Perception & Psychophysics
1983, 33 (4), 333-337

Duplex perception: Confirmation of fusion

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

and

CHRISTINA MILBURN and JOHN ASHKENAS
Yale College, New Haven, Connecticut

Duplex perception—the simultaneous perception of a speech syllable and of a nonspeech "chirp"—occurs when a single formant transition and the remainder (the "base") of a synthetic syllable are presented to different ears. Two experiments were conducted to test whether the speech percept derives from the dichotic fusion of the transition with the base or from phonetic information extracted directly from the isolated transition. Experiment 1 showed that subjects were unable to assign speech labels to isolated transitions in a consistent manner, although the same transitions led to accurate identification when paired with the constant base in the other ear. Experiment 2 used an AXB paradigm to show that selective attention to the ear receiving the base does not prevent the contribution of the contralateral transition to the speech percept. Both experiments support the hypothesis that the speech percept in the duplex situation results from dichotic fusion at a relatively early stage in processing.

The phenomenon of duplex perception has been taken to support the existence of a specialized phonetic mode for perceiving speech (Liberman, 1979; Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, in press). Duplex perception occurs when a synthetic consonant-vowel syllable is split in a certain way and presented dichotically (Rand, 1974). If the initial formant transition which identifies the consonant is removed from the acoustic context of the rest of the syllable and played in isolation, listeners report hearing a nonspeech "chirp." When the rest of the syllable without the transition, the "base," is played in isolation, listeners report hearing a syllable, sometimes beginning with the same consonant as the whole syllable and sometimes not. If the chirp is now presented to one ear and the base to the other ear, with the two stimuli timed to coincide as they would in the whole syllable, listeners report a duplex percept. In the ear to which the chirp was presented, they report hearing a nonspeech sound—the chirp, more or less as it sounds when played in isolation. In the other ear, they report hearing speech which they correctly identify as the original syllable from which the two stimuli were derived.

The standard explanation given for this phenomenon is that the base and the chirp are fused to form the whole syllable which is heard in one ear, while the chirp alone is also heard separately in the other ear (Cutting, 1976; Liberman et al., 1981). According to this account, the chirp is heard simultaneously as part of the fused speech syllable and as nonspeech (as it sounds in isolation). The duplex phenomenon therefore supports the existence of two distinct modes for perceiving sound: one auditory, for nonspeech sounds, and the other phonetic, a mode of perception specialized for processing speech (Liberman et al., 1981; Mann & Liberman, in press). Both modes seem to be engaged simultaneously in the duplex situation.

The above account is based on listeners' introspections and has never been tested directly. There are alternative theoretical possibilities, however, which make such a test desirable. It has been suggested (Nusbaum, Schwab, & Sawusch, Note 1) that although the formant transition in isolation sounds like a nonspeech chirp, it may contain enough phonetic information for listeners to identify the consonant it cues. In the duplex situation, listeners may then identify the syllable correctly on the basis of the chirp alone, and, since the base in the other ear sounds like (perhaps ambiguous) speech, listeners merely attribute the speech percept to that ear. According to this hypothesis, no fusion of the chirp and base occurs, and the formant transition is perceived in exactly the same (simplex) way when it is presented with the base as when it is not.

Two easily testable predictions follow from this nonfusion hypothesis: (1) Isolated formant transitions should be identifiable as the consonants they are intended to cue, and (2) when listeners focus their attention on the ear in which the base occurs,

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. We thank Alvin Liberman, Virginia Mann, Richard Pastore, and David Pisoni for their advice and for comments on the manuscript. B. H. Repp's mailing address is: Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

they should be able to identify it the way it is identified in isolation, regardless of whether or not a chirp occurs in the other ear. We conducted two experiments to examine these issues.

EXPERIMENT 1

The hypothesis that subjects might be able to assign phonetic labels to isolated formant transitions is in apparent contradiction to claims in the literature that these stimuli are pure nonspeech sounds (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971). However, these claims may have been exaggerated. Investigators familiar with stimuli of this kind will have noted that, for example, isolated second-formant transitions derived from /ba/ and /ga/ sound vaguely like /wə/ and /yə/, respectively. Since these glides share place of articulation with the relevant stop categories, subjects may be able to associate the two manner classes and thereby arrive at consistent labeling responses. To make such an association is different from actually hearing /ba/ and /ga/ (which is what subjects report experiencing in the duplex condition). Nevertheless, a recent demonstration that subjects indeed can label isolated second-formant transitions in a consistent manner (Nusbaum et al., Note 1) raises the question of whether or not the speech percept in the duplex situation is similarly derived from the chirps alone.

Experiment 1 used synthetic stimuli which formed a /da/-/ga/ continuum and were distinguished only by the transition of the third formant. These transitions are in a much higher frequency range than the second-formant transitions employed by Nusbaum et al. (Note 1) and sound considerably less speechlike. Duplex perception has been obtained with similar stimuli by Mann and Liberman (in press). The present study attempted to replicate this finding and tested, in addition, whether subjects could label third-formant chirps consistently as /da/ or /ga/. The goal of the experiment was to demonstrate that duplex perception can be obtained with chirps that, by themselves, are not readily associated with phonetic categories.

Method

Subjects. A total of 12 subjects participated. Eight of them were student volunteers with little or no previous experience in speech-perception experiments. The other four were familiar with the purpose of the experiment and included two relatively experienced (B.R. and V.M.) and two relatively inexperienced listeners (C.M. and J.A.).

Stimuli. The stimuli were six three-formant synthetic syllables created on the Haskins Laboratories parallel resonance synthesizer and forming a /da/-/ga/ continuum. All syllables were 250 msec in duration and had linear 50-msec initial transitions in all three formants, followed by a 200-msec steady state. The first formant rose from 285 to 771 Hz, the second formant fell from 1770 to 1233 Hz, and the third formant, which alone distinguished the six syllables, started at a variable frequency and went to 2525 Hz. The onset frequencies of the third formant in the six stimuli were 2862, 2694, 2525, 2348, 2180, and 2018 Hz. The "chirps" consisted of the 50-msec transition of the third formant in isolation; the "base" consisted of a syllable without that distinctive transition, that is, with no energy in the third-formant region during the first 50 msec. Consequently, there were six different chirps but only one base.

Three tapes were recorded. On the first, the six chirps occurred in isolation. On the second, the six full syllables were recorded, with the base thrown in as a seventh stimulus. The third tape contained the six duplex syllables, with the chirp on one channel and the base on the other. On each tape, the stimuli were repeated 20 times in random sequence, with interstimulus intervals of 3 sec.

Procedure. The subjects listened in groups, over TDH-39 earphones, in a quiet room. The isolated chirps were presented first, to avoid any effects of experience. The subjects were told that they would hear chirp-like sounds but should do their best to label these sounds as "d" or "g," guessing if necessary. The chirps were presented monaurally to the right ear. Next, the full syllables and the base were presented monaurally to the left ear. The subjects were instructed to identify the consonant in these syllables as "d" or "g." This was followed by the duplex tape, with the base always in the left ear and the chirps in the right ear. The subjects were told to ignore the chirps and identify the syllables in the left ear. Finally, the eight inexperienced subjects listened to the isolated chirps for a second time, to determine whether exposure to the duplex condition had any beneficial effect on chirp identification.

Results and Discussion

A first inspection of the data revealed no differences between the results of the first and second chirp-identification tests for the naive subjects, so both were combined. Furthermore, there were no systematic differences between the results for naive and experienced or informed listeners, so their results also were pooled. The average results of all 12 subjects are displayed in Figure 1.

The results are very clear. First, in both the full-syllable and duplex conditions, the stimuli were

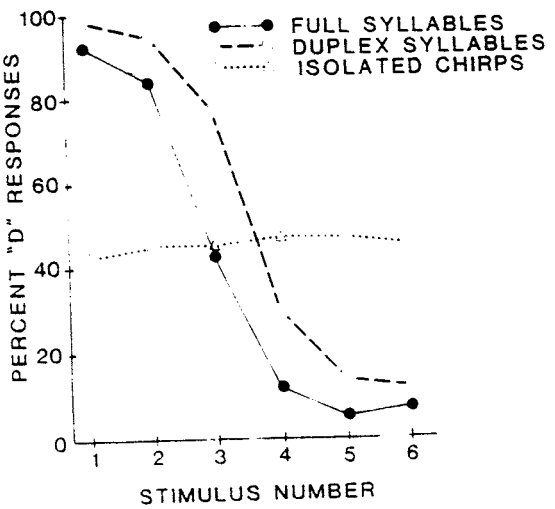

Figure 1. Labeling functions for three conditions of Experiment 1.

labeled quite consistently, whereas labeling of isolated chirps was totally random for the subject group as a whole. Second, there was a sizable difference between the full-syllable and duplex labeling functions; there were generally more "d" responses in the duplex condition [$F(1,11) = 11.7$, $p < .01$].

The poor labeling performance for isolated chirps was expected because these stimuli bore no resemblance to speech. While some of them sounded discriminably different, at least to some listeners, they could not be consistently associated with the two phonetic categories, "d" and "g." Individual subject data are presented in Table 1. There was only one listener (Subject 7) who labeled the stimuli in a consistent way: He called stimuli 1-3 "g" and stimuli 4-6 "d." Note that this assignment of categories is reversed with respect to the duplex condition. This subject presumably used the category labels to arbitrarily designate the psychoacoustic categories of rising and falling pitch. (Stimulus 3 had a level pitch.) Three other subjects (5, V.M., and B.R) showed a tendency to label chirps the "correct" way, but none of them was nearly as consistent as in the duplex condition. (The two experienced listeners, V.M. and B.R., probably could have achieved more consistent performance by adopting psychoacoustic categories, like Subject 7, but they deliberately refrained from doing so and, instead, tried hard to follow the instructions to hear the stimuli as "d" or "g.")

Since *all* subjects gave orderly labeling responses

Table 1
Individual Subject Data for Isolated Chirp Identification: Number of "d" Responses (of 20)

Subject	Run	Chirp Stimulus Number 1	2	3	4	5	6
1	1	10	10	4	6	15	11
	2	15	15	9	12	10	8
2	1	12	9	14	9	13	9
	2	11	9	10	11	13	9
3	1	12	13	13	11	15	13
	2	17	18	19	17	18	19
4	1	5	13	13	12	11	15
	2	6	7	12	13	10	10
5	1	13	9	11	5	7	4
	2	18	14	6	3	7	1
6	1	0	1	0	6	7	6
	2	1	2	2	3	3	1
7	1	1	1	3	20	20	19
	2	0	0	0	20	20	20
8	1	9	8	11	7	8	10
	2	5	6	3	6	4	4
C.M.		9	8	13	10	7	12
J.A.		1	0	1	1	1	0
V.M.		15	16	16	12	8	7
B.R.		10	17	14	10	5	9

in the duplex condition, these data strongly suggest that the speech percepts in the duplex situation were due to dichotic fusion and not to phonetic labeling of the chirps. By implication, dichotic fusion may be assumed to occur also in duplex situations involving somewhat more speechlike (e.g., second-formant) chirps.

The finding of a difference in labeling functions between the full-syllable and duplex conditions is in need of explanation. One possibility is that, in the duplex condition, fusion was not complete, so that the phonetic category associated with the base exerted a bias on identification. The base on the full-syllable tape was identified as "d" on 87.1% of the trials; that is, it sounded essentially like /da/. The shift of the duplex labeling function in favor of "d" responses is consistent with the hypothesis just proposed. However, other data (Mann, Note 2; Nusbaum et al., Note 1) do not seem to follow this pattern. An alternative possibility is that the duplex condition favored the category associated with a falling critical formant transition over the category associated with a rising transition. It has long been known that the first formant exerts an "upward spread of masking" effect on the perception of the higher formants; indeed, this effect motivated the original research using duplex and split-formant stimuli (Nye, Nearey, & Rand, 1974; Rand, 1974). This "masking" may be partially due to an incompatibility in the direction of formant transitions: Since the first formant in initial stop consonants is always rising in frequency, the perception of simultaneous falling transitions in the higher formants may be selectively impaired. Dichotic presentation may reduce this incompatibility effect, and this may explain the increase in responses corresponding to the category cued by falling formant transitions. This explanation seems to be in agreement with the data reported by Nusbaum et al. (Note 1) but may not be universally valid (Mann, Note 2). A final possibility is that the abrupt offset of the isolated formant in one ear and the abrupt onset of the corresponding formant in the base in the other ear introduce spectral effects that persist in the (apparently smooth) fused percept and have some influence on phonetic perception, although the direction of that influence is difficult to predict.

EXPERIMENT 2

Experiment 2 examined the hypothesis that subjects, when selectively attending to the ear containing the base, might actually perceive the syllable represented by the base and not the one thought to result from the fusion of the base with a contralateral chirp. A failure to find support for this hypothesis would argue not only against the pro-

posal (Nusbaum et al., Note 1) that the speech percept in the duplex situation derives exclusively from the chirp, but also against a weaker version of that view which is compatible with the results of Experiment 1 and according to which the base provides a relevant background against which the chirp yields its phonetic information, but without fusion occurring. Despite instructions to ignore the chirp, the labeling task of Experiment 1 may not have provided a sufficient incentive for directing full attention to the ear containing the base. In the present study, an AXB forced-choice paradigm was used instead, which required subjects to make similarity judgments about stimuli in one ear only, with contralateral chirps occurring only from time to time. Subjects' inability to recover the base under these more stringent conditions would provide further support for early dichotic fusion as the cause of the reported speech percept.

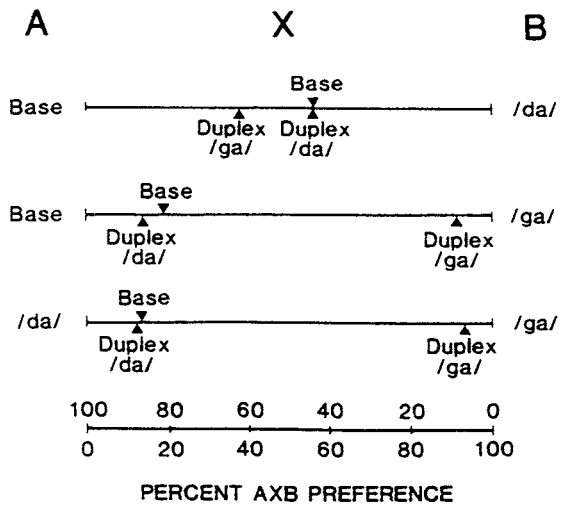

Figure 2. AXB similarity judgments (Experiment 2).

Method

Subjects. The same subjects as in Experiment 1 participated in this test, which was administered at the end of the same single session.

Stimuli. The stimuli were the two endpoints of the /da/-/ga/ continuum, their duplex versions, and the isolated base. These five stimuli were arranged into AXB triads in the following way: The A and B stimuli, which were always different from each other, were either the two full syllables or one of them and the base, in either order. The X stimuli inserted into these six possible frames were the two duplex syllables and the base. This resulted in 18 different triads, which were recorded five times in random order, with interstimulus intervals of 1 sec within triads and of 4 sec between triads. All stimuli were recorded on the left channel except for the chirps of the duplex syllables, which occurred on the right channel.

Procedure. The subjects were instructed to pay attention only to the left ear and to judge in each triad whether the middle stimulus sounded more similar to the first (response "1") or to the third stimulus (response "3"), guessing if necessary. Note that the A and B stimuli were always monaural, which forced attention to the ear receiving the base of the duplex X stimuli.

Results and Discussion

The majority of the stimulus triads were uninformative and merely provided the background for the critical triads. Since it was known from Experiment 1 that the base by itself sounded like /da/, it was to be expected that for a triad such as "full /da/, duplex /da/, base" subjects' judgments would be fairly random, for they would hear "/da/, /da/, /da/." The critical triads were those in which duplex /ga/ occurred between full /da/ and full /ga/, or between the base and full /ga/. Because the base of duplex /ga/ sounds like /da/, duplex /ga/ should be judged to be more similar to either full /da/ or to the base than to full /ga/ if fusion can be avoided. The fusion hypothesis, of course, predicts exactly the opposite.

There were no systematic differences between experienced and inexperienced subjects, although the former provided somewhat more consistent results. The results for all 12 subjects combined are displayed in Figure 2. The figure shows the percentages of trials on which each of the three X stimuli was judged to be more similar to either A or B. Each line shows one of the three A-B frames, combining the two possible orders. The results are unambiguous. When both A and B sounded like /da/ (line 1), the subjects responded randomly, although the duplex /ga/ was judged to be somewhat more similar to the base than to the full /da/. When one frame stimulus sounded like /da/ and the other like /ga/ (lines 2 and 3), the base and the duplex /da/ were judged to be more similar to /da/, whereas the critical duplex /ga/ was judged to be more similar to /ga/. Note in particular that, in the sequence "base, duplex /ga/, full /ga/," the attended ear received two identical stimuli (the base) followed by a different one; nevertheless, the subjects chose the second stimulus as being significantly more similar to the third than to the first, indicating that the perception of the second stimulus was significantly altered through fusion with the contralateral chirp.

CONCLUSION

The present results strongly support the hypothesis that chirp and base fuse at a relatively early stage in processing (see Cutting, 1976). This fusion seems to be obligatory and, unlike some higher level dichotic fusions (Sexton & Geffen, 1981), to be unaffected by selective-attention strategies. The present findings definitely refute the hypothesis that the phonetic percept in the duplex paradigm derives from the assignment of speech labels to the unfused

chirp. The standard interpretation of duplex perception, discussed and elaborated most recently by Liberman et al. (1981) and by Mann and Liberman (in press), therefore appears valid and provides a basis for further demonstrations of a dissociation between phonetic and auditory modes of perception (Mann, Note 2).

REFERENCE NOTES

1. Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. *Duplex perception: Dichotic integration or "chirp" identification?* Paper presented at the meeting of the Eastern Psychological Association, New York, April 1981.

2. Mann, V. A. Personal communication, 1982, and manuscript in preparation with S. Bentin.

REFERENCES

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

LIBERMAN, A. M. Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jørgensen, J. Rishel, & N. Thorsen (Eds.), *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen Press, 1979.

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.

MANN, V. A., & LIBERMAN, A. M. Some differences between phonetic and auditory modes of perception. *Cognition*, in press.

MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131-157.

NYE, P. W., NEAREY, T. M., & RAND, T. C. Dichotic release from masking: Further results from studies with synthetic speech stimuli. *Haskins Laboratories Status Report on Speech Research*, 1974, **SR-37/38**, 123-138.

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

SEXTON, M. A., & GEFFEN, G. Phonological fusion in dichotic monitoring. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, **7**, 422-429.

(Manuscript received July 6, 1982;
revision accepted for publication January 31, 1983.)