Perception & Psychophysics
1984, 35 (1), 89-93

Against a role of "chirp" identification in duplex perception

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

Duplex perception occurs when a single formant transition (or a pair of such transitions) of a synthetic syllable is isolated and presented to one ear while the remainder of the syllable (the "base") is presented to the opposite ear (Rand, 1974). Listeners report hearing a nonspeech "chirp" in the ear receiving the transition and, at the same time, a syllable in the other ear; the perceived identity of the syllable-initial consonant is determined by the contralateral formant transition. Previous accounts of this phenomenon have attributed the speech percept to dichotic integration or fusion of the transition with the base (e.g., Cutting, 1976; Liberman, Isenberg, & Rakerd, 1981). The nonspeech "chirp" percept was thought to reveal the simultaneous operation of distinct phonetic and auditory modes of perception (Liberman et al., 1981; Repp, 1982).

In a recent article, Nusbaum, Schwab, and Sawusch (NSS; 1983) proposed a new explanation. According to their "chirp-identification hypothesis," the speech percept does not derive from fusion, but from phonetic identification of the chirp *without reference to the base.* NSS also reported two experiments whose results seem consistent with their hypothesis. Although counterevidence was published simultaneously be Repp, Milburn, and Ashkenas (1983), it was not accepted as such by NSS (see their Footnote 3). The purpose of this note is to examine the arguments and data presented by NSS and to expose their weaknesses. The conclusion will be that the chirp-identification hypothesis is not a viable explanation of duplex speech perception and should be laid to rest.

Motivation for the Chirp-Identification Hypothesis

From a brief review of some earlier research, NSS conclude that "taken together, the available evidence favors the dichotic integration explanation of duplex perception" (pp. 324-325). Nevertheless, to prepare the ground for their chirp-identification hypothesis, NSS cite two findings that they consider to be at variance with the dichotic integration view.

Preparation of these comments was supported by NICHD Grant HD01994 to Haskins Laboratories. I am grateful to Alvin Liberman, Virginia Mann, Richard Pastore, Sigfrid Soli, and Michael Studdert-Kennedy for helpful comments on an earlier draft. The author's mailing address is: Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695.

One finding is Rand's (1974) observation that attenuation of second- and third-formant (F2 and F3) transitions in an intact syllable is more detrimental to phonetic perception than attenuation of the same transitions when they are removed from the base and presented to the opposite ear. NSS conclude that "this result demonstrates that the transitions are processed differently in an intact syllable and on the speech side of the duplex percept" (p. 325). They neglect the fact that Rand's (1974) and many subsequent split-formant studies (e.g., Danaher & Pickett, 1975; Hannley & Dorman, 1983; Nearey & Levitt, 1974; Perl & Haggard, 1974) were undertaken to investigate the effects of "upward spread of masking" due to the first formant (F1). Release from this form of masking consequent upon dichotic separation of formants is well documented. Within the framework of the dichotic integration hypothesis, then, there has been a widely accepted *psychoacoustic* explanation of the perceptual differences between intact and fused syllables, which does not imply that they are "processed differently."[1]

The second finding NSS cite as being incompatible with the dichotic integration hypothesis is Cutting's (1976) result that large differences in fundamental frequency do not substantially alter duplex perception. NSS argue that different fundamental frequencies signify different articulatory sources, and that the "phonetic processor" should not be able to integrate stimuli that appear to come from different sources. Several counterarguments may be offered, however: (1) The dynamic articulatory information conveyed by the time-varying properties of the chirp is likely to be much more important than that conveyed by fundamental frequency. (2) The chirp is not sufficiently speechlike to suggest any specific articulatory origin by itself. (3) Other forms of dichotic fusion are similarly unaffected by differences in fundamental frequency (Cutting, 1976; Repp, 1976a; Tartter & Blumstein, 1981).

Thus, contrary to NSS's arguments, there do not appear to be any serious problems for the dichotic integration explanation of duplex perception. The possibility remains that the chirp identification hypothesis might account equally well for the data in the literature. That it does not, however, is immediately evident from findings that NSS themselves cite as support for the dichotic integration hypothesis: How, for example, can the chirp-identification hypothesis account for the fact that duplex speech identification deteriorates with increasing temporal asynchrony of chirp and base (Cutting, 1976)? Or for the finding that, with selective attention to the speech side of the duplex percept, the chirp receives a different perceptual interpretation depending on the base it

is paired with (Liberman et al., 1981)? If there is no integration of chirp and base, it should not matter what the base is and when it occurs. NSS simply bypass these difficulties, which are painfully obvious.

The chirp-identification hypothesis rests on three assumptions. The first one is reasonable: "With the appropriate instructions, subjects might at least be able to 'guess' from which consonant or place of articulation a chirp was derived" (p. 325). The second assumption, however, is bizarre: "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" (p. 325). The base by itself sounds like a perfectly acceptable syllable (at least when the stimuli are derived from stop-consonant-vowel syllables), and if listeners could avoid fusing it with the chirp, they would surely respond to it the same way in which they respond to it in isolation. Indeed, NSS's own data show that, when the base is presented repeatedly in isolation, subjects are not reluctant at all to give the same response over and over. The third assumption is that the speech-like character of the base leads listeners to "identify the phonetic response with the base instead of with the transition" (p. 324). However, an inability to attribute the response to its correct stimulus would be expected only in the case of fusion. Moreover, if there is no dichotic integration, as NSS maintain, listeners should be able to attend to the base and hear it the way it sounds in isolation. In other words, the chirp and the base should be perceived as separate and unrelated stimuli, which they most decidedly are not (e.g., Liberman et al., 1981; Repp et al., 1983).

In summary, it is evident that the chirp-identification hypothesis is not only inconsistent with most data in the literature, but also rests on extremely implausible assumptions.

The Nusbaum et al. (1983) Data

NSS's Experiment 1 confirmed the crucial prediction that isolated chirps can be identified consistently as phonetic segments. The stimuli were the synthetic two-formant syllables [ba] and [ga], which are distinguished by a rising versus falling F2 transition. Repp et al. (1983) have pointed out that rising and falling F2 chirps bear an auditory resemblance to the glides [w] and [j]. Thus, subjects may have arrived at their (surprisingly consistent) responses by perceiving the chirps not as [b]-like or [g]-like but as [w]-like or [j]-like, and by subsequently choosing the response category that most resembled the quasi-phonetic glide percept. Such a relatively straightforward association may not exist, however, for stimuli used by others in earlier duplex perception experiments. Perhaps unwittingly, NSS chose stimuli that were uniquely suited to chirp identification.

Even though the isolated chirps could be associated with phonetic labels, it by no means follows that the subjects of NSS also relied on chirp identification in the duplex condition of Experiment 1. The relative similarity of the overall response proportions for isolated chirps and duplex stimuli (shown in Figure 3 of NSS) is very weak evidence indeed; it not only amounts to accepting the null hypothesis but also merely reflects similar response consistency—not necessarily similar response strategies—in the two experimental conditions. In fact, it is not unlikely that whatever speech-like attributes chirps may possess in isolation (e.g., [w]-like, [j]-like), they lose them in the duplex situation, due to competition from the fused speech percept. It is significant, in this connection, that NSS never asked their subjects to *identify the chirps in the duplex condition* while ignoring the bases (or, perhaps, some irrelevant syllables substituted for the bases). Without any demonstration that subjects actually can identify chirps phonetically in the presence of distracting contralateral speech stimuli, the results of Experiment 1 are inconclusive.

Experiment 2 was conducted to determine what NSS call the "labeling characteristics of the perceptual process (or processes)" (p. 328) used in the duplex paradigm. A six-member acoustic continuum from [ba] to [ga] was constructed by varying the onset frequency of the F2 transition in the presence of a constant F3 (with a rising transition, to inhibit [da] percepts). These stimuli were presented as full syllables, in a duplex condition, and in an isolated-chirp condition in which the isolated chirps included both the variable F2 and (for no apparent reason) the fixed F3 transition.

According to NSS, the dichotic integration hypothesis predicts that, "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" (p. 328). This prediction ignores once again the potential influence on the category boundary of release from masking due to F1 (as well as other possible psychoacoustic factors) in split-formant presentation (cf. Rand, 1974). While the direction of that influence is difficult to predict, there is no strong basis for expecting identical category boundaries in the two conditions. NSS further predict that, "since the isolated transitions must be processed differently from normal speech . . . , the category boundary for isolated transitions should be different from the duplex and intact boundaries" (p. 328). This is simply a non sequitur. The boundaries on entirely unrelated continua may coincide, particularly when they fall near the center of the stimulus range. Unless an experiment is designed to permit the prediction of specific boundary locations (see Bailey, Summerfield,

& Dorman, 1977), there is simply no logical connection between category boundaries and "manner" or mode of processing.

Although NSS do not state the predictions of the chirp identification hypothesis in detail, they apparently expected that the boundaries for isolated chirps and duplex stimuli would be the same, since both were thought to involve chirp identification, and different from the boundary for intact syllables because of the purported difference in "manner of processing." The results of Experiment 2 fit these predictions and thus were taken by NSS to support the chirp-identification hypothesis. It should be clear from the foregoing discussion, however, that the results are just as compatible with the dichotic integration hypothesis, and that the experiment is logically flawed.

In their General Discussion, NSS make a surprising (and confusing) turnabout by considering the possibility of dichotic fusion without abandoning the chirp-identification hypothesis which, of course, postulates the absence of fusion. They suggest, however, that "this dichotic fusion might not occur *prior* to phonetic labeling. Rather fusion should [sic!] occur *after* the phonetic features have been separately identified in the two ears" (p. 331). However, there is little evidence in favor of this new hypothesis. 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, for example [bga] or [bja]—which never happens in duplex presentation. A weakened version of the hypothesis, which does not permit such double-consonant percepts, would be indistinguishable from the dichotic integration view.

NSS also suggest that duplex perception experiments should include an isolated-chirp control condition, to be able "to determine how much more information is contributed by hearing the acoustic attribute in the appropriate syllabic context" (p. 331). If this methodological recommendation were all that NSS wished to convey, there would be little to disagree with. Clearly, despite the implausibility of the chirp-identification hypothesis, there might be some value in demonstrating that chirp identification can *not* account for the results of a particular study. The experiments of NSS could then be accepted as carefully contrived situations in which it seemed *as if* chirp identification had occurred in duplex perception. The problem with NSS's account, of course, is their insistence that chirp identification actually does occur. The correct conclusion should have been that there was no support for this hypothesis.

The Repp et al. (1983) Data

The data of Repp et al. (1983) were collected for the explicit purpose of refuting the chirp-identification hypothesis, as described in an early version of the NSS paper (Nusbaum, Schwab, & Sawusch, 1981). In Experiment 1, stimuli from a [da]-[ga] continuum varying in the F3 transition were used in a design similar to that of Experiment 2 of NSS. All subjects but one were unable to label the isolated F3 transitions consistently, and that one subject consistently reversed the category assignment. All subjects, however, labeled the syllables accurately in the duplex condition. Thus, this study demonstrated that phonetic identifiability of isolated chirps is not a necessary condition for duplex speech perception. In Experiment 2 of Repp et al., an AXB similarity judgment task was employed to facilitate selective attention to the ear receiving the base. Perception continued to be strongly influenced by the unattended contralateral chirp. This study disconfirmed a prediction that follows directly from the chirp-identification hypothesis, namely that subjects should be able to "recover" the base by selective attention to the ear receiving it.

In a footnote added in proof (Footnote 3, p. 332), NSS comment on Experiment 1 of Repp et al. (1983). Five points are made: (1) Instead of fusion of the chirp with the base, "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." Note that this is yet another hypothesis, different from the chirp-identification hypothesis, which postulates that duplex speech identification proceeds without reference to the base. In fact, the only way in which this unannounced "facilitation hypothesis" seems to differ from the dichotic fusion hypothesis is that it predicts that selective attention to the base should be possible. However, Experiment 2 of Repp et al. (1983) (on which NSS do not comment) refutes that prediction. (2) NSS point out that the results of Repp et al. do not prove "that it is *impossible* for subjects to extract phonetic information from these isolated chirps." This is correct but irrelevant, because the point of the demonstration was that poorly identified chirps nevertheless lead to accurate consonant identification when paired with a base. (3) "Repp et al. did not establish the level at which this fusion occurs." Indeed, this was not the purpose of their study. (4) "According to the chirp-identification hypothesis, if fusion does occur, it should take place after some phonetic processing of the chirp." How can a prediction about fusion be derived from a hypothesis that explicitly postulates the nonoccurrence of fusion? (5) 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." However, parsimony demands that a common account be provided for all duplex perception and split-formant experiments, and dichotic fusion is a highly satisfactory general explanation.

Moreover, there is no evidence at all that the chirps in earlier duplex studies could be identified in isolation, since this was not tested and different types of stimuli were used. In summary, these comments of NSS do nothing to weaken the results of Repp et al., which clearly disconfirm the chirp-identification hypothesis.[2]

Conclusion

To be sure, a lot more is to be learned about dichotic fusion and auditory segregation in speech stimuli. While fusion clearly takes place in duplex perception, we do not know at what level in the auditory system it occurs, what kinds of neural mechanisms it involves, and whether or not it is specific to phonetic perception. These interesting questions should be pursued without further distraction.

REFERENCES

AINSWORTH, W. A. (1978). Perception of speech sounds with alternate formants presented to opposite ears. *Journal of the Acoustical Society of America*, **63**, 1528-1534.

BAILEY, P. J., SUMMERFIELD, Q., & DORMAN, M. (1977). *On the identification of sine-wave analogues of certain speech sounds* (Haskins Laboratories Status Report on Speech Research, SR-51/52, pp. 1-25). New Haven, CT: Haskins Laboratories.

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**, S52-53. (Abstract)

BROADBENT, D. E. (1955). A note on binaural fusion. *Quarterly Journal of Experimental Psychology*, **7**, 46-47.

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.

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

DANAHER, E. M., & PICKETT, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *Journal of Speech and Hearing Research*, **18**, 261-271.

DARWIN, C. J., HOWELL, P., & BRADY, S. A. (1978). Laterality and localization: A "right ear advantage" for speech heard on the left. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.

HANNLEY, M., & DORMAN, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, **74**, 40-51.

ISENBERG, D. S., & LIBERMAN, A. M. (1978). Speech and nonspeech percepts from the same sound. *Journal of the Acoustical Society of America*, **64** (Suppl. No. 1), S20. (Abstract)

JUSCZYK, P. W., SMITH, L. B., & MURPHY, C. (1981). The perceptual classification of speech. *Perception & Psychophysics*, **30**, 10-23.

LIBERMAN, A. M. (1974). The specialization of the language hemisphere. In F. O. Schmidt & F. G. Worden (Eds.), *The neurosciences: Third study program*. Cambridge: M.I.T. Press.

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. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, **14**, 211-235.

NEAREY, T. M., & LEVITT, A. G. (1974). *Evidence for spectral fusion in dichotic release from upward spread of masking* (Haskins Laboratories Status Report on Speech Research, SR-39/40, pp. 81-89). New Haven, CT: Haskins Laboratories.

NUSBAUM, H. C., SCHWAB, E. C., & SAWUSCH, J. R. (1981, April). *Duplex perception: Dichotic integration or "chirp" identification?* Paper presented at the 52nd meeting of the Eastern Psychological Association, New York.

NUSBAUM, H. C., SCHWAB, E. C., & SAWUSCH, J. R. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics*, **33**, 323-332.

NYE, P. W., NEAREY, T. M., & RAND, T. C. (1974). *Dichotic release from masking: Further results from studies with synthetic speech stimuli* (Haskins Laboratories Status Report on Speech Research, SR-37/38, pp. 123-137). New Haven, CT: Haskins Laboratories.

PASTORE, R. E., SZCZESIUL, R., ROSENBLUM, L. D., & SCHMUCKLER, M. A. (1982). When is a [p] a [t], and when is it not? *Journal of the Acoustical Society of America*, **72** (Suppl. No. 1), S16. (Abstract)

PERL, N., & HAGGARD, M. (1974). *Masking versus hemispheric sharing of processing for speech sounds* (*Speech Perception*, 2:3, 47-54). Belfast: The Queen's University, Department of Psychology.

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

REPP, B. H. (1975). Dichotic masking of consonants by vowels. *Journal of the Acoustical Society of America*, **57**, 724-735.

REPP, B. H. (1976a). Effects of fundamental frequency contrast on identification and discrimination of dichotic CV syllables at various temporal delays. *Memory & Cognition*, **4**, 75-90.

REPP, B. H. (1976b). Dichotic "masking" of voice onset time. *Journal of the Acoustical Society of America*, **59**, 183-194.

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

SCHWAB, E. C. (1981). *Auditory and phonetic processing of tone analogs of speech*. Unpublished doctoral dissertation, SUNY at Buffalo.

STUDDERT-KENNEDY, M. (1981). The emergence of phonetic structure. *Cognition*, **10**, 301-306.

TARTTER, V. C., & BLUMSTEIN, S. E. (1981). The effects of pitch and spectral differences on phonetic fusion in dichotic listening. *Journal of Phonetics*, **9**, 251-259.

NOTES

1. NSS later dismiss the possibility of (upward spread of) masking effects on the grounds that "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)" (p. 330). This reflects a serious misunderstanding: By the same token, these proponents would presumably have to argue that the intelligibility of speech should remain unimpaired in the presence of loud noise! Obviously, distortions due to interactions in the peripheral auditory system must precede any phonetic processing. The point of the authors cited by NSS was that phonetic *classification* cannot be explained by general auditory processes; however, perceptual changes may well result from factors that affect the internal spectrotemporal representation of speech signals. NSS also cite an unpublished dissertation by Schwab (1981) as showing that auditory masking is absent when stimuli are perceived as speech. While Schwab's results are intriguing, they are not directly applicable to the duplex situation because they did not rest on a comparison of monaural and dichotic presentation conditions. To conclude from Schwab's findings that auditory masking cannot occur in speech stimuli would be absurd.

2. There are a variety of other observations that speak directly or indirectly against the chirp-identification hypothesis. To mention only one particularly damaging result, both Cutting (1976) and Rand (1974) have found that duplex speech perception is resistant to severe attenuation of the chirp; in fact, Bentin and Mann (1983) recently demonstrated that speech identification is still good when chirp detection and discrimination scores are at chance. For other relevant results, see Ainsworth (1978), Bentin and Mann (1983), Broadbent (1955, 1957), Darwin, Howell, and Brady (1978), Isenberg and Liberman (1978), Jusczyk, Smith, and Murphy (1981), Mann and Liberman (1983), Nye, Nearey, and Rand (1974), Pastore, Szczesiul, Rosenblum, and Schmuckler (1982), and Repp (1975, 1976b).

(Manuscript received February 28, 1983;
revision accepted for publication June 7, 1983.)