The role of "chirp" identification in duplex perception

HOWARD C. NUSBAUM and EILEEN C. SCHWAB
Indiana University, Bloomington, Indiana

and

JAMES R. SAWUSCH
State University of New York, Buffalo, New York

Duplex perception occurs when the phonetically distinguishing transitions of a syllable are presented to one ear and the rest of the syllable (the "base") is simultaneously presented to the other ear. Subjects report hearing both a nonspeech "chirp" and a speech syllable correctly cued by the transitions. In two experiments, we compared phonetic identification of intact syllables, duplex percepts, isolated transitions, and bases. In both experiments, subjects were able to identify the phonetic information encoded into isolated transitions in the absence of an appropriate syllabic context. Also, there was no significant difference in phonetic identification of isolated transitions and duplex percepts. Finally, in the second experiment, the category boundaries from identification of isolated transitions and duplex percepts were not significantly different from each other. However, both boundaries were statistically different from the category boundary for intact syllables. Taken together, these results suggest that listeners do not need to perceptually integrate F2 transitions or F2 and F3 transition pairs with the base in duplex perception. Rather, it appears that listeners identify the chirps as speech without reference to the base.

Speech consists of a collection of chirps, buzzes, and clicks. However, as users of spoken language, we are normally unaware of these bizarre sounds that constitute the acoustic structure of speech. Instead of hearing this cacophony, we perceptually integrate these sounds to form a structured sequence of linguistic units. Thus, speech has an inherently dualistic nature. First and foremost, speech is a *linguistic* signal. But, at the same time, the speech waveform is an *acoustic* stimulus—a series of acoustic events distributed in time. This distinction has been very important in speech research because the relationship between the acoustic waveform and corresponding perceptual categories has seldom displayed a simple one-to-one mapping (see Liberman, 1970). There is considerable evidence that no single static portion of the waveform carries all the acoustic information that cues a particular phonetic percept (Dorman, Studdert-Kennedy, & Raphael, 1977; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Walley & Carrell, in press; Kewley-Port, Note 1). Instead, successive phonemes are "encoded" into the speech wave-

form by articulation (see Liberman, 1970). The information corresponding to a particular phoneme is therefore distributed over several acoustic segments, overlapping and interacting with cues to other temporally adjacent phonemes. One implication of the encoded nature of speech is that a perceptual mechanism must integrate information from several acoustic events in order to derive phonetic segments. Moreover, it has been suggested that this perceptual mechanism must be *specialized* for its task, since the principles that support the acoustic-to-phonetic integration process ultimately stem from the principles that govern speech production (see Liberman, 1970, 1974, 1982; Liberman & Studdert-Kennedy, 1978; Repp, 1982).

One source of evidence for this theoretical position comes from a study by Mattingly, Liberman, Syrdal, and Halwes (1971). They compared *discrimination* of different second formant (F2) transitions when these transitions were presented in the context of an entire syllable with *discrimination* of the same transitions presented in isolation. The full syllables that served as context were distinguishable only by differences in the F2 transitions. The syllables were heard as [bae], [dae], or [gae]. In contrast, the isolated F2 transitions sounded more like bird chirps than speech. However, subjects were *never* asked to *identify* these transitions as speech.

Mattingly et al. found that discrimination of the

This research was supported by NIMH Grant MH 31468-01 to SUNY/Buffalo. Preparation of this manuscript was supported, in part, by NIH Training Grant NS-07134 to Indiana University. The authors would like to thank David B. Pisoni for many helpful suggestions and comments. Reprint requests should be sent to Howard C. Nusbaum, Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

transitions in syllables was governed by the phonetic categories of speech, while discrimination of the isolated transitions (i.e., the chirps) seemed to be independent of these categories. They interpreted these results as evidence for two distinct perceptual modes. Perception of the isolated transitions (as chirps) was mediated by general auditory processes. However, the perception of transitions in syllables was assumed to be carried out by a specialized speech processor that correctly integrated the transition information with the remainder of the syllable.

More recently, it has been claimed that this specialized speech processor can integrate acoustic transition information with the rest of a syllable, even if these two sources of information are presented to different ears (see Liberman, 1979, 1982; Repp, 1982). The assertion is that transitions heard in one ear can be properly *combined* with acoustic information in another ear to produce a unified phonetic percept. This claim comes from research on a perceptual phenomenon called "duplex perception" (Liberman, 1979, 1982; Repp, 1982). Rand (1974) first demonstrated this phenomenon by separating a syllable into two complementary parts. One of these parts contained the acoustic information that cued the phonetic distinction of place of articulation (e.g., the difference between [b], [d], and [g]). This acoustic segment consisted of the isolated second and third formant transitions with associated bursts. According to Mattingly et al. (1971), these isolated transitions should be heard as nonspeech chirp-like sounds. The remainder of the syllable, without these distinguishing cues, constituted the other acoustic segment. When presented alone, without the critical transitions, this portion of the syllable—called the "base" (Liberman, Isenberg, & Rakerd, 1981)—is heard as containing a [b]-like consonant (see Cutting, 1976). When Rand presented the transitions and the base (with synchronized onsets) to different ears, subjects appeared to integrate these segments, since they correctly identified the consonant as [b], [d], or [g]. However, the subjects did not simply "fuse" these two acoustic segments together to form a single unitary percept. Rand reported that the listeners simultaneously heard *two* distinctly different sounds. One of these sounds was a nonspeech chirp heard in the ear with the transition information. The second sound was the correctly integrated syllable heard in the other ear (the ear that received the base). Thus, it appears that the transitions played two completely different perceptual roles. While being perceived in one ear as nonspeech, the same transitions also cued the correct phonetic percept in the opposite ear. It is this duality of perceptual function that is referred to as duplex perception.

The demonstration of duplex perception has provided further evidence for the existence of two sep-

arate perceptual modes. A general auditory mode is presumed to mediate chirp perception while a separate specialized speech mode is assumed to mediate the phonetic integration of the transitions in one ear with the rest of the syllable in the other ear (Cutting, 1976; Liberman, 1979, 1982).

What is the evidence for this explanation of duplex perception? First, there is the subjective experience of duplex perception (see Liberman, 1979, 1982; Liberman et al., 1981). Listeners perceive speech and nonspeech simultaneously in different ears. Furthermore, listeners report hearing the correct phonetic percept (as cued by the transitions) on the speech side of the dichotic stimulus pair, that is, in the ear with the phonetically undistinguished portion of the syllable (i.e., the base). In other words, listeners hear two sounds, and they identify the phonetic response with the base instead of with the transitions. To some extent, the compelling nature of this phenomenology may be the strongest rationale for the dichotic integration hypothesis—that the transitions are integrated into the base prior to phonetic labeling.

Despite the compelling experience of duplex perception, phenomenology does not constitute strong scientific evidence. There is, however, some support for this hypothesis from experiments using duplex perception. Cutting (1976) demonstrated that changing the temporal relationship between the transition information in one ear and the base in the other ear reduced the proportion of correct phonetic responses. It is as if the transitions and the syllabic base must be correctly time-aligned (as they are in a normally produced intact syllable) to permit accurate phonetic processing. This is just the result that might be expected if the putative specialized speech processor integrated acoustic cues by taking into account their common articulatory origins (see Best, Morrongiello, & Robson, 1981, and Repp, 1982, for discussions of the articulatory integration hypothesis).

Another experiment supporting dichotic integration was recently reported by Liberman et al. (1981), who had subjects discriminate AX pairs of duplex stimuli. In one condition, subjects discriminated A from X using the chirp side of the duplex percept. In a second condition, subjects were instructed to discriminate A from X using the speech side of the duplex percept. Liberman et al. found that chirp discrimination was independent of the type of base presented in the other ear. However, discrimination on the speech side depended on both the chirps and the bases. Liberman et al. concluded that the subjects must have perceptually integrated the transitions in one ear with the bases in the other ear to form intact syllables. They also claimed that this perceptual integration was mediated by a specialized speech processor.

Taken together, the available evidence favors the

dichotic integration explanation of duplex perception. One reason for this, however, is the absence of any alternative explanation that could account for duplex perception. Furthermore, there are some results that cannot be easily explained by the hypothesis that a specialized speech processor directly integrates acoustic information across the two ears. Rand (1974), for example, reported that attenuating the amplitude of the F2 and F3 transitions from a consonant-vowel syllable produced a greater decrement in consonant identification when the transitions were presented in an intact syllable than when the transitions were paired dichotically with the base. This result demonstrates that the transitions are processed differently in an intact syllable and on the speech side of the duplex percept.

In a subsequent study, Cutting (1976) showed that fundamental frequency (F0) differences between the transitions in one ear and the base in the other ear did not affect phonetic judgments in duplex perception. Thus, even though F0 indicated that the transitions and base were articulated by *different sources*, phonetic labeling was not disturbed. If the specialized phonetic processor integrates acoustic features according to their *common articulatory origins* (cf. Best et al., 1981; Repp, 1982), the transitions that differed from the base in F0 should not have been integrated with the base. These two acoustic segments (the transitions and the base) should have been treated as arising from different articulatory sources, and the proportion of correct phonetic judgments should have decreased with increasing F0 disparity. Thus, Cutting's results contradict this prediction.

Clearly, these findings pose problems for a speech processor that integrates information across ears with respect to articulatory source. However, there is an alternative account of these results that does not depend on the dichotic integration assumption. Subjects may be able to directly derive some phonetic information from the isolated transitions without integrating the transitions with the base. While this hypothesis may seem to be contradicted by the results obtained by Mattingly et al. (1971) for chirp discrimination, it is important to remember that, in this study, subjects were not asked to identify the isolated transitions.¹ From the experimenter's point of view, this might have seemed reasonable given the extremely nonspeech character of these sounds. But, at the same time, these chirps carry the distinctive phonetic information necessary for phoneme identification. With the appropriate instructions, subjects might at least be able to "guess" from which consonant or place of articulation a chirp was derived.

Indeed, there is some support for the proposal that subjects can identify nonspeech signals as speech under certain experimental conditions. Recently, Remez, Rubin, Pisoni, and Carrell (1981) demonstrated

that complex time-varying nonspeech sounds can be perceived as speech when listeners are given appropriate instructions. The sounds were sine-wave analogs of speech matched to the center frequencies of the formants in a spoken sentence. Few, if any, subjects spontaneously heard these sounds as speech; most listeners identified them as "computer noises" or "science fiction sounds." However, when instructed to listen to these stimuli as speech, subjects were able to determine the linguistic content of the signal.

Similar studies, using sine-wave analogs of isolated CV syllables, have obtained different patterns of results when subjects were instructed to listen to these sounds as speech and when they were instructed to listen to the same sounds as nonspeech (Grunke & Pisoni, 1982; Schwab, 1981). These findings clearly demonstrate that listeners who are given appropriate instructions can extract phonetic information from nonspeech signals. Of course, listeners do not suddenly hear these sounds as natural utterances just because they are instructed to process the sounds as speech. Instead, phonetic information is perceived in a distinctly nonspeech-sounding signal. Similarly, subjects might be able to identify the phonetic information inherent in a nonspeech chirp (isolated transitions) under the appropriate circumstances.

How can this explain the phenomenon of duplex perception? When subjects are asked to consciously describe the duplex percept, they will respond with the most general perceptual feature that distinguishes the members of the dichotic stimulus pair—the different auditory qualities of the sounds. The transitions-alone side of the duplex percept has a distinctly nonspeech quality since it is an extremely impoverished stimulus much like the sine-wave analogs. In contrast, the base side seems very speech-like and, when heard alone, it sounds like an intact syllable (Cutting, 1976). The most reasonable description that subjects could provide of the duplex percept is that they hear a speech syllable (the base) and a nonspeech chirp (the transitions). However, when asked to identify the speech, subjects can no longer rely solely on the speech-like, but phonetically constant, base for responding. In order to avoid responding the same way on every trial, subjects must use the transitions (in some way) to produce a phonetic response. It is possible that this phonetic response is directly derived from the chirp without integration of the transitions and base across ears. When subjects are asked to produce responses using the nonspeech side of the duplex percept (e.g., in a discrimination task), they might process the chirp in an auditory mode. But when instructed to use the speech side of the percept, subjects might directly extract some phonetic information from the chirp. Thus, the subjects might interpret instructions to attend to one side of the duplex percept or the other as a cue to process the chirp in either

an auditory mode or a speech mode. Instructions to use the speech side or the nonspeech side of the duplex percept may induce different perceptual expectations about the transitions in much the same way that manipulating instructions affected the perception of sine-wave analogs of speech.

Thus, the chirp-identification hypothesis can account for the same duplex-perception results that were explained by dichotic integration. Moreover, while an articulation-based theory of dichotic integration cannot explain all the results obtained by Rand (1974) and Cutting (1976), these results can be accounted for by the chirp-identification hypothesis. The insensitivity of phonetic judgments to differences in amplitude and F0 between the transitions and the base may simply indicate that subjects identified the chirps without reference to the base.

EXPERIMENT 1

The chirp-identification hypothesis is, of course, predicated on the putative ability of subjects to extract the phonetic information encoded into isolated transitions under the experimental conditions used in duplex research. Since previous research has demonstrated that isolated transitions are not heard as speech (Liberman et al., 1981; Mattingly et al., 1971), it is important to determine if subjects can *identify* chirps as speech under the appropriate experimental conditions (cf. Jusczyk et al., 1981). The integration hypothesis predicts that, while transitions may be labeled phonetically when combined with a base in an intact syllable or in a dichotic presentation, isolated chirps should not be identifiable as phonetic segments because the transitions represent only a portion of the acoustic cues to phonetic identity within a syllable (Liberman, 1970). Therefore, while each subject should be able to reliably assign isolated transitions to different responses (since these chirps are discriminable), the integration hypothesis claims that this mapping is arbitrary without regard for the phonetic source of the transitions. Thus, there should be no consistency across subjects for labeling isolated transitions as speech (see Jusczyk et al., 1981); subjects should not be able to determine the place of articulation of isolated transitions.

In contrast, the chirp-identification hypothesis predicts that subjects should be able to derive some place of articulation information from the isolated chirps. In duplex perception, when specifically instructed to respond to the speech side of the duplex percept, subjects may simply identify the phonetic content of the chirp side. Thus, subjects should be able to correctly extract the phonetic information encoded in transitions when these chirps are accompanied by the base in the other ear *and* when they are presented alone.

Method

Subjects. The subjects were 17 undergraduate students at the State University of New York at Buffalo. All subjects were right-handed native English speakers with no reported history of either speech or hearing disorder. They participated to fulfill part of a course requirement.

Stimuli. The stimuli were six computer-generated sounds that were produced using the parallel branch of a software speech synthesizer (Klatt, 1980) modified by Kewley-Port (Note 2). By using the parallel branch of the synthesizer, the amplitudes of each formant could be specified independently so that the amplitude of the F2 transitions would be the same in isolation and in syllabic context. Two of the stimuli were two-formant versions of the syllables [ba] and [ga]. These stimuli were 295 msec in duration and were based on the two-formant [ba] and [ga] syllables used by Cutting (1976). Fundamental frequency (F0) was set to a constant 100 Hz for both syllables. The first formant (F1) was also the same for both syllables. The F1 transition started at 400 Hz and increased to 740 Hz over the first 45 msec of the syllable and then was held constant at 740 Hz for the remaining 250 msec. For [ba], the second formant transition increased from 1386 to 1620 Hz in the first 65 msec of the syllable. Following the transition, F2 was constant at 1620 Hz for the remainder of the syllable. In the [ga] syllable, F2 started at 1996 Hz and decreased over the first 65 msec to 1620 Hz.

The remaining four stimuli were derived from the intact [ba] and [ga] syllables. Figure 1 shows schematic spectrograms of these syllables. The dashed circles indicate the portion of the syllables used for the two chirps. These F2 chirps were synthesized in isolation from the rest of the syllable. The amplitude of these chirps was the same as the amplitude of the F2 transitions in the corresponding intact syllables. This was determined by comparing the amplitudes of these F2 transitions using LPC spectra.

Two bases were synthesized without the F2 transitions and were acoustically identical to each other. Thus, the six stimuli in this experiment consisted of the intact [ba] and [ga] syllables, the [ba] and [ga] bases, and the isolated [ba] and [ga] F2 transitions. These stimuli were converted to analog form at 10 kHz, low-pass filtered at 4.8 kHz, and then presented in real time under computer control. The stimuli were presented at 76 dB SPL over matched and calibrated headphones (TDH-39).

Procedure. Small groups of two to six subjects participated in a single experimental session, which lasted for 1 h. At the start of each session, the subjects received practice in identifying the intact syllables. During practice, five repetitions of the two intact syllables were presented in random order. The subjects responded by pressing a button marked "ba" or "ga" on a computer-

Figure 1. Schematic spectrograms of the two-formant [ba] and [ga] syllables used for the test stimuli in Experiment 1. The F2 transitions that were isolated as chirps are indicated by the dashed circles. The remainder of the syllables outside the dashed circles provided the bases.

controlled response box. After all subjects had responded, the computer turned on a light over the correct response button. This feedback was presented only during the practice trials.

Following the practice trials, the subjects participated in five experimental conditions. These conditions are depicted in Figure 2. The order of presentation in these conditions was random for different experimental sessions. The test stimuli were each presented 30 times in random order for each condition. The subjects identified all the stimuli in every condition with the labels "ba" and "ga."

In one condition, the intact [ba] and [ga] syllables were presented binaurally. In a second condition, the isolated F2 transition was electronically mixed with its corresponding base and the resulting stimulus was presented binaurally. This condition provided a control to demonstrate that synthesizing the transitions and bases separately did not change the perceptual relationship present in the intact syllables. In the duplex condition, the F2 transitions and bases were presented as dichotic pairs with synchronized onsets. For this condition, the subjects were told that, on each trial, they would hear two sounds, one in each ear, and that they were to identify the speech sound they heard. In a fourth condition, the isolated F2 transitions were presented binaurally. And in the final condition, the isolated bases were presented to both ears. In these two latter conditions, the subjects were told that the stimuli might not sound very speech-like, but these stimuli were, in fact, generated from speech syllables. The subjects were encouraged to identify the stimuli as "ba" or "ga." In addition, the subjects were instructed to guess whether the stimuli were derived from "ba" or "ga" if they heard these stimuli only as nonspeech sounds.

Figure 2. The five conditions in Experiment 1. In the duplex condition, the chirp and base were presented to different ears. In all other conditions, the stimuli were presented binaurally.

Results and Discussion

For all conditions, the data are plotted as the mean percentage of "ba" responses made to each syllable. The identification data for the intact syllables are presented in the left panel of Figure 3 (shown by the dashed line with the open squares). Also, in the left panel of this figure are the data from the condition in which the F2 transitions and bases were electronically mixed together to form whole syllables (shown by the solid line with the solid squares). All of the 17 subjects correctly identified the [ba] and [ga] syllables in both the intact ($p < .002$, two-tailed sign test) and mixed conditions ($p < .002$, two-tailed sign test). Synthesizing the two complementary parts of the syllable separately does not impair phonetic labeling.

The left panel of Figure 3 also shows the identification of the [ba] and [ga] bases when presented in isolation (indicated by the dashed line with the open triangles). As expected, none of the subjects could correctly identify *both* the [ba] and [ga] bases consistently. Of course, this is not surprising because these stimuli were acoustically identical once the distinguishing F2 transitions were removed. Both bases were labeled as "ba" as previously reported by Cutting (1976).

The right panel of Figure 3 shows the data from the duplex condition (the dashed line with the open circles) and the data from identification of the isolated F2 transitions (the solid line with the solid circles). In the duplex condition, where the F2 transitions and bases were presented as a dichotic pair, 14 of the 17 subjects correctly identified both consonants ($p < .02$, two-tailed sign test). Similarly, 14 subjects correctly identified the consonants from which the isolated F2 transitions were derived ($p < .02$, two-tailed sign test). There was no significant difference in the percentage of "ba" responses given to the [ba] chirp in isolation and the percentage of "ba" responses given to the [ba] transition presented with the base in the duplex condition [$t(16) = .178$, n.s.]. Also, there was no significant difference in identification of the [ga] chirp presented alone and in the duplex condition [$t(16) = .735$, n.s.].

These results demonstrate that sufficient phonetic information is perceptually available in the F2 transitions to allow subjects to phonetically label these chirps when presented alone. Since the identification of the chirps was consistent and accurate, it appears that some phonetic information can be directly derived from the F2 transition in the absence of an appropriate acoustic context. These results are of interest because they are contrary to previous claims that have been made concerning the inability of listeners to derive any phonetic information from isolated acoustic features (e.g., Liberman, 1970; Mattingly et al., 1971). Moreover, the similarity of the identification data in the duplex condition and the condition in which the chirps were heard in iso-

Figure 3. In the left panel, the mean percentage of [b] responses for the intact syllables (dashed line, open squares), for the electronically mixed whole syllables (solid line, solid squares), and for the bases with the distinguishing transitions removed (dashed line, open triangles). In the right panel, the mean percentage of [b] responses for the duplex condition (dashed line, open circles) and for the isolated [ba] and [ga] F2 transitions (solid line, solid circles).

lation suggests that subjects might simply have identified the chirps as speech in *both* conditions. From these results, there is certainly no evidence to indicate that phonetic processing of the F2 transitions was affected by the presence of the base in the other ear. Thus, there is no reason to assume that subjects necessarily integrated the F2 transitions (i.e., the chirps) with the base in previous duplex perception experiments (cf. Cutting, 1976; Liberman et al., 1981; Rand, 1974). In short, these results substantially weaken the claim that duplex perception constitutes evidence for separate auditory and phonetic modes of perception.

EXPERIMENT 2

In the first experiment, we found that, with the appropriate instructions, listeners were able to directly derive phonetic information from an isolated F2 transition. Furthermore, the lack of a significant difference in phonetic labeling of the isolated chirps and the duplex stimuli suggests that these sounds may be processed as speech by the same perceptual mechanism. These results clearly contradict the assertion that phonetic processing of the duplex percept is necessarily based on the integration of the transitions with the base to form a perceptually intact syllable (cf. Liberman, 1979, 1982; Repp, 1982). However, proponents of the integration hypothesis might still claim that physically intact syllables and the speech side of the duplex percept are processed by the same perceptual mechanism, while phonetic labeling of the chirps is mediated by an entirely different process.² To test these predictions, it was necessary to determine more precisely the labeling characteristics of the per-

ceptual process (or processes) used in identifying the intact syllables, the speech side of the duplex percept, and the isolated transitions. To accomplish this, we evaluated the effects of systematic changes in stimulus structure on consonant identification. Specifically, we compared the location of the perceptual boundary between [b] and [g] for the three types of test stimuli.

If the chirp and base are truly perceptually integrated in the duplex condition, this fused percept should be processed in the same manner as the intact syllables. Thus, the category boundaries should not differ in these two conditions. If the critical transitions are perceptually fused with the base, using the same specialized process that interprets transitions in intact syllables, this fusion should not be any more impoverished or degraded than the intact syllables. Also, according to the integration hypothesis, since the isolated transitions must be processed differently from normal speech (perhaps by some cognitive association between the sound of the chirp and the phonetic category), the category boundary for isolated transitions should be different from the duplex and intact boundaries. Both the intact syllables and the duplex syllables (transitions fused with base) have F1 transitions, while the isolated chirps would be deprived of this feature. Since Jusczyk et al. (1981) have already shown that the presence of an F1 transition significantly affects place of articulation classification, the labeling boundary for isolated chirps should be different from the boundaries for the duplex and intact syllables. This prediction assumes that, for the duplex syllables, the F2 and F3 chirps are perceptually fused with the base containing the F1 transition.

Method

Subjects. The subjects were 12 undergraduates at the State University of New York at Buffalo. These subjects met the same requirements as the subjects used in Experiment 1.

Stimuli. The stimuli were derived from a series of six consonant-vowel syllables that varied perceptually from [ba] to [ga]. Schematic spectrograms of the endpoints of this series are shown in Figure 4. For all six syllables, F0 and F1 were identical to the values of F0 and F1 in the first experiment. Also, the syllables were 295 msec in duration. In each of the six syllables, F3 started at 2300 Hz and increased to 2600 Hz over the first 65 msec of the syllable and was then constant to the end of the syllable. For the [ba] endpoint (Stimulus 1), F2 started at 1240 Hz and increased to a final value of 1620 Hz. Following the transition, F2 was constant at 1620 Hz for all six syllables. For the [ga] endpoint (Stimulus 6), the initial value of F2 was 2000 Hz. The F2 transition was 65 msec long in all the syllables. The test series was constructed by systematically increasing the onset frequency of the F2 transition in 152-Hz steps from the initial value in the [ba] endpoint until the [ga] endpoint was reached.

The test stimuli were generated with the software synthesizer used in the first experiment. Two complementary parts of each syllable were synthesized. These stimuli are indicated in Figure 4 for the series endpoints. The F2 and F3 transitions were synthesized together to form six chirps, one for each syllable. The six bases were generated from the remainder of the syllables (without

Figure 4. Schematic spectrograms of the three-formant [ba] and [ga] endpoints of the six-syllable series used for test stimuli in Experiment 2. The endpoint F2 and F3 transition pairs that were isolated as chirps are indicated by the dashed lines. The remainder of the syllables outside the dashed lines provided the bases.

the critical F2 and F3 transitions). The test stimuli were presented using the procedures described in Experiment 1.

Procedure. The basic experimental procedure was the same as that used in Experiment 1. The major difference was that the subjects participated in only four of the five conditions previously used. The subjects heard electronically mixed syllables, duplex syllables, isolated transitions, and the bases for these conditions. Each condition consisted of a set of practice trials and a set of experimental trials. In the practice trials, subjects identified three repetitions of each stimulus in random order. No feedback was given on the practice or experimental trials. The data from the practice trials were discarded. In every condition, following the block of practice trials the subjects were presented with 20 repetitions of each of the six test stimuli in random order. The subjects identified the stimuli in all conditions using a 6-point rating scale. A rating of "1" was used to indicate a good example of [ba]; a rating of "6" indicated a good [ga]. Variations between the exemplar [ba] and [ga] were assigned ratings "2" through "5." The subjects responded by pressing the appropriately marked button on a computer-controlled response box. As in the first experiment, the subjects were instructed to identify all the stimuli in every condition as the syllables [ba] or [ga], using the rating scale. If the subjects heard any test stimulus as nonspeech, they were encouraged to "guess" from which syllable the sound was derived.

Results and Discussion

As expected, the six bases were not differentially identified when presented in isolation. The mean rating of the [ba] endpoint base was 1.3, and the mean rating of the [ga] endpoint base was 1.4. As a result, the data from this condition were excluded from subsequent analyses.

As in the first experiment, subjects were able to reliably identify the isolated transitions. The isolated F2 and F3 transitions from the [ba] endpoint syllable were rated 1.3 (i.e., "ba") and the transitions from the [ga] endpoint syllable were rated 5.4 (i.e., "ga") in isolation. Thus, subjects could extract some pho-

netic information from the chirps in the absence of an appropriate syllabic context. Moreover, the mean ratings of the [ba] and [ga] endpoints in the duplex condition were quite similar to the ratings of the isolated transitions. For the [ba] endpoint, the mean rating in the duplex condition of 1.2 was not significantly different from the rating of the isolated [ba] endpoint transitions [$t(11) = .48$, n.s.]. For the [ga] endpoint, the mean rating of 5.9 in the duplex condition was not significantly different from the mean rating of the isolated [ga] endpoint transitions [$t(11) = .91$, n.s.]. These results replicate the findings of the first experiment, using a more sensitive measure of identification and using F2 and F3 isolated transitions. Therefore, these results support the conclusion that subjects can directly extract phonetic information from the isolated transitions. In addition, the similarity of the endpoint ratings in the duplex and isolated transition conditions implies that, in the duplex condition, subjects could have ignored the base in making their judgments.

The category boundaries were determined by linear interpolation using the rating data. The top panel of Figure 5 shows the rating data from the condition in which the transitions and bases were electronically mixed together to form intact syllables (the solid line) and the mean ratings from the duplex condition (the

Figure 5. In the top panel, the mean rating function for the whole syllables in the electronically mixed condition (solid line) and mean ratings of the duplex percepts (dashed line). In the bottom panel, the mean rating function for the whole syllables in the mixed condition (solid line) and the mean ratings of the isolated F2 and F3 transition pairs (dashed line).

dashed line). According to the integration hypothesis, the category boundaries in these conditions should be the same, since, in the duplex condition, the chirp and base should be perceptually fused across ears. The rating data from the mixed condition (the solid line) and the mean ratings of the isolated F2 and F3 transitions (the dashed line) are shown in the bottom panel of Figure 5. A one-way analysis of variance was conducted for the category boundaries in the mixed, duplex, and isolated transition conditions. The analysis revealed a significant effect of condition on the location of the category boundaries [$F(2,22) = 6.86$, $p < .005$]. The prediction made by the integration hypothesis was that the category boundary for the isolated transitions should be different from the boundaries for the duplex percepts and the whole syllables, while the duplex and whole syllable boundaries should be the same. Post hoc Newman-Keuls tests indicated that the obtained data were inconsistent with these predictions. The category boundary in the duplex condition was significantly different from the boundary in the mixed condition, in which whole syllables were heard ($p < .05$). In addition, the category boundaries for the isolated transitions and the electronically mixed syllables were significantly different ($p < .05$). Finally, the boundaries for the isolated transitions and the duplex percepts were not statistically different.

These results demonstrate that phonetic perception in the duplex condition is quite different from perception of intact syllables. These findings suggest that listeners do not simply fuse the transitions and base in the duplex condition to form a perceptually intact syllable. Instead, listeners may produce phonetic responses without attending to the base.

One explanation of the perceptual difference between duplex and intact syllables is that there are differential auditory masking effects in the binaural (intact) and dichotic (duplex) presentations of syllables (Rand, 1974). In fact, this auditory masking explanation could also account for the difference in the boundary for isolated transitions and the boundary for intact syllables. However, this explanation cannot be invoked for the articulation-based dichotic integration hypothesis, since proponents of this position have explicitly stated that general auditory processes have no role in mediating phonetic perception (Liberman, 1974; Repp, 1982; Studdert-Kennedy, 1981). To acknowledge an auditory masking account of our findings would entail implicit acceptance of the proposition that phonetic processes are subsidiary to and dependent upon generalized auditory processes. Of course, this is the antithesis of the view that phonetic judgments are mediated by mechanisms that are distinct and independent from auditory processes (cf. Schouten, 1980).

Furthermore, recent experiments conducted by

Schwab (1981) have shown that auditory masking effects in sine-wave analogs of syllables occurred only when the stimuli were heard as nonspeech signals in an auditory mode of perception. When these same stimuli were perceived as speech and labeled phonetically, no auditory masking effects were found. Thus, since these stimuli were all labeled phonetically, it would be hard to reconcile Schwab's results with a masking explanation of the difference between duplex and intact boundaries.

Since masking effects cannot be invoked to explain the boundary differences, these results argue against a dichotic integration account of duplex perception. The difference in the location of the duplex and intact labeling boundaries might reflect identification of the chirps without the bases in the duplex condition. Without the bases, the influence of the F1 transition would be eliminated from phonetic labeling of the chirps in both the duplex and isolated transition conditions. The absence of the F1 transition might explain the shift in the rating functions for these conditions compared with the intact syllables (see Jusczyk et al., 1981, for a discussion of the importance of the F1 transition in place of articulation judgments).

It is important to emphasize that both the auditory masking explanation and the missing F1 explanation of our results are incompatible with the hypothesis that dichotic integration in duplex perception is mediated by a specialized speech processor. However, both of these explanations are consistent with the chirp-identification hypothesis. Thus, the results of this experiment, in which subjects were able to identify the isolated transitions as speech, support the claim that listeners do not utilize the base in the duplex condition. Moreover, the similarity of the rating functions for the duplex and the isolated transition conditions suggests that perception in these conditions occurs in much the same way—by identifying the chirps without the base.

GENERAL DISCUSSION

During speech perception, a listener must integrate various acoustic segments to form phonetic percepts. Early research had indicated that isolated acoustic segments do not provide sufficient information to support speech perception (e.g., Mattingly et al., 1971). That is, these acoustic segments do not sound like speech when removed from the appropriate context. However, our results indicate that despite the nonspeech quality of isolated F2 transitions and F2 and F3 transition pairs, subjects can derive enough phonetic information to produce correct phonetic responses under the appropriate task constraints.

Of course, it is not our intention to claim that listeners perceive fluent natural speech by identifying as phonemes only some of the acoustic segments in

an utterance. It is apparent that the phonetic load carried by a single acoustic attribute in *natural* speech varies greatly as a function of phonetic context (see Dorman et al., 1977). In listening to natural speech, isolated transitions may not always provide the same amount of phonetic information. However, in duplex perception, the stimuli are synthetic syllables in which most of the phonetic load is carried by the isolated acoustic attribute. Furthermore, in these experiments, the alternative responses available to the subjects are constrained by the experimenter and only a single vowel context is used. As a result of these constraints, subjects may be able to produce correct phonetic responses from even a small amount of phonetic information derived from the chirp. Thus, in any speech experiment (duplex or otherwise) in which isolated acoustic features are presented as nonspeech, it is important to include a control condition that assesses the amount of phonetic information subjects can directly extract from an isolated acoustic segment. Without this condition, it is impossible to determine how much more information is contributed by hearing the acoustic attribute in the appropriate syllabic context.

Our results provide support for an alternative interpretation of duplex perception and thus weaken the conclusions from previous duplex perception research that used isolated F2 transitions (Cutting, 1976) and F2 and F3 transition pairs (Liberman et al., 1981; Rand, 1974) as chirps. In these experiments, the chirps provided sufficient information to account for the subjects' phonetic responses even without the base. Therefore, it is not necessary to claim that dichotic fusion of the complementary acoustic segments (the chirp and base) ever occurred to account for duplex perception. Duplex perception may simply demonstrate the same type of processing that occurs when listeners report that they can understand the linguistic content of a distinctly nonspeech-sounding stimulus, such as the sine-wave analogs of speech (cf. Remez et al., 1981).

Finally, our results do not completely rule out the possibility of dichotic fusion's occurring in duplex perception.³ However, these results do suggest that this dichotic fusion might not occur *prior* to phonetic labeling. Rather, fusion should occur *after* the phonetic features have been separately identified in the two ears (see Pisoni, 1975, for a discussion of phonetic feature fusion). To some extent, fusion of the phonetic features identified in each ear may account for the subjective experience in duplex perception that the speech stimulus is heard in the ear that receives the base. Also, this type of phonetic feature fusion could explain the apparent use of both the base and transitions for producing phonetic responses when the phonetic identity of the base varies (cf. Liberman et al., 1981).

In summary, two conclusions can be drawn from our results. First, under the experimental conditions used by Mattingly et al. (1971) and in previous duplex research (Cutting, 1976; Liberman et al., 1981; Rand, 1974), we found that subjects can extract some phonetic information directly from a nonspeech-sounding chirp in the absence of an appropriate context. Thus, this result contradicts earlier claims surrounding chirp perception made by those studies. Second, we have provided evidence to support an alternative explanation of duplex perception. Since subjects can use isolated transitions to produce phonetic responses, there is no need to invoke dichotic fusion to explain duplex perception in those previous studies. While it is true that we cannot directly show that subjects ignore the base in duplex perception, we have demonstrated that the presence of the base is not necessary to produce correct phonetic responses. However, it is also true that the results produced by different instructions (attend to the chirp or attend to the base) in previous duplex research do not necessarily demonstrate that duplex perception entails fusion. Rather, the effects of the different instructions in these experiments could be due to the induction of different perceptual expectations about the chirps. Taken together, our results indicate that when presented with an F2 chirp or an F2 and F3 chirp in duplex perception, subjects may simply identify the chirp and ignore the base.

REFERENCE NOTES

1. Kewley-Port, D. *Representations of spectral change as cues to place of articulation in stop consonants* (Research on Speech Perception Tech. Rep. No. 3). Bloomington, Ind: Department of Psychology, Indiana University, 1980.
2. Kewley-Port, D. *KLTEXC: Executive program to implement the KLATT software speech synthesizer* (Research on Speech Perception Progress Report No. 5, pp. 327-346). Bloomington, Ind: Department of Psychology, Indiana University, 1978.

REFERENCES

- BEST, C. T., MORRONGIELLO, B., & ROBSON, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, 29, 191-211.
- CUTTING, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 1976, 83, 114-140.
- DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, 22, 109-122.
- GRUNKE, M. E., & PISONI, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Perception & Psychophysics*, 1982, 31, 210-218.
- JUSCZYK, P. W., SMITH, L. B., & MURPHY, C. The perceptual classification of speech. *Perception & Psychophysics*, 1981, 30, 10-23.
- KLATT, D. H. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 1980, 67, 971-995.
- LIBERMAN, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, 1, 301-323.

- LIBERMAN, A. M. The specialization of the language hemisphere. In F. O. Schmidt & F. G. Worden (Eds.), *The neurosciences: Third study program*. Cambridge: M.I.T. Press, 1974.
- LIBERMAN, A. M. Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen, 1979.
- LIBERMAN, A. M. On finding that speech is special. *American Psychologist*, 1982, 37, 148-167.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- LIBERMAN, A. M., ISENBERG, D., & RAKERD, B. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, 1981, 30, 133-143.
- LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8): *Perception*. New York: Springer, 1978.
- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, 2, 131-157.
- PISONI, D. B. Dichotic listening and processing phonetic features. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.
- PISONI, D. B., & SAWUSCH, J. R. Some stages of processing in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. Berlin: Springer, 1975.
- RAND, T. C. Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, 1974, 55, 678-680.
- REMEZ, R. E., RUBIN, P. E., PISONI, D. B., & CARRELL, T. D. Speech perception without traditional speech cues. *Science*, 1981, 212, 947-950.
- REPP, B. H. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 1982, 92, 81-110.
- REPP, B. H., MILBURN, C., & ASHKENAS, J. Duplex perception: Confirmation of fusion. *Perception & Psychophysics*, 1983, 33, 333-337.
- SCHWAB, E. C. *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral dissertation, S.U.N.Y. at Buffalo, 1981.
- SCHOUTEN, M. E. H. The case against a speech mode of perception. *Acta Psychologica*, 1980, 44, 71-98.
- STUDDERT-KENNEDY, M. The emergence of phonetic structure. *Cognition*, 1981, 10, 301-306.
- WALLEY, A. C., & CARRELL, T. D. Onset spectra vs. formant transitions as cues to place of articulation. *Journal of the Acoustical Society of America*, in press.

NOTES

1. In a recent series of experiments, Jusczyk, Smith, and Murphy (1981) attempted to induce subjects to identify chirps as speech. The chirps were either isolated F2 and F3 transitions or isolated F1, F2, and F3 transitions taken from full syllables. The subjects were unable to learn to classify the two-formant (F2 and F3) chirps as phonetic segments, but they could classify the three-formant chirps correctly. At first glance, these results demonstrate that listeners cannot extract any phonetic information from isolated F2 and F3 transitions. However, these results are not directly relevant to previous duplex research for two reasons. First, Jusczyk et al. presented chirps that were only 30 msec in duration. In con-

trast, in the duplex experiments, subjects heard chirps that were either 70 msec (Cutting, 1976; Liberman et al., 1981) or 100 msec (Rand, 1974) long. It is possible that if Jusczyk et al. had presented longer chirps, their subjects might have been more successful in identifying the two-formant stimuli phonetically. Second, Jusczyk et al. presented chirps derived from consonants in four vowel contexts. In contrast, only one vowel environment (i.e., [a]) has ever been used in duplex experiments (cf. Cutting, 1976; Liberman et al., 1981; Rand, 1974). Since this vowel context was not included in the experiments reported by Jusczyk et al., it is hard to generalize from their results because the phonetic load carried by transitions varies across vowel environments (see Dorman et al., 1977). Indeed, Kewley-Port (Note 1) has demonstrated that when vowel identity is known in advance, the F2 and F3 transitions together provide an invariant cue to place of articulation for stop consonants paired with [a] in natural speech.

2. Admittedly, it is very difficult to conceive of a mechanism that is different from the normal process of speech perception and yet could be capable of producing phonetic responses. One possibility is that subjects can associate some perceptual attribute of the isolated chirps with the same transitions heard in intact syllables. This type of association might then mediate the phonetic responses used by subjects to label the chirps. If this associative attribute is phonetic, then subjects can derive some phonetic information from the chirps, which would be the same as the chirp identification hypothesis. Of course, it is possible that the associative attribute is auditory instead of phonetic. While this proposal would be consistent with perceptual theories that propose that phonetic labeling is predicted on general auditory features of speech (e.g., Pisoni & Sawusch, 1975), this account would be inconsistent with the more traditional view that phonetic processing is totally independent of more general auditory mechanisms. One of the fundamental claims of this position, that phonetic and auditory processes are independent, is that listeners have no access to the auditory components of a stop consonant—listeners perceive only the phonetic dimensions of consonants (e.g., Liberman, 1974). If listeners cannot decompose a phonetic segment into its auditory constituents, then they would have no basis for forming the auditory association between the chirps and the syllables from which the transitions were derived. Moreover, to propose that listeners form auditory associations that serve as the basis for phonetic responses is tantamount to accepting an auditory theory of phonetic perception (cf. Schouten, 1980.)

3. In a recent study, Repp, Milburn, and Ashkenas (1983) found that duplex perception could be obtained even when the chirps were F3 transitions that were not identifiable in isolation. While this study purportedly demonstrates that duplex perception results from dichotic fusion, this conclusion is not the only interpretation possible. For example, it is possible that the context of the base in one ear facilitates the extraction of phonetic information from the chirp in the other ear. Even though Repp et al. demonstrated that subjects did not label the isolated chirps correctly in one condition, this does not mean that it is impossible for subjects to extract phonetic information from these isolated chirps. Furthermore, even if we accept the hypothesis that duplex perception results from dichotic fusion, Repp et al. did not establish the level at which this fusion occurs. According to the chirp-identification hypothesis, if fusion does occur, it should take place after some phonetic processing of the chirp. Finally, although dichotic fusion may be a reasonable explanation of the results obtained by Repp et al., there is still no reason to assume that such fusion occurred when the chirps could be identified in isolation, as in the earlier duplex research and the present experiments.