Possible mechanisms of duplex perception: "chirp" identification versus dichotic fusion

HOWARD C. NUSBAUM

Indiana University, Bloomington, Indiana

Within the past few years, the phenomenon of duplex perception has received a great deal of attention from researchers interested in speech perception. To produce duplex perception, a listener is dichotically presented with two complementary components of a syllable. An isolated formant transition (or transition pair) is presented to one ear while the other ear receives the remainder of the syllable (called the "base"). Synchronous dichotic presentation of these stimuli results in two percepts—a brief nonspeech "chirp" or glissando corresponding to the isolated transition and an intact syllable correctly cued by the isolated transition (see Cutting, 1976; Rand, 1974). It is the dual role of the isolated transition in both the nonspeech "chirp" and the perceptually intact syllable that gave rise to the appellation of "duplex perception" (Liberman, 1979).

The reason for the interest in duplex perception is quite straightforward. The phenomenon of duplex perception has been interpreted as one important source of evidence for the claim that speech perception is mediated by perceptual mechanisms that are uniquely specialized for phonetic processing (see Liberman, 1982; Repp, 1982; Studdert-Kennedy, 1983). According to this view, perception of the isolated transition as a chirp is mediated by a generic auditory information processor, whereas a phonetic processor is responsible for perception of the perceptually intact syllable (see Liberman, 1982; but cf. Pastore, Schmuckler, Rosenblum, & Szczesniul, 1983).

Dichotic Fusion and Duplex Perception

One question that can be raised about duplex perception concerns how the listener derives phonetic information from the dichotic presentation of an isolated transition and a base. The answer to this question has generally been that the phonetic percept results from *dichotic fusion* of the two stimuli (e.g., see Cutting, 1976; Liberman, Isenberg, & Rakerd, 1981; Repp, Milburn, & Ashkenas, 1983). That is, the transition and base become fused to form a per-

ceptually complete syllable that is processed phonetically. Although at first glance this answer seems to provide an adequate explanation, on closer inspection there appear to be several problems with this account, contrary to the claims made by Repp (1984).

First and foremost, just what does it mean to assert that the phonetic percept results from dichotic fusion of the transition with the base in duplex perception? According to Repp et al. (1983), "the chirp and base fuse at a relatively early stage in processing" (p. 336). Unfortunately, Repp et al. remain vague on the precise nature of this early processing stage, although, by referring to research conducted by Cutting (1976), they imply that fusion occurs at an auditory (rather than phonetic) level of processing. This suggests that the transition and base are fused to form a unitary auditory representation that is subsequently labeled by phonetic processes (see Cutting, 1976). However, Repp et al. also cite work by Liberman and his colleagues (Liberman et al., 1981; Mann & Liberman, 1984) as providing the "standard interpretation" of dichotic fusion in duplex perception. (Presumably, this is an interpretation that Repp et al. accept.) Mann and Liberman (1984) assert that "incorporation of the transitions into stop percepts, and, in particular, the contrast this presents to their perception as chirps, reflects a specialized phonetic process" (p. 232). Similarly, Liberman (1982) has claimed that in duplex perception "integration of the formant transitions into a phonetic percept is due to a special process that makes available to perception a unitary phonetic object well suited to its role in language" (p. 154). In other words, the standard interpretation of duplex perception deferred to by Repp et al. states that fusion of the transitions with the base is carried out by a *phonetic* process, contradicting the claim also made by Repp et al. that fusion occurs at an early *auditory* stage of processing. This contradiction clearly points out a major problem with current explanations of duplex perception: Vague and imprecise appeals to "dichotic fusion" are not sufficient to account for phonetic processing in duplex perception.

A second problem with the dichotic fusion account of duplex perception concerns the application of the basic concept of fusion to this phenomenon. The concept of fusion (in audition or vision) generally refers to the combination of physically independent stimulus elements to form a unitary percept. In other words, perceptual fusion (either dichotic or binocular) produces a single perceptual object from separate stimuli. Indeed, of the six dichotic fusions described by Cutting (1976), duplex perception is the one least likely to produce a single fused percept. In fact, while most dichotic fusions are characterized by

Preparation of this manuscript was supported by NIH Training Grant NS-07134 to Indiana University. I am grateful to David B. Pisoni and Eileen C. Schwab for their helpful comments and suggestions. Reprint requests should be sent to Howard C. Nusbaum, Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405.

a single percept (e.g., Halwes, 1969; Repp, 1976), the hallmark of duplex perception is the recognition of the two separate stimulus elements (i.e., the transition and the base) as two distinct perceptual objects (i.e., a nonspeech chirp and a syllable). Thus, despite the assertion that fusion occurs in duplex perception (e.g., Repp et al., 1983), each of the two stimuli maintains its own distinctive auditory characteristics and is perceived separately. Since dichotic fusions occurring as early in the perceptual system as sound localization and as late as phonological fusion generally result in a single percept (see Cutting, 1976), it is not clear why dichotic fusion in duplex perception should result in two percepts.¹ Of course, this suggests the possibility that some mechanism (or mechanisms) other than dichotic fusion may be operating in duplex perception.

Furthermore, despite arguments by Repp (1984) to the contrary, there are two empirical findings that argue against an account of duplex perception that depends on dichotic fusion at an early auditory stage of processing. First, Rand (1974) found that attenuating F2 and F3 formant transitions impaired phonetic identification of intact syllables more than it impaired phonetic labeling of the duplex percept. Rand interpreted this result as evidence for release of masking of the formant transitions when the transitions were presented dichotically with the base. Nusbaum, Schwab, and Sawusch (1983) cited this as evidence "that transitions are processed differently in an intact syllable and on the speech side of the duplex percept" (p. 325). The rationale for this conclusion was quite straightforward and stems from two possible interpretations of this dichotic release from masking. One interpretation is that the perceptual representation formed from the dichotic presentation of the transitions and the base is more resistant to masking than the representation resulting from an intact syllable. This suggests that the process that produces this representation in duplex perception is different from the process that generates this representation from physically intact syllables. The alternative interpretation is that the dichotic combination of transition information with the base occurs at a level of perception that is later than (or beyond) the point of auditory processing where masking normally occurs for an intact syllable. Clearly, then, the speech side of the duplex percept must be processed in a manner different from that of an intact syllable. Moreover, invoking a "psychoacoustic" explanation of this difference (Repp, 1984) does not weaken or modify this claim, since it is quite likely that the differences in processing an intact syllable and a duplex percept arise at an auditory stage of processing. Indeed, one possible psychoacoustic account is that the chirp and the base are "streamed" into different auditory channels of analysis and do not fuse at all (cf. Steiger & Bregman, 1982), thereby reducing the

effectiveness of the F1 transition in the base as a mask for the chirp information in the other ear.

Nusbaum et al. (1983) also discussed a second result that argues against an early auditory fusion account of duplex perception. Cutting (1976) found that increasing the disparity in fundamental frequency (F0) between the isolated transition and the base did not measurably affect duplex perception. Nusbaum et al. argued that increasing the F0 disparity between the two stimuli should reduce the probability of fusion by indicating that these stimuli were produced by different sources. Conversely, the lack of an effect of F0 disparity on fusion responses in duplex perception suggests the possibility that the transition and base may not be fused at all in duplex perception. Although Repp (1984) claimed that F0 is a poor index of articulatory source, and therefore should not affect fusion of speech stimuli, several experiments have demonstrated that F0 disparity between dichotically presented speech stimuli does attenuate early auditory fusion (see Broadbent & Ladefoged, 1957; Cutting, 1976; Darwin, 1981). Thus, the lack of an F0-disparity effect in duplex perception does argue against early auditory fusion of the chirp and base.

While these problems with the dichotic fusion account of duplex perception do not conclusively reject this interpretation, these problems do imply that the fusion hypothesis is neither as well supported nor as well defined as Repp would have us believe (Repp, 1984; Repp et al., 1983). Indeed, it is simply not clear how dichotic fusion operates to produce the duplex percept, and, until the details of this hypothesis are clarified, it will be difficult to derive any specific tests of this hypothesis. Moreover, until the dichotic fusion account is described more precisely, duplex perception cannot really be taken as strong support for the existence of a specialized processor dedicated to and optimized for phonetic perception (cf. Liberman, 1982; Repp, 1982).

"Chirp" Identification and Duplex Perception

The dichotic fusion account of duplex perception is considered to provide an important argument for the claim that speech perception is mediated by a specialized phonetic process (e.g., Liberman, 1982). According to this position, in speech production, the acoustic cues to one phoneme overlap and interact with the acoustic cues to other, adjacent phonemes resulting in an encoded acoustic representation of the phonemes in speech (see Liberman, 1970). The phonemic interpretation of each acoustic segment should therefore be dependent on the acoustic context in which the segment was produced. Thus, speech perception is viewed as a context-dependent process that must decode the acoustic structure of speech into phonetic segments (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). One source

of support for this view was the finding that an F2 transition was perceived as a nonspeech chirp in isolation and as a phonetic cue when embedded in an intact syllable (Mattingly, Liberman, Syrdal, & Halwes, 1971).² In the case of duplex perception, however, subjects can phonetically label the duplex percept even though the transition and its context (i.e., the base) are physically separated and presented to different ears (e.g., Liberman et al., 1981). This ability to produce phonetic responses to a cue separated from its context (in duplex perception) should pose a problem for the view of speech perception that requires acoustic cues to be interpreted phonetically only within an appropriate syllabic context. This potential problem was solved by the assumption that dichotic fusion reunites the isolated transition with its context.

However, Nusbaum, Schwab, and Sawusch (1983) offered an alternative account of duplex perception. Rather than assume the operation of dichotic fusion, Nusbaum et al. were interested in investigating the separate contributions of the individual stimulus elements (i.e., the transition and the base) to phonetic processing of the duplex percept. Since the base is phonetically redundant (constant), we suggested that perhaps subjects do not need to fuse the transition with the base to generate a phonetic response. Instead, subjects might be able to derive some phonetic information directly from the isolated transition (i.e., the chirp) itself.

Of course, this perspective represented a change from the view that phonetic information could be decoded only from acoustic cues in the appropriate context. In part, this shift in perspective was facilitated by research on phonetic processing of tone analogs of speech (see Grunke & Pisoni, 1982; Remez, Rubin, Pisoni, & Carrell, 1981; Schwab, 1981). This research has demonstrated that phonetic information can be derived from nonspeech signals when subjects expect to hear these sounds as speech. (Naturally, this assumes that phonetically appropriate cues exist in the nonspeech stimulus.) In fact, listening to the sine-wave analogs is something akin to duplex perception in that the listener perceives phonetic information and, at the same time, hears the distinctly nonspeech auditory quality of the signal (see Remez, Rubin, & Pisoni, 1983).

Nusbaum et al. (1983) hypothesized that subjects might be able to derive phonetic information directly from an isolated transition under the appropriate conditions, even though the isolated transition has a distinct nonspeech quality. In two experiments testing this hypothesis, Nusbaum et al. asked subjects to use phonetic categories to identify isolated transitions, bases, intact syllables, and duplex percepts. Contrary to previous claims (e.g., Mattingly et al., 1971), the results clearly demonstrated that subjects could accurately identify the phonetic information encoded in the isolated transitions (i.e., the chirps) in the ab-

sence of an appropriate context. These results indicate that it is not necessary to assume that a transition must be fused with its base as a prerequisite for phonetic processing. Instead, listeners may simply identify the phonetic information that is inherent in the chirp.

Recently, Repp (1984) has raised a number of criticisms of this study in an attempt to reject the chirp-identification hypothesis. In essence, Repp has argued that the assumptions of the chirp-identification hypothesis are faulty and that there are problems with the design of the experiments testing this hypothesis. Furthermore, Repp has argued that several experiments disconfirm the chirp-identification hypothesis. Although he has presented a wide range of criticisms, none of these is difficult to answer, and therefore, it would be quite premature to reject the chirp-identification hypothesis.

Several of Repp's criticisms seem to result from a basic misunderstanding of the scope of the chirp-identification hypothesis. For example, Repp claimed that one assumption of the chirp-identification hypothesis was that listeners incorrectly locate the phonetic response derived from the chirp onto the base because of the speech-like quality of the base. He then went on to criticize this purported assumption. In fact, this is not an assumption of chirp identification at all. Rather, this was an attempt made by Nusbaum et al. to speculate on a possible mechanism that, once the phonetic response was determined, would assign this response to the base rather than the chirp. Similarly, Repp also criticized Nusbaum et al. for considering the possibility that fusion could occur after phonetic processing of the chirp. However, contrary to Repp's claim that this was a contradiction of the chirp-identification hypothesis, this was simply further speculation about the type of process that might mislocate the phonetic response onto the base. In fact, as Nusbaum et al. pointed out, chirp identification and fusion are not completely incompatible hypotheses as long as fusion is used only to account for the *illusion* that the phonetic response arises from the base and not the chirp. Furthermore, Repp suggested that "since both the base and the chirp carry place-of-articulation and manner information, fusion after labeling would frequently result in the perception of *two* consonants" (p. 91). This is a surprising assertion to make given that, in phonetic feature fusion, two intact CV syllables (both containing manner and place of articulation information) can be fused to form a single percept containing only one consonant (e.g., Halwes, 1969).

The dichotic fusion interpretation of duplex perception was proposed to account for both the phonetic response *and* the localization of that response in the ear that received the base. However, the chirp-identification hypothesis was proposed to account only for phonetic identification of the duplex percept. A separate mechanism would be required to ac-

count for the illusion that the phonetic response arises from the ear with the base. This separation of duplex perception into a phonetic recognition process and a localization process is similar to the distinction made by Deutsch and Roll (1976) between *what* and *where* decision mechanisms in general auditory perception—the *what* mechanism is responsible for object identification and the *where* mechanism is responsible for object location. Indeed, with respect to perception of tone sequences and the octave illusion, Deutsch (1980) has claimed that “these *what* and *where* mechanisms operate at some stage so independently that we can end up perceiving a stimulus that does not exist, that is, with its pitch taken from one source and its location from another” (p. 587). Substituting “phonetic identity” for “pitch” in that quote would produce a reasonable description of duplex perception.

However, Repp (1984) has stated that “an inability to attribute the response to its correct stimulus would be expected only in the case of fusion” (p. 90). This seems to be quite an overstatement, because there are several perceptual phenomena in which subjects incorrectly attribute an identification response to the wrong stimulus. One example is the octave illusion described by Deutsch (1980; Deutsch & Roll, 1976) which cannot be explained simply by dichotic fusion. Similarly, Treisman and Schmidt (1982) have reported illusory conjunctions in vision, whereby perceptual features from one object are mislocated onto a different object. These illusory conjunctions are attributed to a failure of selective attention (see Treisman & Gelade, 1980) rather than “fusion,” and are similar to the “blend” errors that occur in dichotic listening (see Pisoni, 1975). Thus, in illusory conjunctions, subjects do not fuse two stimuli into one percept, but instead incorrectly assign perceptual features of one stimulus to another.

Repp (1984) has also questioned another assumption made by the chirp-identification hypothesis. We suggested that when subjects are instructed to identify the duplex percept using phonetic categories, they rely on the phonetic information inherent in the chirp rather than on the constant base. Repp called this assumption bizarre and asserted that subjects have no difficulty using the same response on every trial. The implication was that, without fusion, if they were asked to classify the duplex percept using phonetic categories, subjects would respond using the most speech-like stimulus—the base. Thus, they would produce the same response for every dichotic presentation. By referring to the data collected by Nusbaum et al. (1983) for phonetic identification of the isolated base, Repp attempted to support his assertion that subjects were not hesitant to use the same response on every trial. However, it is not clear why Repp believes these data support his contention. An examination of the responses made by individual subjects in the first experiment reported by Nusbaum

et al. indicated that 13 of the 17 subjects used *both* phonetic categories to identify the acoustically constant base [$\chi^2(1) = 18.13$, $p < .001$]; the subjects just responded with one category (i.e., [b]) more frequently. Thus, the actual data directly contradict Repp's claim. Clearly, the subjects were trying to identify the place of articulation of the bases despite the fact that they heard the same stimulus on each trial. Therefore, it does not seem bizarre at all to assume that the subjects, in attempting to comply with the experimenter's instructions, would use the phonetically informative chirp (when it was present) as the basis for their responses, since there was no distinctive information contained in the base.

Repp (1984) also criticized the design of the experiments conducted by Nusbaum et al. (1983). First, Repp suggested that the stimuli used in that research were somehow optimized for chirp identification. In fact, in the first experiment, the stimuli “were based on the two-formant [ba] and [ga] syllables used by Cutting (1976)” (Nusbaum et al., 1983, p. 326). Thus, these stimuli were chosen specifically to *replicate* earlier duplex research. Part of the reason for this choice was the difficulty we encountered in pilot experiments trying to reproduce the phenomenon of duplex perception. In our pilot work, using syllables modeled on spectrograms of a male talker's CV syllables, we found that few, if any, of our subjects experienced duplex perception. Instead, the subjects reported hearing the base *and* the chirp on each trial, with the base remaining constant across trials. In order to obtain duplex perception, it was necessary to use stimuli for which duplex perception had been reported previously. Furthermore, these two-formant syllables and chirps were similar to the stimuli used by Mattingly et al. (1971) to claim that isolated transitions are perceived only as nonspeech. In addition, similar stimuli were used recently in another duplex experiment reported by Bentin and Mann (1983). Clearly, these two-formant syllables and chirps (or their equivalents) have been chosen for several other experiments concerned with chirp and duplex perception. As a consequence, and despite Repp's comments, there was considerably justification for this choice of stimuli, especially since the results reported by Nusbaum et al. have important implications for other research using these (or similar) stimuli. In the second experiment reported by Nusbaum et al., these two-formant syllables were modified slightly and a third formant was added to create a series from [ba] to [ga]. These three-formant stimuli also allowed us to generalize the conclusions from the first experiment which used F2 chirps and two-formant syllables to duplex research that used F2 and F3 transition pairs for the chirps (e.g., Liberman et al., 1981; Rand, 1974).

Repp also criticized the first experiment reported by Nusbaum et al. for accepting the null hypothesis entailed by the similarity of response patterns for

phonetic identification of the duplex percept and isolated transitions. However, the purpose of our first experiment was to test the major premise of the dichotic fusion hypothesis, which asserts that isolated acoustic cues must be fused with an appropriate acoustic context in order to be perceived as speech. The prediction of the dichotic fusion hypothesis was that F2 transitions should provide phonetic information only in intact syllables and duplex percepts. By contrast, the chirp-identification hypothesis predicted that phonetic labeling of isolated transitions should not depend on the presence of an appropriate syllabic context either in the form of an intact syllable or the duplex percept. Indeed, the results of this experiment provided an existence proof demonstrating that phonetic information in isolated transitions can be recovered without dichotic presentation of the base. In other words, listeners could accurately classify the place of articulation of isolated F2 transitions, in direct contradiction to previous claims (e.g., Mattingly et al., 1971). Given that the isolated chirp is sufficient to support phonetic labeling, it is not necessary to postulate the operation of dichotic fusion in duplex perception.

The second experiment reported by Nusbaum et al. was criticized by Repp as being poorly motivated. That experiment was designed to investigate the process of phonetic labeling for intact syllables, duplex percepts, and isolated transition pairs. The dependent measure of interest in this experiment was the location of the phonetic category boundary within a series of stimuli varying in onset frequency of the F2 transition. The category boundary is generally interpreted as the point in a stimulus series at which two alternative phonetic interpretations are equally probable (see Eimas & Corbit, 1973). The locus of this boundary may be affected by the interaction of various acoustic cues in phoneme recognition (see Oden & Massaro, 1978; Sawusch & Pisoni, 1974). In our second experiment, the F1 transition was the critical cue relevant to the phonetic decision. Stevens and Blumstein (1978) have suggested that the F1 transition may provide a cue to place of articulation. In addition, previous research has indicated that the F1 transition is an important cue to manner of articulation (e.g., Suzuki, 1970), and several studies have demonstrated the influence of manner of articulation on perception of place of articulation (see Carden, Levitt, Jusczyk, & Walley, 1981; Eimas, Tartter, & Miller, 1981). In our second experiment, for the isolated chirp series, the F1 transition was absent, whereas for the intact syllable series, this cue was present. However, the critical stimulus continuum was the duplex series. In duplex perception, if the chirp fused with the base, the presence of the F1 transition should affect phonetic labeling, as in the intact syllables, and the duplex boundary should have been in precisely the same location as the boundary for the intact syllables. By contrast, if

chirp and base were not fused, the presence of the F1 transition in the base should not influence phonetic labeling of the duplex series (as in the chirp series). The duplex boundary therefore should be in the same location as the chirp boundary. In fact, the results indicated that the dichotic presentation of the F1 transition in the base did not influence phonetic labeling of the duplex stimuli; the chirp boundary and the duplex boundary were both significantly displaced from the location of the intact boundary. Even if release of the F2-F3 transition pairs from masking by the F1 transition could account for this result (as suggested by Repp, 1984), this interpretation would still require that the duplex percept be processed differently from the intact syllable, as discussed earlier. However, it is not clear that masking should affect the location of the phonetic category boundary, although place-manner interactions at the category boundary have been obtained previously (Carden et al., 1981). Also, the presence of the F1 transition has been linked to the perceptual stability of phonetic categories across different vowel contexts (see Jusczyk, Smith, & Murphy, 1981)—a result that can hardly be attributed to masking of higher formant transitions by the F1 transition.

Finally, Repp cited several studies that he claimed provided data against the chirp-identification hypothesis. First, Repp questioned whether the chirp-identification hypothesis could account for the finding that correct phonetic responses decrease with increasing temporal asynchrony between the isolated chirp and the base (see Cutting, 1976). However, this result may simply reflect dichotic recognition masking between perceptual processing of the base and the chirp (cf. Pisoni, 1975).

Repp also questioned how the chirp-identification hypothesis could account for the finding that instructions to respond to the chirp side of the duplex percept produce a different pattern of responses than instructions to attend to the speech side of this percept. Indeed, this is one question that Repp raised several times both in connection with earlier research (Liberman et al., 1981; Mann & Liberman, 1984) and the second experiment reported by Repp et al. (1983). The assumption made by Repp is that instructing subjects to respond to the chirp or the base is sufficient to insure *selective attention* to that percept. However, there is no evidence that listeners (especially untrained and unpracticed listeners) are able to selectively attend to one side of the duplex percept or the other. Indeed, if duplex perception results from a failure of selective attention such as is found in illusory conjunctions (cf. Treisman & Gelade, 1980), this assumption may be wrong. Instead, subjects may interpret selective attention instructions as instructions to respond to either the auditory properties of the duplex percept or the phonetic properties of this percept (see Nusbaum et al., 1983). For example, the results obtained by Mann and Liberman

(1984) showing different patterns of discrimination for the chirp side of the duplex percept and the speech side may reflect differences between auditory processing of the chirp and phonetic processing of the chirp. To date, no duplex experiment has attempted to determine if subjects are indeed selectively attending to one side or the other of the duplex percept. Of course, this would require measuring decrements in performance of a secondary task contingent upon different selective attention conditions (cf. Nusbaum, 1981). The flexibility shown by subjects in adopting different perceptual strategies in response to task demands is well established, and any claims regarding selective attention should be supported by appropriately designed experiments (see Duncan, 1980). The claims made by Repp (1984) and Repp et al. (1983) regarding selective attention in duplex perception are completely unsupported empirically. Until the appropriate conditions are investigated, the results of this research that purportedly manipulated selective attention (Liberman et al., 1981; Mann & Liberman, 1984; Repp et al., 1983) cannot be interpreted unequivocally.

The final study that bears directly on the chirp-identification hypothesis is the first experiment reported by Repp et al. (1983). This experiment was conducted to test the hypothesis that "isolated formant transitions should be identifiable as the consonants they are intended to cue" (Repp et al., 1983, p. 333). In this experiment, Repp et al. attempted to replicate the results obtained by Nusbaum et al. (1983) using different stimuli—isolated F3 transitions, intact syllables, and duplex percepts. The results of this experiment suggested that their subjects were unable to reliably identify the F3 chirps using phonetic categories, even though the subjects could phonetically label intact syllables and duplex percepts that were cued by these transitions. Repp et al. concluded that fusion of the F3 transition with the base was necessary to identify the phonetic information encoded in the F3 chirps. However, drawing this conclusion from a null result (i.e., the inability of subjects to identify the chirps with phonetic categories) is specious.³ Numerous explanations for any null result, such as the inability to perform a task, are possible. For example, subjects are quite sensitive to an experimenter's expectations about the outcome of an experimental condition, even when these expectations are never stated explicitly. These tacitly communicated expectations can have an extremely powerful influence on performance (see Intons-Peterson, 1983). The poor labeling performance obtained for F3 chirps by Repp et al. could reflect the subjects' response to the researchers' strong interest in disconfirming the chirp-identification hypothesis. Indeed, an examination of the F3-chirp labeling performance reported by Repp et al. (1983) for individual subjects is quite interesting for the inconsistency shown by

most subjects. Since the F3 chirps should be discriminable, if the subjects were appropriately motivated to identify these stimuli using phonetic categories, each subject should adopt a consistent (although possibly wrong) labeling strategy. By contrast, the individual identification functions indicated that subjects generally did not even differentially identify the endpoints of the test series. Thus, it appears that the subjects might not have been sufficiently motivated by the instructions to identify the chirps as speech.

Another possible explanation of the inability of subjects to identify the F3 chirps using phonetic labels is that for some acoustic cues, phonetic labeling is too insensitive a procedure to measure the amount of phonetic information directly extracted from some isolated cues. Previous research has demonstrated that in some speech perception tasks, a single forced-choice identification response may underestimate the amount of information available to the subject (see Green & Swets, 1974). Thus, it is possible that, for some cues such as the F3 chirp, more sensitive procedures are required to determine the amount of phonetic information directly available to the listener.

However, even if the results reported by Repp et al. are taken at face value, it is still not necessary to assume the operation of dichotic fusion in duplex perception. The only conclusion that can be drawn from these results is that somehow the presence of the base facilitates identification of the phoneme cued by the F3 chirp. Although this might argue against the strongest form of the chirp-identification hypothesis, in which phoneme recognition should be completely independent of the base, this result still does not completely rule out chirp identification as a viable alternative to dichotic fusion.

Previous research has indicated that the isolated F3 transition is not as reliable a cue for place of articulation as is the isolated F2 transition (see Kewley-Port, 1982). In the experiments reported by Repp et al., in which isolated F3 transitions served as chirps, the match between the F3 transitions and the subjects' phonetic prototypes might have been much weaker than in the Nusbaum et al. (1983) experiments, which used F2 and F2-F3 chirps. Recently, Keyser, Stevens, and Kawasaki (1983) have suggested that redundant information (such as the base) may enhance phoneme recognition when the distinctive cue (such as the F3 chirp) is weak. Thus, the subjects might have used the base to facilitate phonetic identification in the duplex condition. However, it is certainly not necessary to postulate that the chirp and base are fused, even if subjects did consider the base in producing phonetic responses. Instead, duplex perception may still be characterized by the independent extraction of phonetic information from the chirp and the base, and both of these sources of information may contribute to the phonetic percept.⁴

Conclusion

The chirp-identification hypothesis was proposed by Nusbaum et al. to account for duplex perception without the assumption that dichotic fusion was necessary for phonetic processing. Indeed, Nusbaum et al. demonstrated that, contrary to previous widely believed claims (e.g., Mattingly et al., 1971; Liberman, 1982), it was possible to directly extract phonetic information from an isolated acoustic cue. Thus, these results support the basic contention of the chirp-identification hypothesis, and therefore, they suggest that, contrary to Repp's claims, the results of previous duplex experiments (e.g., Liberman et al., 1981; Mann & Liberman, 1984; Repp et al., 1983) can be accounted for without the assumption of dichotic fusion. Of course, these findings do not imply that all acoustic cues are sufficient to support phonetic processing. By virtue of its encoded nature, speech can be viewed as a multidimensional signal in which many cues may cooperate and contribute to phoneme recognition. However, our results do suggest that, through the redundancy inherent in the speech signal, some cues may provide sufficient information to support phonetic processing in isolation, and perhaps these cues contribute more to phoneme recognition in some contexts than other cues.

Finally, if it is not necessary to assume that the chirp and base must be fused for phonetic labeling to occur, duplex perception cannot be taken as evidence for the operation of a specialized phonetic processor. Indeed, Pastore et al. (1983) have recently demonstrated that duplex perception can be obtained with musical (nonspeech) stimuli. Their results clearly demonstrate that duplex perception does not depend upon the operation of a specialized phonetic subsystem. Furthermore, Pastore et al. also suggest the possibility that duplex perception may reflect a more general context effect in auditory perception. Thus, duplex perception may not represent dichotic fusion, but may instead demonstrate some form of contextual enhancement of chirp identification through the integration of information from several independent sources. Despite the criticisms raised by Repp (1984), we believe our earlier study was well motivated and that the results are directly relevant to the central claim that duplex perception reflects the operation of a perceptual subsystem that is uniquely specialized for phoneme perception.

REFERENCES

- BENTIN, S., & MANN, V. (1983). Using the duplex phenomenon to distinguish speech and nonspeech perception of second formant transitions. *Journal of the Acoustical Society of America*, 73, S52-S53. (Abstract)
- BROADBENT, D. E., & LADEFOGED, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, 29, 708-710.
- CARDEN, G., LEVITT, A., JUSCZYK, P. W., & WALLEY, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics*, 29, 26-36.
- CUTTING, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 83, 114-140.
- DARWIN, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset time. *Quarterly Journal of Experimental Psychology*, 33A, 185-207.
- DEUTSCH, D. (1980). The octave illusion and the what-where connection. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.
- DEUTSCH, D., & ROLL, P. L. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 23-29.
- DUNCAN, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology*, 12, 75-96.
- EIMAS, P. D., & CORBIT, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- EIMAS, P. D., TARTTER, V. C., & MILLER, J. L. (1981). Dependency relations during the processing of speech. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. New York: Krieger.
- GRUNKE, M. E., & PISONI, D. B. (1982). Some experiments on perceptual learning of mirror-image acoustic patterns. *Perception & Psychophysics*, 31, 210-218.
- HALWES, T. G. (1969). *Effects of dichotic fusion on the perception of speech*. Unpublished doctoral dissertation, University of Minnesota.
- INTONS-PETERSON, M. J. (1983). Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception and Performance*, 9, 394-412.
- JUSCZYK, P. W., SMITH, L. B., & MURPHY, C. (1981). The perceptual classification of speech. *Perception & Psychophysics*, 30, 10-23.
- KEWLEY-PORT, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, 72, 379-389.
- KEYSER, S. J., STEVENS, K. N., & KAWASAKI, H. (1983). *Phonetic and phonological investigation of redundant features*. Manuscript in preparation.
- LIBERMAN, A. M. (1970). The grammars of speech and language. *Cognitive Psychology*, 1, 301-323.
- LIBERMAN, A. M. (1979). Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen.
- LIBERMAN, A. M. (1982). On finding that speech is special. *American Psychologist*, 37, 148-167.
- 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., HARRIS, K. S., KINNEY, J. A., & LANE, H. L. (1961). The discrimination of relative onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, 61, 379-388.
- LIBERMAN, A. M., ISENBERG, D., & RAKERD, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, 30, 133-143.
- MANN, V. A., & LIBERMAN, A. M. (1984). Some differences between phonetic and auditory modes of perception. *Cognition*, 14, 211-235.
- MASSARO, D. W., & ODEN, G. C. (1980). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3). New York: Academic Press.

- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 2, 131-157.
- NUSBAUM, H. C. (1981). *Capacity limitations in phoneme perception*. Unpublished doctoral dissertation, SUNY at Buffalo.
- NUSBAUM, H. C., SCHWAB, E. C., & SAWUSCH, J. R. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics*, 33, 323-332.
- ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85, 172-191.
- PASTORE, R. E., SCHMUCKLER, M. A., ROSENBLUM, L., & SZCZESIUŁ, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics*, 33, 469-474.
- PISONI, D. B. (1975). 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, NJ: Erlbaum.
- PISONI, D. B. (1979). Some remarks on the perception of speech and nonspeech signals. *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen.
- PISONI, D. B., ASLIN, R. N., PEREY, A. J., & HENESSY, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.
- RAND, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, 55, 678-680.
- REMEZ, R. E., RUBIN, P. E., & PISONI, D. B. (1983). Coding of the speech spectrum in three time-varying sinusoids. In C. W. Parkins & S. W. Anderson (Eds.), *Cochlear prostheses: An international symposium*. New York: New York Academy of Sciences.
- 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. (1976). Identification of dichotic fusions. *Journal of the Acoustical Society of America*, 60, 456-469.
- REPP, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92, 81-110.
- REPP, B. H. (1984). Against a role of "chirp" identification in duplex perception. *Perception & Psychophysics*, 35, 89-93.
- REPP, B. H., MILBURN, C., & ASHKENAS, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics*, 33, 333-337.
- SAWUSCH, J. R., & PISONI, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, 2, 181-194.
- SCHWAB, E. C. (1981). *Auditory and phonetic processing of tone analogs of speech*. Unpublished doctoral dissertation, SUNY at Buffalo.
- STEIGER, H., & BREGMAN, A. S. (1982). Competition among auditory streaming, dichotic fusion, and diotic fusion. *Perception & Psychophysics*, 32, 153-162.
- STEVENS, K. N., & BLUMSTEIN, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, 1358-1368.
- STUDDERT-KENNEDY, M. (1983). On the dissociation of auditory and phonetic structure. In R. Carlson & B. Granstrom (Eds.), *The representation of speech in the peripheral auditory system*. Amsterdam: Elsevier.
- SUZUKI, H. (1970). *Mutually complementary effect of rate and amount of formant transition in distinguishing vowel, semi-vowel, and stop consonant* (Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics, N. 96). Boston: M.I.T.
- TREISMAN, A., & GELADE, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- TREISMAN, A., & SCHMIDT, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.

NOTES

1. Cutting (1976) has suggested that in duplex perception the brief duration of the transition relative to the base causes the chirp to stand out as a figure against the background of the syllable. Although this is certainly an interesting possibility, it is extremely speculative without any empirical support. It is quite difficult to think of another case in which one stimulus element serves as the entire figure and, at the same time, combines with a second stimulus to define the perceptual identity of the background. Thus, while Cutting's account is possible, it does seem somewhat implausible.

2. It is important to note that there is a fairly long history of research comparing perception of speech and nonspeech analogs that dates back to a study by Liberman, Harris, Kinney, and Lane (1961). However, in all this research, few if any studies attempted to obtain identification data for the nonspeech analogs. The vast majority of these experiments instead compared discrimination performance for speech and nonspeech, and on the basis of different patterns of discrimination for these stimuli, the conclusion has generally been reached that perception of speech and nonspeech is mediated by different "modes" of processing (e.g., Mattingly et al., 1971). However, when identification data have been obtained for nonspeech analogs, the labeling functions have been remarkably similar to the functions obtained for identification of the corresponding speech stimuli (e.g., see Pisoni, 1979).

3. The conclusion that subjects could not access the phonetic information encoded in the F3 chirps, based on the inability of subjects to label these chirps phonetically (Repp et al., 1983) is even more tenuous since these subjects received no training or feedback about their performance. It is entirely possible that feedback would have allowed the subjects to focus on the relevant attributes of the chirp for phonetic labeling. Indeed, with regard to an analogous question, Pisoni, Aslin, Perey, and Henessy (1982) demonstrated that appropriate training procedures and feedback can promote the acquisition of new phonetic categories by focusing the listener's attention on the relevant attributes of the contrast. Similarly, the conclusion that phonetic information cannot be directly extracted from the F3 chirp should be drawn only if subjects could not be trained to label the F3 chirps phonetically. In fact, it is entirely possible that in the duplex condition reported by Repp et al., the base, by temporarily directing phonetic processing to phonetically relevant attributes of the F3 chirp, could have served a function similar to feedback.

4. Massaro and Oden (1980) have proposed a model of speech perception that integrates information extracted from several acoustic cues to produce phonetic decisions. According to this model, phonetic information is independently derived from different sources and is then integrated according to the rules of fuzzy logic—not by dichotic fusion. This model also predicts that the effects of context (e.g., the base) on phonetic perception should be greatest when the relevant phonetic cue is ambiguous (e.g., the F3 chirp). Thus, this model can easily account for the facilitation of phonetic labeling by the base in duplex perception, without any assumption of dichotic fusion. Furthermore, since this model postulates independent extraction of phonetic information from acoustic cues, this model is entirely consistent with the chirp-identification hypothesis.

(Manuscript received September 12, 1983;
revision accepted for publication September 29, 1983.)