precursors captured more of the signal than less intense ones did. However, the signal in the vowel was of constant intensity, while our sinusoidal transitions were at the same intensity as the precursor, whatever that might have been for a particular condition. It seems likely that the match of the precursor's amplitude to that of the tone to be captured is important, but that cannot be known on the basis of the data so far available.

EXPERIMENT 2

The nonphonetic precursors of Experiment 1 that reduced the phonetic effectiveness of the sinusoidal transitions were continuous with those transitions and preceded them in time. Apparently, those precursors provided the listeners with an alternative, nonphonetic pattern of which the sinusoidal transitions might be a part. Indeed, on the basis of source characteristics, the transition ought to have been only a part of the sinusoidal tone. The purpose of this experiment was to see if such a reduction in phonetic effectiveness could be obtained with correspondingly simple signals that might include the transition as part of a nonphonetic pattern but not precede it in time. To that end, we thought it would be reasonable to make the sinusoidal transitions part of a harmonic series, because such series are known to form a coherent source from an auditory point of view. In addition, Darwin and Sutherland (1984) have provided compelling evidence of the power of harmonic relationships in experiments like those being pursued here: When they preceded the vowel not just by the precursor tone itself, but by that tone plus its second harmonic, the capturing of the vowel's harmonic by the precursor was reduced to nothing. That is, the capturing tone was itself captured by one of its harmonics. We thus had reason to think that making a harmonic complex on the basis of the transition was a reasonable way to make the cue part of a coherent auditory—that is, nonphonetic—pattern.

As another way to reduce phonetic integration of sinusoidal transition and base, we thought it reasonable to extend the sinusoid beyond its normal end. Darwin and

Sutherland (1984) found a capturing effect even by following tones, if the vowel was short. The critical information in the present stimuli was equivalently short (50 msec), although the syllable did continue beyond that point. In our case, therefore, the following tone would overlap the syllable, but it might nevertheless be expected to capture the transition.

Method

Stimuli and Equipment. The stimuli included those of Experiment 1 and three new sets. All of the new stimuli contained more sinusoids, and these were created with the synthesizer used before. One set, the "continuation" sinusoids, comprised straight-line extensions of the *F*3 transitions 50 msec beyond the end of the original *F*3 transition and at the same amplitude (Figure 2c). Another, the "three-harmonic" sinusoids, took each *F*3 transition as a fundamental frequency and added the two next higher harmonics (Figures 2d and 2e). Those added harmonics were coextensive with the transitions and had the same input amplitude; they were, however, passed through the deemphasis filter of the PCM system, which reduced them by approximately 7 dB. They also were attenuated by the same system used to set the transitions above or below the duplexity threshold. For the third set, the "five-harmonic" sinusoids, the *F*3 transition was treated as the second harmonic rather than the fundamental (Figures 2f and 2g). Thus, they consisted of the second through sixth harmonic of a (missing) fundamental frequency corresponding to one half the frequency of the transition itself. The five-harmonic stimuli were, like the three-harmonic ones, coextensive with the transition and had the same input amplitude. They went through the deemphasis filter and the filter for setting the transition above or below the duplexity threshold.

Procedure. Except as specified below, the procedure was as in Experiment 1. The duplexity thresholds were similar to those in Experiment 1. The [d] transition became audible at lower intensities (−7.6 dB relative to the intensity of the *F*3, averaged across the subjects) than the [g] transition (−3.2 dB). For the identification task in the present experiment, all of the stimulus sets were included in a single randomized sequence for each condition (above and below duplexity). Twenty repetitions of each of the transitions ([d] or [g]) for each stimulus set (syllable-only, preceding, continuation, three-harmonic, or five-harmonic) were presented in random order for identification of the consonant as "d" or "g." This resulted in 200 responses per condition, with the below duplexity coming first.

Subjects. The subjects were 12 undergraduate students from Yale University, none of whom had participated in the first experiment. They had no reported hearing problems and were paid for their participation. Two failed the screening test and were excluded from the experiment.

Results

The parts of Experiment 2 that dealt with the syllable-only and precursor stimuli were identical in procedure with Experiment 1, except that a wholly different set of subjects was used in Experiment 2. As can be seen in Tables 2 and 3, the results of the two experiments were much the same. In both, there was a highly significant amount of phonetic integration for both kinds of stimuli in the above- and below-duplexity conditions alike, and in both experiments, the amount of phonetic integration was significantly reduced by the precursor stimuli. This increases our confidence that these effects are reliable. For the continuation, three-harmonic and five-harmonic stimulus sets, there was significant phonetic integration in all cases; it was reduced only for the three- and five-harmonic stimulus sets in the below-duplexity condition.

Discussion

In Experiments 1 and 2, the precursor tone reduced the amount of phonetic integration, but the continuation tone of Experiment 2, which was identical to the precursor in intensity, duration, and continuity with the transition, did not. Interpretation of this difference is complicated by the fact that the continuation tone overlapped the formants of the vowel in time and therefore might have been masked by them. A further test might be to compare precursor tones and continuation tones for utterance-final stops, where the continuation would not overlap the formants of the vowel.

In the remaining cases, reductions in phonetic integration were produced in some circumstances though not

Table 2
Percent Correct on the Phonetic Identification Task for Five Stimulus Sets, Above the Duplexity Threshold, Experiment 2

Subject	Syllable Only	Precursor	Continuation	Three-Harmonic	Five-Harmonic
1	95.0*	85.0*	97.5*	92.5*	90.0*
2	100.0*	82.5*	97.5*	95.0*	92.5*
3	100.0*	82.5*	100.0*	100.0*	97.5*
4	87.5*	57.5	90.0*	87.5*	87.5*
5	80.0*	52.5	90.0*	97.5*	97.5*
6	95.0*	90.0*	95.0*	100.0*	100.0*
7	97.5*	77.5*	92.5*	85.0*	65.0*
8	95.0*	50.0	95.0*	90.0*	92.5*
9	62.5	47.5	62.5	55.0	52.5
10	97.5*	75.0*	97.5*	97.5*	87.5*
M	91.0*	70.0*	91.8*	90.0*	86.3*
SEM	±3.7	±5.2	±3.4	±4.2	±4.9
t	—	5.83	−0.61	0.39	1.18
p		<.001	n.s.	n.s.	n.s.

Note—The *t* test compares the syllable-only stimulus set with each of the other four sets. There are 9 *df* for this test. *Percentages that are individually better than chance.

Table 3
Percent Correct on the Phonetic Identification Task for Five Stimulus Sets, Below the Duplexity Threshold, Experiment 2

Subject	Syllable Only	Precursor	Continuation	Three-Harmonic	Five-Harmonic
1	100.0*	72.5*	100.0*	80.0*	92.5*
2	90.0*	57.5	92.5*	62.5	62.5
3	90.0*	75.0*	97.5*	45.0	50.0
4	100.0*	77.5*	100.0*	97.5*	90.0*
5	70.0*	52.5	75.0*	50.0	55.0
6	97.5*	95.0*	100.0*	90.0*	57.5
7	77.5*	75.0*	75.0*	60.0	50.0
8	100.0*	90.0*	100.0*	97.5*	97.5*
9	72.5*	75.0*	57.5	57.5	52.5
10	72.5*	62.5*	57.5	47.5	52.5
M	87.0*	73.3*	85.5*	68.8*	66.0*
SEM	±4.0	±4.2	±5.6	±6.5	±6.1
t	—	3.80	0.62	4.51	5.16
p		<.01	n.s.	<.01	<.001

Note—The *t* test compares the syllable-only stimulus set with each of the other four sets. There are 9 *df* for this test. *Percentages that are individually better than chance.

in all, but in no case is it clear what the reduction is to be attributed to. One possibility is that the nonphonetic stimulus additions captured the sinusoidal cue, causing it to be integrated with the nonphonetic stimulus into a coherent auditory percept, and thus precluding its phonetic integration with the base. Perhaps the most direct evidence for that possibility would be to find in the precursor case, for example, that when it precedes the transition cue listeners perceive the precursor tone to be twice as long as it is when presented in isolation. Unfortunately, that test and the appropriately analogous tests for the harmonic-complex cases are hardly possible, because the perceptual judgments that are called for are virtually impossible to make. Another, and equally convincing, kind of evidence for capture was obtained by Darwin and Sutherland (1984) when they showed that, with the precursor in place, perception of the vowel was the same as that evoked when the harmonic corresponding to the precursor had been reduced in amplitude. Similar evidence about our consonants would require that, with the nonphonetic stimuli in place, listeners perceive the syllable as they do when the base is presented in isolation. We cannot make that test, however, because perception of the base has been found in informal tests to be quite variable, even within a testing session. There is a more indirect test of whether the transitions are captured, however, and it was the purpose of Experiment 3.

EXPERIMENT 3

As an explanation for the reduction of phonetic integration in Experiments 1 and 2, an alternative to the assumption that the nonphonetic stimuli captured the phonetic cue—perhaps *the* alternative—is to suppose that they acted as distractors. Surely, it is possible that the nonphonetic stimuli attracted the listener's attention, and so interfered with the reporting of these phonetic percepts, which can be assumed to be less robust than the percepts generated by natural speech. That is, despite the consistent performance on the part of our subjects, the syllables lacked much of the information normally present and effective. The purpose of Experiment 3 was to test for that possibility. A simple way to do that is to pair each transition with the appropriate nonspeech pattern for one of the sinusoidal cues, as in Experiment 2, and, in a new condition, with the nonspeech pattern appropriate for the other. While it has been found that inharmonic tones can group perceptually (Bregman & Pinker, 1978), such groupings are less effective than harmonic ones. Thus, if the degree of interference is the same for the harmonic and inharmonic versions, it will be likely that something other than perceptual grouping is responsible.

Method

Stimuli and Equipment. The stimuli and equipment were the same as those in Experiment 2, except for the following: There were no "continuation" stimuli, and each transition was paired not only with the nonspeech stimulus appropriate to its trajectory and thus calculated to capture (the "coherent" stimuli; see Figures 2b, 2d–2g), but also with the one appropriate to the trajectory of the other and thus calculated not to capture or to capture more weakly (the "noncoherent" stimuli: see Figures 3a, 3b, 3c). Thus for a [d] transition, the coherent precursor would be the linear extension from the [d] pattern, while the noncoherent signal would consist of the precursor based on the [g] pattern. A similar manipulation was made for the three- and five-harmonic complexes.

Procedure. The screening test, the determination of the duplexity thresholds, and the establishing of the levels for the above- and below-duplexity conditions were carried out as for Experiments 1 and 2. The duplexity thresholds were somewhat higher than those in Experiments 1 and 2. The [d] transition became audible at lower intensities (+2.2 dB relative to the intensity of the *F*3, averaged across the subjects) than the [g] transition (+6.8 dB). In each condition (above and below duplexity), 10 repetitions of each of the transitions ([d] or [g]) with each of the coherent signals and each of their noncoherent counterparts were presented in random order for identification of the consonant as "d" or "g." The below-duplexity condition was run first.

Subjects. The subjects were 11 members of the Yale University community with no reported hearing problems; they were paid for

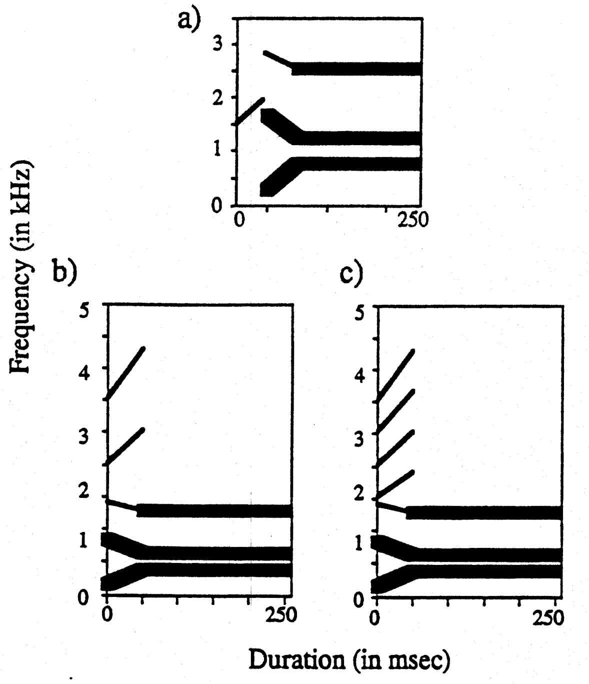

Figure 3. Schematic spectrograms of the noncoherent stimuli, Experiment 3. Only the version for [da] is shown. (a) [da] stimulus with preceding, noncoherent sinusoid. (b) Three-harmonic noncoherent stimulus for [da]. (c) Five-harmonic noncoherent stimulus for [da].

their participation. None had participated in the previous experiments. One failed the screening test and was excluded from the experiment.

Results

In Tables 4 and 5, one can see in all stimulus sets, both above and below the duplexity level, that there is significant phonetic integration between the base and the acoustically disparate sinusoidal transition. One also sees, at both levels, that the coherent precursor tone reduced phonetic integration, replicating Experiments 1 and 2. At neither level, however, did the noncoherent version of the precursor reduce phonetic integration.

With the harmonic complexes, the result is quite different. In the above-duplexity condition, all four tone complexes failed to disrupt integration—in fact, for each of those, phonetic integration was greater than that of the syllable-only stimulus set. This is consistent with the results of Experiment 2. In the below-duplexity condition, the three-tone complexes did not affect integration, which is not consistent with the results of Experiment 2. But the five-tone complexes did reduce integration, just as they had in Experiment 2. However, they did so in both the coherent and the noncoherent versions. That is, the noncoherent patterns were just as likely to reduce phonetic integration as the coherent versions.

Discussion

The results of Experiment 3 provide indirect evidence that the precursor sinusoid was effective in reducing phonetic integration, not by acting as a distractor, but by capturing the sinusoidal cue. This seems a reasonable inference from the fact that only the coherent precursor had an effect. On the other hand, such effects as the harmonic complexes had were the same for coherent and noncoherent stimuli, which implies that the complexes were acting simply as distractors. The noncoherent complexes had, if anything, even more of an effect than the coherent ones, indicating that perceptual grouping is unlikely to be at work here. It may seem odd that the quieter distractors (i.e., those in the below-duplexity case) were the ones that affected the judgments. However, in these stimuli, the transition is also weak, and thus the whole pattern is more vulnerable to the effects of distraction. It may also be that grouping via common frequency is weak for such short stimuli at such high frequencies. Experiment 4 will help us address that concern.

Table 4
Percent Correct on the Phonetic Identification Task for Seven Stimulus Sets, Above the Duplexity Threshold, Experiment 3

		Precursor		Three-Harmonic		Five-Harmonic	
Subject	Syllable Only	Coherent	Noncoherent	Coherent	Noncoherent	Coherent	Noncoherent
1	100*	80*	100*	100*	100*	95*	75*
2	100*	95*	100*	100*	100*	95*	100*
3	100*	100*	100*	100*	100*	100*	100*
4	95*	70*	100*	100*	100*	100*	95*
5	80*	50	100*	90*	95*	95*	85*
6	90*	40	80*	95*	100*	90*	100*
7	65*	75*	65*	70*	80*	75*	65*
8	75*	55	55	95*	85*	85*	100*
9	100*	55	100*	95*	100*	90*	100*
10	95*	45	85*	100*	85*	95*	90*
M	90.0*	66.5*	88.5*	94.5*	94.5*	92.0*	91.0*
SEM	±3.9	±6.6	±5.3	±2.9	±2.5	±2.4	±3.9
t	—	3.55	0.45	−2.08	−1.78	−0.80	−0.25
p		<.01	n.s.	n.s.	n.s.	n.s.	n.s.

Note—The competing signals occurred in both coherent and noncoherent versions. The *t* test compares the syllable-only stimulus set with each of the other six sets. There are 9 *df* for this test. *Percentages that are individually better than chance.

Table 5
Percent Correct on the Phonetic Identification Task for Seven Stimulus Sets, Below the Duplexity Threshold, Experiment 3

		Precursor		Three-Harmonic		Five-Harmonic	
Subject	Syllable Only	Coherent	Noncoherent	Coherent	Noncoherent	Coherent	Noncoherent
1	100*	75*	100*	100*	95*	95*	60
2	100*	100*	100*	100*	100*	95*	100*
3	100*	100*	100*	100*	100*	95*	100*
4	100*	100*	100*	100*	100*	95*	95*
5	85*	85*	100*	90*	90*	85*	75*
6	100*	90*	100*	100*	100*	95*	100*
7	100*	70*	100*	60	70*	65*	60
8	90*	55	65*	45	65*	60	50
9	100*	55	100*	100*	100*	95*	85*
10	90*	85*	100*	100*	60	80*	65*
M	96.5*	81.5*	96.5*	89.5*	88.0*	86.0*	79.0*
SEM	±1.8	±5.5	±3.5	±6.3	±5.2	±4.3	±6.1
t	—	2.76	0.00	1.16	1.93	2.79	3.19
p		<.05	n.s.	n.s.	<.10	<.05	<.05

Note—The competing signals occurred in both coherent and noncoherent versions. The *t* test compares the syllable-only stimulus set with each of the other six sets. There are 9 *df* for this test. *Percentages that are individually better than chance.

The part of our experiment that dealt with the coherent and noncoherent precursors finds a parallel in the dichotic experiment by Ciocca and Bregman (1989), referred to earlier. For what we would call the coherent condition, they preceded and followed the transition with frequency sweeps that were linear extrapolations of it. For what we would call the noncoherent version, the three preceding tones in those extrapolations were shifted by 2000 Hz. The coherent signals reduced phonetic integration, though not to zero, but the noncoherent signals did not. Thus, our results agree with theirs in assigning an important role to temporal priority in reducing phonetic integration, presumably by the capturing of the transition cue.

EXPERIMENT 4

Although the likeliest explanation for the pattern of results in Experiment 3 is that the precursors captured the transitions, there is still some room for doubt. One way to remove those doubts is to show that the capturing tone can itself be captured by another competing signal that begins at the same time as the precursor and ends as the syllable begins, as in Darwin and Sutherland (1984). If the reduction in phonetic effectiveness is due to capture and not to, say, distortion of the onset spectrum for the stop, then the precursor should be capturable. To determine if it can, in fact, be captured, we made it part of harmonic complexes similar to those used in Experiment 3. Such complexes allow us also to test our assumption (in Experiments 2 and 3) that the harmonically related tones do, in fact, form a coherent auditory pattern with the transition cue, even at the high frequencies they had in our experiments. While harmonicity is a strong perceptual factor, it is likely to be less strong in the higher frequency regions than in the 1- to 3-kHz range.

Method

Stimuli and Equipment. The stimuli and equipment were much the same as those in Experiment 3. The base, syllable-only, and precursor stimuli were used. In addition, there were four new stimuli, designed to capture the precursor tone. As in the tonal complexes of Experiment 2, we added tones that coincided with the precursor in time and were harmonically related to it (see Figure 4). One set ("three-tone complex") used the precursor as a fundamental for two additional harmonics. The other set used frequencies one-half the value of the precursor as a (missing) fundamental for five harmonics (Nos. 2–6). These tones were coextensive with the precursor and ended as the syllable began. The analog of the noncoherent series from Experiment 3 was not possible because the "d" transition overlapped with at least one tone of the "g" harmonic series.

Procedure. Since we were measuring release from capture in this experiment, it is clear that we could test only that subset of subjects that does show a capture effect. We also needed to run subjects who could hear the phonetic difference intended. Thus, each test included a screening test, but one that was somewhat different from that used before. Instead of using the patterns made exclusively of formants, we used the patterns with the sinusoid transitions that would appear in the main test. In order to make the session as short as possible, there was no break between the screening test and the main test.

The main test consisted of 20 repetitions of nine stimuli. One was the base by itself. Inclusion of the base allowed us to determine, for those subjects who found the base unambiguous, whether the perception in the capturing cases returned to the base or was simply random. The other eight stimuli came in pairs, one having "d" transitions and one having "g" transitions. They were the syllable-only, precursor, three-tone precursor complex, and five-tone precursor complex stimuli. These occurred in random order (the same order across subjects). A single presentation level was chosen, namely, 0 dB relative to $F3$ for the transition or transition plus precursor(s). This was a level that was above duplexity for some subjects but well within the phonetic range for all subjects. Some of the subjects in this experiment, then, can be presumed to be above, and some below the duplexity threshold. Since this manipulation had no effect on the precursor case, collapsing the data in this way seemed justified.

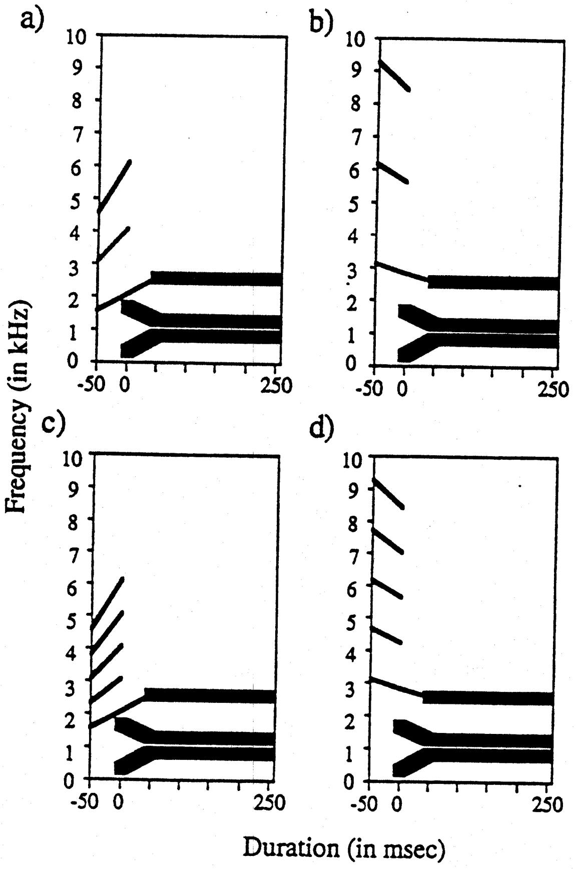

Figure 4. Schematic spectrograms of the tone-complex stimuli, Experiment 4. (a) [ga] stimulus with three-tone, preceding, coherent sinusoid complex. (b) Three-tone stimulus for [da]. (c) Five-tone stimulus for [ga]. (d) Five-tone stimulus for [da].

Subjects. Twenty-five colleagues at Haskins Laboratories were run. These subjects had various levels of phonetic experience (from none to a great deal) and various language backgrounds (seven had a language other than English as their first language). Neither of these factors appeared to explain the patterning of results, although there were too few subjects to test this directly. Eleven were female and 14 were male.

Seven subjects failed the screening test. Two heard all "da"s; two heard all "ga"s; one heard all "bla"s; and two were random. An 8th subject (Subject 7 in Table 6) failed the screening test but was included in the final analysis due to his improved performance on the main part of the test. (This subject spontaneously reported that he began finding the distinction easier to perceive as the test wore on.) This late recognition of the distinction reduced his capturing effect if we look only at the precursor versus the syllable-only stimuli, but the capturing is significant compared with both the three- and five-tone complexes. Six subjects had excellent performance on the screening task and also had high performance throughout, indicating that the precursor failed to capture for them. These subjects also had to be excluded. Results are reported for the remaining 12 subjects.

Results

The results for the 12 "capturing" subjects are shown in Table 6. For individual subjects, responses to two stimulus sets are significantly different (at .05) if they are 10 or more percentage points apart. Three subjects show significant recovery from the capturing effect, while 8 show no effect. Additionally, one other subject (No. 11) had significant recovery for the three-tone stimuli but not for the five-tone. For the group statistics, then, there is no significant difference between the tone complex stimuli and the precursor stimuli [$F(1,11) < 1$, n.s., for the average of the two complexes; the comparisons for each of the tone complexes are also not significant]. For some subjects, although apparently a minority, there is significant capture of the precursor tone, resulting in restored phonetic perception.

There is some room to think, then, that some subjects group these harmonic complexes more strongly than do other subjects. This suggests that the precursor tones might interfere with phonetic identification either by capturing the transition or by some other mechanism, since all 12 subjects showed decreased phonetic performance with the precursor tones. If the precursor was, in fact, capturing the transition, we should expect the subjects' percepts to revert to the base. Of the 12 subjects who showed reduced phonetic integration, and thus presumably capture, with the precursor stimuli, 2 had found the base to be ambiguous. For them, therefore, there is no way to tell whether or not the precursor was itself captured. The remaining 10 subjects can be divided into the 4 who showed recovery of the phonetic percept (from which it can be assumed that the precursor was itself captured by the harmonic complex) and the 6 who did not. It is of interest, then, to see whether responses to the *simple* precursor stimuli reverted to the base, as should happen for captured transitions. For that purpose, we have calculated the deviation from 50% response for each cat-

Table 6
Percent Correct on the Phonetic Identification Task for Four Stimulus Sets in Experiment 4, and the Identification of the Base Alone

Subject	Base	Syllable (R) Only	Precursor	Three-Harmonic	Five-Harmonic
1	90–g	100.0	90.0	87.5	80.0
2*	90–d	95.0	77.5	90.0	90.0
3*	100–d	92.5	77.5	97.5	95.0
4	75–d	100.0	87.5	87.5	90.0
5	100–d	100.0	90.0	92.5	90.0
6	55–d	92.5	82.5	82.5	82.5
7*	80–d	82.5	75.0	87.5	85.0
8	80–d	100.0	72.5	65.0	77.5
9	95–d	100.0	77.5	72.5	82.5
10	55–g	90.0	75.0	70.0	72.5
11(*)	100–d	92.5	65.0	80.0	62.5
12	65–d	97.5	72.5	77.5	45.0
M	—	95.2	78.5	82.5	79.4
SEM	—	±1.6	±2.2	±2.8	±4.0
t		—	7.96	3.72	3.64
p			<.001	<.01	<.01

Note—The Base column shows the percentage of the majority judgment and the category of that judgment. For individuals, values that differ 10% or greater are significantly different. The *t* test has 11 *df*.
*The 4 subjects who had significant recovery of phonetic perception (with parentheses where only one condition showed recovery).

egory, with positive numbers being toward the category of the base perception and negative numbers being away from the base perception. The 4 restoring subjects had an average of 13.8% (individually, 12.5, 7.5, 5, 30), while the 6 nonrestoring subjects averaged −11.7% (individually, 0, −12.5, −5, −27.5, −17.5, −7.5). Despite the small number of subjects, these means differ significantly by a t test [$t(8) = 3.79, p < .01$]. Thus, the subjects who showed no recovery actually perceived more of the *opposite* category, indicating that, whatever was responsible for their lower performance in the precursor case, it was unlikely to be capturing.

Discussion

Experiment 4 provides support for two of our earlier conclusions. First, it is clear that some subjects perceptually group the precursor tone with its harmonics, and thus show that the precursor is itself captured. Moreover, these same subjects revert significantly more to perception of the base with the precursor stimuli than those who showed no evidence that the precursor was captured, further indicating that capture is active for them but not for the other subjects. While it would have been more compelling if all of the subjects had had the same effect, the fact that any of the subjects showed this restoration indicates that this perceptual grouping is available, and that the original precursor effect is, at least in part, due to capturing of the transition. Second, by this same token, the fact that the precursor can be captured by these tones indicates that the perceptual grouping that we assumed for the tone complexes in Experiments 2 and 3 are ones that are, in fact, used under some circumstances. While not being a direct test of the susceptibility of the transition to capture by these tones, this experiment does show that such capture can occur. Thus our previous interpretation is supported in two distinct ways.

GENERAL DISCUSSION

In the introduction, we embraced the claim that speech perception is the business of a phonetic module, specialized to process the acoustic signal so as to recover the coarticulated gestures that were its distal causes. Accordingly, as we said there, the primary perceptual representations are immediately phonetic; they are not, as on a common view, auditory percepts that are invested with phonetic significance at some secondary, cognitive stage. One kind of evidence for our view is that, in duplex perception, listeners integrate into a coherent phonetic percept acoustic information from sources that are, on auditory grounds, quite disparate. In the experiments reported here, the evidence for separate sources was in the disparity between a sinusoidal $F3$ transition cue for "da" versus "ga," on the one hand, and the remainder of a synthetic syllable that was made of resonances (i.e., formants), on the other. Integrating across that disparity is telling, because the mechanisms of scene analysis that assign sounds to sources must occur at the level of the primary perceptual representation, not at some cognitive remove. So, if the phonetic system can ignore those mechanisms, as it does in duplex perception, its representations must be primary in the same way that ordinary auditory representations normally are.

In the experiments reported here, we found, as others had, that phonetic integration across disparate sources can be reduced when the phonetically relevant information is made part of a coherent auditory pattern, and so "captured." For some listeners, the capturing signal could itself be captured when it was made part of a coherent pattern, in which cases, the phonetic integration was restored. We were unable to effect capture of the phonetically relevant cue, and thus reduce phonetic integration, with harmonically related signals that were presented at the same time. The two instances in which phonetic performance was reduced with simultaneous complexes appeared to be due to distraction rather than capture. But just because the cue can, in some circumstances, be captured and presumably made part of an auditory pattern, it does not follow that the phonetic percept is therefore dependent on, and secondary to, the auditory. Like all perceptual specializations, the phonetic system is elastic in the sense that it will accept stimuli that go beyond the bounds of what is ecologically possible. But that elasticity is not infinite; as the evidence for disparity is increased, there must be a point at which the phonetic system is stretched beyond its limits, and integration fails. Apparently, the limit on elasticity is different for different listeners.

In the reduction of integration that we observed, it is as if the information was somehow shared between the auditory and phonetic systems. But surely that does not mean, as some have implied (Ciocca & Bregman, 1989; Darwin & Culling, 1990) that the systems are therefore not independent. One possibility, which seems to us unlikely, is that the information is used twice, once by the one system and then again, but independently, by the other. A more likely arrangement, in our view, would have the information divided between the system in such a way that what the one system gets the other can't have. In either case, however, the phonetic system is independent of the auditory in the important sense that its psychological units are not formed of primary auditory representations. Rather, just as the first-stage percepts of the auditory system are distinctly auditory, those of the phonetic system are distinctly phonetic.

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.

Bentin, S., & Repp, B. H. (1986). *Central masking effects in duplex speech perception*. Unpublished manuscript.

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

BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

BREGMAN, A. S., & PINKER, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, **32**, 19-31.

CIOCCA, V., & BREGMAN, A. S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics*, **46**, 39-48.

CIOCCA, V., & DARWIN, C. J. (1993). Effects of onset asynchrony on pitch perception: Adaptation or grouping? *Journal of the Acoustical Society of America*, **93**, 2870-2878.

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

DARWIN, C. J. (1995). Perceiving vowels in the presence of another sound: A quantitative test of the "Old-plus-New" heuristic. In C. Sorin, J. Mariani, H. Méloni, & J. Schoentgen (Eds.), *Levels in speech communication: Relations and interactions: A tribute to Max Wajskop* (pp. 1-12). Amsterdam: Elsevier.

DARWIN, C. J., & CULLING, J. F. (1990). Speech perception seen through the ear. *Speech Communication*, **9**, 469-475.

DARWIN, C. J., & SUTHERLAND, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology*, **36A**, 193-208.

DIEHL, R. L., & KLUENDER, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, **1**, 121-144.

EIMAS, P. D., & MILLER, J. D. (1992). Organization in the perception of speech by young infants. *Psychological Science*, **3**, 340-345.

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

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.

HALL, M. D., & PASTORE, R. E. (1992). Musical duplex perception: Perception of figurally good chords with subliminal distinguishing tones. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 752-762.

HUKIN, R. W., & DARWIN, C. J. (1995). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception & Psychophysics*, **57**, 191-196.

KUHL, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.

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.

MILLER, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals* (Life Sciences Research Report 5, pp. 49-58). Berlin: Dahlem Konferenzen.

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.

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. (1984). On the perception of intonation from sinusoidal sentences. *Perception & Psychophysics*, **35**, 429-440.

REMEZ, R. E., & RUBIN, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics*, **48**, 313-325.

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.

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). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 243-335). New York: Academic Press.

REPP, B. H., MILBURN, C., & ASHKENAS, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics*, **33**, 333-337.

RICHARDS, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America*, **61**, 410-414.

SCHOUTEN, M. E. H., & HESSEN, A. J. VAN (1993). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, **92**, 1841-1855.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Identification of consonants and vowels presented to the left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59-63.

STEVENS, K. N., & HOUSE, A. S. (1972). Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (pp. 1-62). New York: Academic Press.

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., & LIBERMAN, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, 237, 169-171.

WHALEN, D. H., WILEY, E. R., RUBIN, P. E., & COOPER, F. S. (1990). The Haskins Laboratories' pulse code modulation (PCM) system. *Behavior Research Methods, Instruments, & Computers*, **22**, 550-559.

(Manuscript received April 3, 1995;
revision accepted for publication November 15, 1995.)