

Duplex perception of cues for stop consonants: Evidence for a phonetic mode

ALVIN M. LIBERMAN, DAVID ISENBERG, and BRAD RAKERD
Haskins Laboratories, New Haven, Connecticut 06510

When the (vocalic) formant transitions appropriate for the stops in a synthetic approximation to [spa] or [sta] are presented to one ear and the remainder of the acoustic pattern to the other, listeners report a duplex percept. One side of the duplexity is the same coherent syllable ([spa] or [sta]) that is perceived when the pattern is presented in its original, undivided form; the other is a nonspeech chirp that corresponds to what the transitions sound like in isolation. This phenomenon is here used to determine why, in the case of stops, silence is an important cue. The results show that the silence cue affects the formant transitions differently when, on the one side of the duplex percept, the transitions support the perception of stop consonants, and when, on the other, they are perceived as nonspeech chirps. This indicates that the effectiveness of the silence cue is owing to distinctively phonetic (as against generally auditory) processes.

A biological specialization for language could become manifest in several ways. In the domain of speech perception, it might appear as a distinctive mode (or subsystem), different from the auditory mode(s), and comprising just those processes that come into play when, on being presented with the sounds of speech, a listener perceives consonants and vowels. That such a mode—let us call it “phonetic”—may, in fact, exist is implied by a variety of circumstantial evidence. (For reviews, see Darwin, 1976; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978; Studdert-Kennedy, 1976; but see Kuhl & Miller, 1975; Miller, 1977.) Our aim here is to offer evidence of a more direct sort. To that end, we take advantage of a procedure, contrived by Rand (1974), that produces an odd, and possibly unique, phenomenon: two quite different percepts arise simultaneously when the acoustic constituents of a (synthetic) syllable are separated and presented dichotically. One of these percepts is a coherent syllable that sounds just like the original, having been formed, evidently, by a fusion of the dichotically presented constituents. The other percept is a nonspeech “chirp” that contrasts

strikingly with the syllable and corresponds closely to what one of the constituents sounds like in isolation. We take this duplex percept to be of interest, not only because it can demonstrate the simultaneous operation of phonetic and auditory modes, but, more particularly, because it can be used to determine why, in the perception of speech, certain aspects of the signal are effective cues. Here, we have used the duplexity to answer questions about silence as a cue for stop consonants. But we will defer those questions until we have seen more clearly just what duplex perception is and what it might represent.

An example of duplex perception, appropriate for purposes of explication, is found in a recent study of the perceived contrast between [ra] and [la] (Isenberg & Liberman, 1978; Liberman, 1979). The procedure for obtaining the phenomenon was like that of Rand, but duplexity was, perhaps, more thoroughly explored and more convincingly demonstrated. First, the syllables [ra] and [la], shown schematically in the top half of Figure 1, were synthesized so as to make the perceived distinction depend entirely on the transition of the third formant. Then, as shown in the bottom half of the figure, these patterns were divided into two constituents. One constituent, labeled “base” and shown at the left, included all aspects of the pattern that were identical in the two syllables. When presented by itself, this common core was perceived as a syllable, almost always as [ra]. The other constituent, shown to the right, was one or the other of the third-formant transitions that, in the undivided syllable, critically distinguished [ra] from [la]. In isolation, these transitions were perceived variously, but in no case did they sound the same as when, in the undivided patterns, they were essential to the difference between the syllables; by most listeners, indeed, they were thought to be not-very-speechlike, but dis-

This research was supported by the following grants from the National Institutes of Health: HD01994 (National Institute of Child Health and Human Development); RR05596 (Biomedical Research Support, Division of Research Resources); NS05493 and NS07040 (National Institute of Neurological and Communicative Disorders and Stroke). We thank Terry Halwes for his help in the preparation of the stimuli, and we thank Michael Studdert-Kennedy, Bruno Repp, and Dennis Klatt for their many useful comments. Alvin M. Liberman is also affiliated with the University of Connecticut and Yale University, Brad Rakerd with the University of Connecticut. David Isenberg is now at the Speech Applications Information Laboratories, Woods Hole, Massachusetts. Address reprint requests to Alvin M. Liberman, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

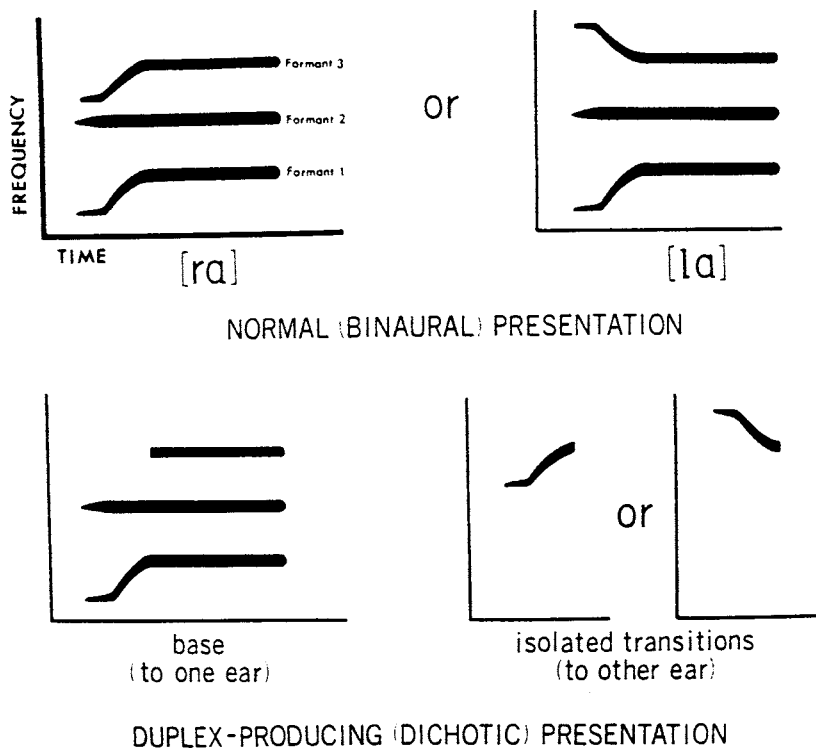


Figure 1. Schematic representations of patterns appropriate for duplex perception of [ra] and [la].

criminally different, "chirps." The last, and critical, step was to put the base into one ear and one or the other of the isolated transitions into the other, being careful, of course, to make the temporal relation between the dichotically presented constituents the same as it had been in the undivided patterns.

The result was a duplex percept. One component was a syllable that listeners "correctly" perceived as [ra] or [la] according to the nature of the third-formant transition. The other component, perceived at the same time as the syllable, was a not-very-speechlike chirp. This percept corresponded to the one that had been produced by the third-formant transition in isolation. The two percepts were not only phenomenally distinct but also dissociable, as could be inferred from the further finding that listeners were able to report changes in the loudness of the syllable or the chirp according as the intensity of the base or the third-formant transition was varied.

What interests us here is not so much that the dichotically presented constituents were fused in perception, but rather that one of them was also perceived as if it had not fused. This is the more interesting because the constituent that both fused and did not fuse is the one of the two that, in isolation, did not sound like speech. Thus, given the third-formant

transition appropriate for [l] but perceived in isolation as a chirp, and given also the base that was perceived by itself as [ra], listeners did not perceive only the result of fusion: the syllable [la]. Had they perceived only [la], we should have supposed that they were experiencing an effect no different from the one that is obtained in ordinary dichotic fusion, as, for example, when all of the first and second formant is put into one ear and all of the third formant into the other (Broadbent, 1955; Broadbent & Ladefoged, 1957; Halwes, 1969; Rand, 1974; Turek, Dorman, Franks, & Summerfield, 1980; Darwin, Howell, & Brady, Note 1). Neither did the listeners perceive all possibilities: the "fused" [la], the "unfused" [ra], and the "unfused" chirp. Had they so perceived the dichotically presented stimuli, we might have supposed that there were, somehow, two consciously available stages (fused and unfused) of auditory processing, or, alternatively, an auditory stage (the two unfused percepts) followed by a phonetic stage (the fused percept). What the listeners did, in fact, perceive was the "fused" [la] and the "unfused" chirp. Thus, perception was not, as it might have been, either unitary or triplex. Quite remarkably, it was duplex, which is to say that it represented two ways of processing the stimuli: as speech and as non-

speech. More to the point, the two ways of perceiving, and the duplex percept that resulted, turned on the [l] transition. On the "chirp" side of the percept, that transition was perceived in a way we will call "auditory," because the conscious impression was of sound but not speech; moreover, it had those characteristics that psychoacoustic considerations would have led us to expect. On the other side, the same transition was perceived as having the singularly different quality, hard to describe in auditory terms, that distinguishes [la] from [ra]. We take that different percept to result from correspondingly different processes; in our view, the mode which those processes serve deserves the name "phonetic," because its percepts have just those characteristics we can be aware of when we listen to consonants and vowels.

In the experiments to be reported in this paper, we have extended duplex perception to the case of fricative-stop-vowel syllables ([spa], [sta]) in which perception of the stop depends critically on an interval of silence positioned between the noise of the fricative and the (appropriate) vocalic transitions (Bailey & Summerfield, 1980; Dorman, Raphael, & Liberman, 1979; Bastian, Note 2). More particularly, we have exploited the auditory and phonetic sides of the duplex percept to determine why, in the perception of these stops, silence should be an important cue.

To understand the importance of silence as a cue for stop consonants, we might invoke either of two quite different accounts, one auditory and the other phonetic. An auditory account could have it, for example, that the silence enables the transition cues to evade (forward) auditory masking by the fricative noise. Failing that, it could fall back on the assumption that the silence and the transitions engage in an auditory interaction of some unspecified sort, from which emerges the percept we know as a stop consonant. But the particular nature of an auditory account need not concern us. For our purposes, the important characteristic of such a view is that the effect of silence would occur for all sounds that have certain attributes, and it would not matter that they were, or were not, perceived as speech.

Alternatively, the stop-producing effect of silence could reflect the operation of a distinctive phonetic process, specialized to perceive the silence in a linguistically appropriate way. Such a process would treat the presence or absence of silence as phonetically relevant information, revealing, for example, that the talker's vocal tract had closed, as it must to produce the stop consonant in [spa] or [sta], or that it had not, as it does not when the talker articulates the "stopless" [sa]. On this interpretation, silence would have its effect only when sounds were heard as speech.

The phonetic interpretation of the role of silence has found support in several recent experiments (Fitch,

Halwes, Erickson, & Liberman, 1980; Repp, Liberman, Eccardt, & Pesetsky, 1978), including one (Dorman et al., 1979) that is most directly relevant to the current study. There, the investigators found that a preceding silent interval was, indeed, necessary if formant transitions were to function as cues for stop consonants, but that silence had little or no effect on the identification of those same transitions when, having been isolated from the rest of the pattern, they were perceived as nonspeech chirps. This result is, of course, open to the reservation that the transition cues were in different acoustic contexts in the two conditions of the experiment. The experiments reported here were designed to speak to that reservation. To that end, they first test for duplex perception with stimuli that comprise the appropriate cues of silence and transitions. That done, the duplexity is used to determine whether the silence affects the speech percept and the corresponding nonspeech differently, when, as in this case, the acoustic patterns that produce both percepts are exactly the same.

EXPERIMENT 1

The primary purpose of this experiment is to test for duplex perception in the case of the cues, silence and formant transitions, that produce the perceived distinctions among [spa], [sta], and [sa]. Secondly, the experiment should provide evidence relevant to the hypothesis that the effectiveness of the silence cue is attributable to distinctively phonetic processes.

Given the transitions in one ear and the remainder of the pattern (the "base") in the other, we ask whether listeners perceive syllables appropriate to the fusion of the two while also perceiving a nonspeech chirp appropriate to the transition cues alone. In other words, do listeners achieve duplex perception? On the speech side of the duplex percept, the test will be to determine if the listeners "correctly" perceive [spa] and [sta] (depending on the nature of the transitions) when silence is present in the base, and [sa] (regardless of the transitions) when silence is not present. On the nonspeech side, the test must take into account that, unlike syllables, the chirps are not readily categorized; they are, however, discriminable. Accordingly, the test on the nonspeech side will be to see if listeners do, in fact, correctly discriminate the chirps (according to the nature of the transitions that evoke them) and, moreover, if they do this regardless of the presence or absence of silence in the base.

The results can establish the duplexity of the percept and the independence of its two sides. They can, in addition, speak to the possibility that the two perceptions of the transitions—as relevant to stop consonants and as chirps—are affected differently by silence, just as we should expect them to be if the

effectiveness of the silence cue is an instance of phonetic perception. However, a further and arguably more severe test of this possibility, employing discrimination tests on both sides of the percept, must wait for Experiment 2.

Method

Stimuli. In the top row of Figure 2 are the synthetic syllables from which the stimuli of the experiment were derived. Shown there is the silent interval that serves in these patterns as a necessary condition for the perception of either of the stop consonants [p] or [t]. Shown also are the contrasting formant transitions that underlie the distinction between these stops.

In the bottom row, we see how the syllables were divided into constituents for dichotic (and duplex-producing) presentation. The constituent shown at the bottom right of the figure is simply the transitions of the second and third formants, the only cues in these patterns that distinguish [spa] from [sta]. These formant transitions were produced on the parallel resonance synthesizer of the Haskins Laboratories. The [p] transitions started at 924 and 1,856 Hz and rose to 1,077 and 2,360 Hz; the [t] transitions started at 1,306 and 2,861 Hz and fell to 1,077 and 2,360 Hz. All transitions had a duration of 40 msec. When presented in isolation, they had 30 msec of steady-state formants appended to their trailing edges. The intensity of these appendages decreased gradually, complementing 30 msec of gradually increasing intensity in the corresponding formants of the "base"; though not essential to the duplex phenomenon, this manipulation of the intensity causes the transitions to "blend" more smoothly with the base. When the transitions were presented in isolation, they were perceived variously: some listeners identified them as speech-like sounds of some kind, but not as [p] or [t]; others heard them as

musical glissandos; most listeners (all in this experiment) found "chirps" an acceptable characterization.

The second constituent is displayed at the lower left of Figure 2 as the pattern labeled "base." This is what remains of the original syllables when the second- and third-formant transition cues have been removed and the transition of the first formant straightened. It consists of a patch of fricative noise, followed by a brief period of silence, and then by three steady-state formants. We straightened the first formant because, in the duplex percept, the rising transition seen in the pattern at the top of the figure is important but not absolutely necessary for the perception of a stop consonant. The result of this maneuver is to make the isolated second- and third-formant transitions carry, not only the distinction between [p] and [t], but also more of the information about stop-consonant manner.

The "fricative noise" of the base was produced from an s-like utterance (made by a human talker) in the following way. A 100-msec section of the noise was excerpted, low-passed at 5 kHz, and digitized. The amplitude envelope was tapered at either end, making the effective duration approximately 90 msec. There were energy peaks at about 3,500 and 4,500 Hz. The remainder of the base—the three formants—was produced on the Haskins parallel resonance synthesizer. The first formant centered at 637 Hz and had a duration of 256 msec. The second and third formants centered at 1,077 and 2,360 Hz, respectively. They began 40 msec after the first formant, with gradually rising intensity over the first 30 msec, and ended simultaneously with it. In the "silence" condition, the fricative noise was separated from the formants by 100 msec, during which there was little or no acoustic energy; in the "no-silence" condition, the corresponding interval was 20 msec.

To see how the base alone was perceived, we asked six members of the Speech Communication Group at M.I.T., none of whom had served in the main part of the experiment, to make phonetic

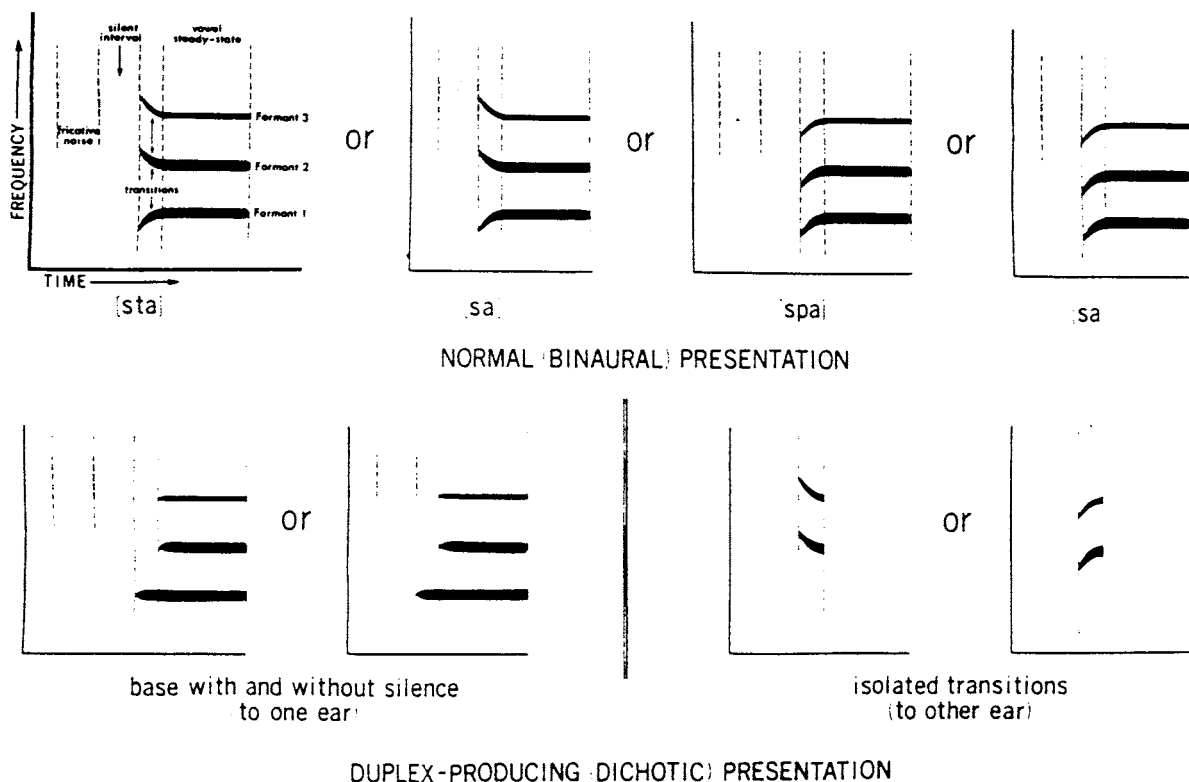


Figure 2. Schematic representations of patterns used as stimuli.

identifications of the base pattern in both its silence and no-silence forms. (This was done as part of the larger task of identifying all of the speech stimuli that were used in this experiment.) Having been invited to use whatever phonetic categories they found most appropriate, these subjects identified the silence version as follows: 16% of the time as [s-a] (i.e., as [s], followed by silence, followed by [a]), 27% of the time as [s?a] (i.e., [s], followed by glottal stop, followed by [a]), 28% of the time as [sa], and 29% of the time as [sta]. They identified the no-silence version 10% of the time as [s-a], 13% of the time as [s?a], and 77% of the time as [sa]. The variability of these judgments was high, the standard error being about 15%.

Procedure. The two constituents described above were always presented to a listener dichotically (base to the left ear, transitions to the right). They were timed to correspond to their occurrence in the normal (binaural) syllables. We will refer to each such dichotically presented pair as a "dichotic stimulus."

A single "experimental trial" consisted of two dichotic stimuli, presented sequentially, 420 msec apart. On each such trial, the base constituent of one of the dichotic stimuli had the silent interval appropriate for a stop consonant and the other did not. Both orders of presentation of the two conditions of the base (silence followed by no silence, no silence followed by silence) were used. The pairing of both conditions of the base (silence, no silence) with both conditions of the transitions ("p" transitions, "t" transitions) exhausted all possible combinations and yielded eight types of experimental trials. Thus, on any given experimental trial, a listener could not know which of the two base constituents would be in either dichotic stimulus or which of the two transitions would be paired with it.

The test tape was organized into 10 blocks of eight trials each. (Each block was a different permutation of the eight types of experimental trials.) The intertrial interval was 5 sec and the inter-block interval was 8 sec. All subjects listened to the test tape three times.

On the first time through the test tape, the subjects were asked to attend to the speech side of the percept and, on each trial, to identify the pair as [spa]-[sa], [sta]-[sa], [sa]-[spa], or [sa]-[sta]. The point of this was to see whether they "correctly" fused the isolated transitions with the base, and also whether the silence cue affected the transitions when, having been fused with the base, those transitions were perceived as speech. Next, the subjects were instructed to attend to the nonspeech side of the percept, and, on each trial, to judge the chirps as "same" or "different." The point here was to make sure that the listeners did also perceive the transitions as nonspeech chirps and, additionally, to determine whether that perception was affected by the silence cue. Lastly, subjects were asked to attend, on each trial, to both sides of the duplex percept, and to respond as before. This tested the possibility that the responses given previously might have depended on the subjects' being able, by an effort of attention, to experience but one side of the percept at a time. Subjects responded by writing [spa], [sta], or [sa] in the speech task and "S" or "D" in the nonspeech task. For the condition in which both tasks were carried out on each trial, the subjects first wrote the speech response, then the nonspeech response.

Subjects. The initial group of subjects comprised 14 college students. None was known to have any defect of hearing; all were native speakers of English; and, most important, all were naive with respect to the nature of the stimuli and the purpose of the experiment.

There was a screening procedure designed to eliminate any subjects who could not perceive the undivided syllables appropriately and also any who, upon being given trial dichotic presentations, did not report at least the impression of duplexity. As a result, 4 of the 14 subjects were excluded: two because they perceived some of the undivided syllables as containing [l] or [w] instead of [p] and [t], one because he apparently could not fuse the base and the isolated transition, and one because he did not hear the "chirp" as a clear and isolable percept.

Results and Discussion

The results of the tests for duplex perception are summarized in Table 1. Looking first at the judgments made on the speech side of the percept, we see that all subjects fused the dichotically presented constituents, "correctly" identifying the syllable pairs with a frequency significantly above chance ($p < .01$, by a binomial test). That is, they identified the syllables as [sa], [spa], or [sta] according to the presence or absence of silence in the base presented to one ear and the nature of the transitions presented to the other. This was so, not only when the subjects were attending exclusively to the speech percept, but also when they were attending to, and reporting on, both percepts. Plainly, the two kinds of isolated transitions did fuse with the base to produce the appropriate stop consonants. Just as plainly, the silence cue grossly affected the perception of the transitions when, on the speech side of the duplex percept, those transitions were serving as cues for stop consonants: given silence in the base, subjects correctly perceived [spa] or [sta] according to the nature of the transitions in the other ear; given no silence in the base, they perceived the "stopless" [sa], even though the transitions were present as before.

At this point, we should emphasize that perception of the syllables on the speech side of the duplex percept is, in a most important respect, like perception of those syllables when, as in the normal case, base and transitions are combined in a single stimulus pattern and presented to both ears. In neither case does a listener consciously perceive a chirp and then consciously decide to interpret it as one or another of the stop consonants. Rather, the listener perceives simply [sta], [spa], or [sa].

Table 1
Individual Subject Performance on the
Speech-Identification and (Nonspeech)
Chirp-Discrimination Tasks

Subject	Percentage of 'Correct' Judgments			
	Speech		Nonspeech	
	S	B	N	B
1	88.8*	71.3*	98.8*	90.0*
2	100.0*	98.8*	78.8*	70.0*
3	98.8*	82.5*	92.5*	71.3*
4	88.8*	73.8*	65.0	80.0*
5	100.0*	100.0*	88.8*	70.0*
6	93.8*	82.5*	72.5*	71.3*
7	97.5*	100.0*	50.0	57.5
8	93.8*	77.5*	50.0	62.5
9	90.0*	92.5*	90.0*	86.3*
10	91.3*	56.3*	93.8*	68.8*
Mean*	94.3	83.5	78.0	72.8

Note—Asterisks indicate above chance performance ($p < .01$) by a binomial test. Chance was 25% on the speech task (four response alternatives) and 50% on the nonspeech task (two alternatives). S = reporting on speech percept only; B = reporting on both percepts; N = reporting on nonspeech percept only.

It should also be noted about the speech side of the duplex percept that the formant transitions apparently provided two kinds of information. The more obvious kind pertains to place of production: given that a stop is heard, the transitions are plainly critical in determining whether it is alveolar ([t]) or labial ([p]). Less obvious, perhaps, is the fact that these same transitions also played a role in determining perceived manner—that is, in determining that the consonant heard was, in fact, a stop of some kind. That this is so may be inferred from the way expert listeners had identified the base component in isolation when, in the pretest described under Procedure, they were invited to apply any phonetic categories they found appropriate. When the base contained the interval of silence, the experts reported a stop consonant of some kind 47% of the time; but, if we exclude glottal stop on the ground that it is different in manner from proper stops such as [p] and [t] (see, e.g., Ladefoged, 1971), then stop manner was perceived only 27% of the time. That the subjects in the (duplex) experiment proper were perceiving stops with a greater frequency can be inferred from the data of Table 1, where it is seen that overall correct syllable identification was 94.3% when listeners were reporting on the speech percept only and 83.5% when they were reporting on both percepts. Since the pretest and the experiment proper are not strictly comparable, the contrasting results do not permit a firm conclusion that the second- and third-formant transitions provided information about manner as well as place. However, they do suggest that this was so.

Finally, we should consider about the speech side of the percept whether the subjects heard not only the result of fusion, as demonstrated above, but also the unfused "base" constituent. For several reasons, we are confident that there was only the "fused" percept. First, of the several expert listeners who attended carefully to the duplex phenomenon, none ever reported hearing two syllables, one corresponding to the base and one to the result of fusion with the transitions. As for the naive listeners who served in the experiment, two observations are pertinent. One is that when, having been presented with a number of sample dichotic stimuli in the pretest (described in Procedure) and asked to identify the speech percept as [sa], [spa], or [sta], none ever reported having heard two syllables at once. The other observation pertains to the perception of [spa]. Recall, first, that the base alone was never perceived by the expert listeners (as described under Procedure) as [spa]. Now, if the listeners in the main part of the experiment had perceived (in the dichotically presented patterns) two syllables, one resulting from the base alone and one resulting from the fusion of base and transitions, then, assuming the choice of either per-

cept was equally likely, they should have been correct with regard to [spa] no more than 50% of the time. That the listeners did considerably better than this in the (duplex) experiment proper can be inferred from the high accuracy of overall syllable identification, as seen in Table 1 and noted in the previous paragraph. Taking all of these considerations into account, we conclude that, as in the [ra]-[la] experiment described in the introduction, the base constituent was absorbed, as it were, in the process of fusion.

Turning now to the nonspeech side of the duplex percept, we should comment first that listeners did hear the transitions as chirps; moreover, they heard them as such whether or not there was silence in the base. More significantly, we see in Table 1 that most of the subjects judged accurately that the chirps were the same or different according as the transitions that gave rise to them were the same or different. When they had to judge only one of the two sides of the duplex percept, 7 of the 10 subjects performed at a level significantly above chance ($p < .01$). When both judgments were required on each trial, 8 of the 10 met the same criterion. (Performance for 7 of those 8 subjects was poorer when both judgments were made on each trial than when they were made on separate trials, although it was above chance on both. In any case, an exact comparison is difficult, because the difference between judgment conditions is confounded with the order of performance of the tasks, and hence with practice.)

As for the three subjects who failed to discriminate the chirps accurately in the attend-chirps-only task, two were put through a further experiment to determine if they nevertheless could hear the chirps, especially when there was no silence in the base. This was done with stimuli in which the transitions and the base were sometimes presented dichotically, as in the experiment proper, and sometimes mixed electronically and presented to both ears. In the former case, the subjects did perceive chirps whether or not the silence was present; in the latter case, in which they would not have been expected to perceive chirps, they did not.

Finally, we should report that subjects were slightly more likely to judge "same" when the stimuli were different than they were to judge "different" when the stimuli were the same. If silence had affected perception of the chirps, we should expect the opposite bias, since, on each experimental trial, silence was always present in one of the dichotic stimuli but not the other.

The principal conclusion that can be drawn from all of these results is that duplex perception of the formant transitions did occur. On the speech side, the transitions were essential to the perceived distinction between [spa] and [sta], but only when there was an appropriate period of silence in the base constit-

uent; without silence in the base, listeners perceived the "stopless" [sa], even though the same transitions had been presented. On the nonspeech side, the transitions were perceived as chirps. More importantly, these chirps were accurately discriminated as same or different according as the transitions that produced them were the same or different, and performance on this task was not measurably affected by the presence or absence of the silence cue in the base constituent presented to the other ear.

Second, these results provide some evidence relevant to the question: Does silence affect the transitions differently on the two sides of the duplex percept? An unequivocal answer is not to be had, if only because the perceptual tasks were different on the two sides of the percept, as, indeed, they had to be in order to test the duplexity. Still, it is surely worth noting that silence did have a gross effect on the transitions when they were being processed as speech, in which case they were critical to the perceived distinction between [spa] and [sta], although the same silence had no measurable influence on those same transitions when, simultaneously, they were being discriminated as nonspeech chirps.

EXPERIMENT 2

The primary result of Experiment 1 was that formant transitions were found to be perceived simultaneously in two phenomenally different ways: as speech, in that they provided information for perception of the stops in [spa] and [sta], and as nonspeech chirps. A further result was that a short interval of silence, known to be a cue for stop consonants, affected the speech but not the nonspeech. The interpretation of that further result was not straightforward, however, because the two sides of the duplex percept were measured in different ways: by identification on the speech side (because only identification could establish that the stimuli were, in fact, heard as speech) and by discrimination on the nonspeech side (because identification of the chirps is rather difficult and also not necessary for the purpose of proving that the subjects did, in fact, perceive the nonspeech appropriately). There was, then, no comparison of the effect of silence on speech and nonspeech percepts when the subjects had to perform the same task in response to both. The purpose of this second experiment is to repair that omission. Accordingly, the subjects will be required to discriminate not only the chirps, but also the speech. Given that duplex perception of the transitions was demonstrated in Experiment 1, these discrimination measures should provide a further test of the hypothesis that in the perception of these stops the effect of silence is phonetic rather than auditory.

Method

Stimuli. The stimuli of this experiment were identical to those of Experiment 1.

Procedure. As in the first experiment, a single experimental trial consisted of the presentation of one dichotic stimulus followed, after 420 msec, by presentation of another. In other respects, however, the procedure of this second experiment differed from that of the first. Most importantly, it differed in the task set for the subjects and in the combinations of dichotic stimuli that were used in the various experimental trials.

Consider, first, the subjects' task. It was, on both sides of the percept, to try to discriminate the successively presented stimuli of each trial. Subjects were asked to listen for a difference in these stimuli and then to report how confident they were that a difference had been detected. In rating confidence, they were instructed to use the following scale: "1" if "not confident" that a difference had been detected, "5" if "completely confident," and "2," "3," or "4" for intermediate degrees of confidence. It was strongly emphasized to all subjects that they were to base their ratings on any difference they could detect. Indeed, subjects were given explicitly to understand that even though two dichotic stimuli might appear to them as tokens of the same type (for example, as tokens of [sa]), they were nevertheless to listen carefully for any difference they might hear and, if confident a difference (of any kind) had been detected, to assign an appropriately high confidence rating.

As for the combinations of dichotic stimuli in the experimental trials, they were so composed as to exhaust all possible pairings of silence/no silence and "p"/"t" transitions. Thus, a single experimental trial had in its two base constituents one of the following three combinations: silence in both, silence in neither, or silence in one but not the other. As for the combinations of transitions, they were, on each experimental trial, either the same (both "p" or both "t") or different (one "p" and the other "t"). There were, then, three combinations of the base times two combinations of the transitions, making a total of six combinations overall. These six are the fundamental conditions of this experiment and will hereafter be so called.

For each of the conditions described above, we made several types of experimental trials. This was done in order to take into account that there were two ways in which the transitions could be the same (both could be "p" or both could be "t"), and also to counterbalance for order whenever the two dichotic stimuli of a trial were different (silence vs. no silence in the base constituents, or "p" vs. "t" in the transition constituents). The result was a total of 16 types of experimental trials. These were recorded onto a test tape in four different randomizations. With this procedure, the experimental conditions with silence in both base constituents were represented on the tape eight times each, as were those with silence in neither base. As a result of counterbalancing, the conditions with silence in one base constituent but not the other were represented 16 times each.

Having satisfied ourselves in Experiment 1 that subjects could, on each experimental trial, judge both sides of the duplex percept, we decided in this experiment to set them the simpler task of judging but one side of the percept at a time. The tape was presented four times. On two of those presentations, subjects were asked to judge the speech side of the percept; on the remaining two, they judged the nonspeech side, the order of speech and nonspeech judgments having been counterbalanced. There were, then, 16 speech and 16 nonspeech judgments made in each experimental condition that had silence in both base constituents or in neither; in the conditions with silence in one base constituent but not the other, 32 speech and 32 nonspeech judgments were made. The dichotic arrangement of the stimuli—the pairing of constituent (base or transitions) with ear (right or left)—was half the time one way and half the other. The order of these arrangements was counterbalanced.

Subjects. Ten college students were in the initial pool of subjects. As in Experiment 1, all were native speakers of English, none had any known hearing loss, and all were naive with respect to the nature of the stimuli and the purpose of the experiment.

The screening procedure employed in this experiment differed from that of Experiment 1. Here, there were two tests: Having been presented (binaurally) with the electronically fused constituents, the subjects were first asked to identify the resulting stimuli as [spa], [sta], or [sa]; then, having been presented (binaurally) with the isolated transitions, they were asked to identify them as patterns that "glided up" or "glided down." On the basis of these tests, 2 of the 10 subjects were eliminated: one because she could not identify the syllables and the other because she could not identify the chirps.

There was also a brief training session, aimed at getting the subjects accustomed to the dichotically presented pairs and to perceiving the two sides of the duplex percept. In this session, the patterns were presented dichotically, and the subjects, having been asked to attend to the speech on some trials and to the nonspeech on others, identified the stimuli as in the screening test. All subjects performed well with the speech stimuli, but two of the eight managed to perform only slightly above chance with the nonspeech chirps. Nevertheless, these two subjects were not eliminated from the experiment.

Results and Discussion

The aim of this experiment, it will be remembered, was to determine whether the silence cue has a different effect on the discriminability of the formant transitions when, on the one side of the duplex percept, they are critical for the perception of stop consonants and when, on the other, they are perceived as nonspeech chirps. In Figure 3, we see the mean confidence ratings that constitute the results of the experiment. These ratings reflect the subjects' confidence that they detected differences in the pairs of dichotic stimuli presented on each experimental trial. (The scale on which those ratings were ordered ranged from 1 to 5.) Plainly, there is a difference in the mean ratings, depending on whether the subjects were judging the speech or the nonspeech sides of the percept.

Consider, first, the leftmost panel of the figure, which displays the results for the condition in which there was no silence in either of the base constituents.

Though such a combination was never presented as such in Experiment 1, we should infer from the results obtained there that the speech side of the duplex percept would have sounded more or less like [sa], regardless of the transitions. Accordingly, we should expect that the transitions would be relatively hard to discriminate when perceived as part of the speech pattern. On the nonspeech side, however, we should suppose that, as in Experiment 1, discriminability would be relatively little affected by the absence of silence. The results of this second experiment confirm these expectations. Given no silence in either base constituent, the speech percepts were not well discriminated, though the ratings were somewhat higher, when the transitions were, in fact, different.¹ On the nonspeech side, the results stand in contrast. There, the transitions were relatively well discriminated when they were, in fact, different, although not, of course, when they were the same. A two-way analysis of variance (with the factors speech-nonspeech and same-different transitions) confirmed that silence did, indeed, affect the discriminability of the transitions differently on the speech and nonspeech sides of the percept [$F(1,7) = 26.17, p < .01$].

Consider, next, the center panel, where we see the results for the condition in which there was silence in one of the base constituents but not in the other. This is the same as the condition that was used throughout Experiment 1, where subjects identified the pattern with silence as [spa] or [sta] (depending on the nature of the transitions in the other ear), while identifying the pattern without silence as [sa] (regardless of the transitions). We are not surprised, therefore, to see, in Experiment 2, that when subjects discriminated the speech percepts they confidently perceived a difference between the "silence" and "no-silence" dichotic stimuli, and they did so whether the transitions were the same or different. (Presumably, they perceived a stop in the one case but not in the other.) The result on the nonspeech side is different. There,

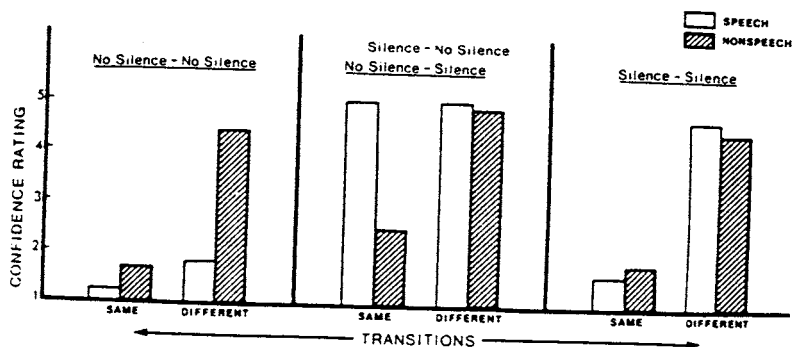


Figure 3. Mean ratings assigned in the conditions of the experiment. Ratings were assigned by the eight subjects to reflect their confidence that the two stimuli of each experimental trial were different.

the stimuli were readily discriminated when the transitions were different but not when they were the same, notwithstanding the fact that silence was always present in one of the dichotic stimuli but not in the other. That silence affected the discriminability of the transitions differently for speech and nonspeech in this condition is confirmed by analysis of variance [$F(1,7) = 40.93, p < .01$].

Finally, there is the condition in which there was silence in both base constituents. Although this condition was not presented as such in Experiment 1, we can infer from the results obtained there that all stimuli would have been perceived, on the speech side, as containing stops. What is more, stops would have been perceived to be the same or different depending on whether the transitions were the same or

different. Not surprisingly, we see this inference supported in the results of Experiment 2: subjects discriminated the speech percepts as different when the transitions were different, but not when the transitions were the same. On the nonspeech side, we should expect the same result, and we see that it was, in fact, obtained. That discriminability of the transitions was not significantly different on the speech and nonspeech sides of the percept was confirmed by analysis of variance [$F(1,7) = .68$].

To see how fairly the group data, as shown in Figure 3 and discussed above, represent the performances of individual subjects, we should examine Table 2. There, we see that seven of the eight subjects conformed quite well to the group result. The single exception (Subject 8) is one of the two subjects who, as noted under Method, performed poorly with the chirps during the training session that preceded the experiment proper.

The results can be summarized quite simply: The silence cue had a different effect on discrimination of the formant transitions depending on whether they supported the perception of stop consonants or whether, alternatively, they were perceived as nonspeech chirps.

SUMMARY AND CONCLUSIONS

The first conclusion, established by the results of Experiment 1, is that the formant transitions simultaneously supported speech and nonspeech percepts. This duplex perception was obtained by presenting a base constituent to one ear and, at the same time, particular formant transitions to the other. The base constituent consisted of a patch of fricative noise appropriate for [s], followed by a vocalic segment appropriate for [a]. These acoustic elements were sometimes separated by a brief period of silence and sometimes not. The formant transitions, presented to the other ear, were appropriate for the stop consonants [p] and [t]. By itself, the base constituent was perceived as a syllable, but never as [spa] and rarely as [sta]; the transitions were perceived in isolation as not-very-speechlike chirps. Given dichotic presentation of the two constituents, there were three consequences in perception: (1) the two constituents (base and transitions) fused to produce the same syllable—[spa], [sta], or [sa]—that would have been perceived had the components been electronically mixed and presented in normal fashion; (2) the transitions that, in isolation, produced nonspeech chirps continued to do so; and (3) the base constituent that, in isolation, produced a syllable different from the "correct" fused syllable was not perceived as such, having been absorbed, as it were, in the process of fusion. What the listeners heard, then, was precisely duplex: a syl-

Table 2
Confidence Ratings Assigned by the Individual Subjects

Subjects	Transitions			
	Same		Different	
	Speech	Non-Speech	Speech	Non-Speech
	No Silence/No Silence* Experimental Condition			
1	1.00	2.00	1.00	4.50
2	1.51	1.57	1.57	4.69
3	1.38	1.44	1.51	4.44
4	1.32	1.07	3.32	4.75
5	1.00	1.88	1.00	4.63
6	1.26	1.00	1.75	4.94
7	1.25	2.25	3.00	4.32
8	1.25	2.13	1.25	2.63
Mean	1.25	1.67	1.80	4.36
	Silence/No Silence and No Silence/Silence** Experimental Condition			
1	5.00	2.88	5.00	4.97
2	5.00	2.04	5.00	4.85
3	5.00	2.69	5.00	4.69
4	5.00	1.23	5.00	5.00
5	5.00	2.72	5.00	4.91
6	4.63	1.00	5.00	4.94
7	5.00	3.91	5.00	5.00
8	5.00	3.38	5.00	4.38
Mean	4.95	2.48	5.00	4.84
	Silence/Silence* Experimental Condition			
1	1.50	2.13	4.94	4.94
2	1.82	1.57	4.38	4.63
3	1.75	1.69	4.82	4.44
4	1.63	1.00	4.75	5.00
5	1.38	2.94	5.00	4.75
6	1.83	1.00	3.88	4.94
7	1.50	2.25	4.75	4.32
8	1.57	2.13	5.00	2.63
Mean	1.62	1.84	4.69	4.46

*Each of these scores is the mean of 16 judgments.
of these scores is the mean of 32 judgments.

**Each

lable in which the transition cues were interpreted in linguistically normal fashion, and a nonspeech chirp that resembled the percept produced when the transitions were presented alone. Evidently, the subjects simultaneously perceived the transitions in two ways. We have termed the one way "auditory," because the percept (the chirp) resembles that produced by sounds other than speech. We have termed the other way "phonetic," because its percept has those characteristics we ordinarily associate with consonants and vowels—that is, it has those characteristics that are manifest when patterns of sound are being perceived as speech.

The second conclusion pertains to the effect of the silence cue on the perception of the transitions, and thus to the role of silence in the perception of stops. In the results of Experiment 1, we saw a large effect of silence on the speech side of the percept: when paired with a base constituent that contained the silence cue, the formant transitions led to the perception of the appropriate stop consonants; when paired with a base that lacked the silence cue, they did not. On the nonspeech side, where the transitions were perceived as chirps, silence had no measured effect: although silence was paired with one transition and not with the other, most subjects correctly perceived that the corresponding chirps were the same when the transitions that produced them were the same and different when the transitions were different. The implication of those results was supported by the outcome of Experiment 2, in which the subjects were required, on both sides of the percept, to discriminate the dichotic stimuli of each experimental trial on the basis of any differences they could detect. There it was found that the effect of silence on discrimination of the transitions was significantly different when, in the one case, those transitions were supplying information critical for the perception of stop consonants, and when, in the other, they were being perceived as nonspeech chirps. Putting these results together, we conclude that the importance of silence in the perception of stop consonants is owing to specifically phonetic (as distinguished from generally auditory) processes, and that the effect of silence in such cases is an instance of perception in a distinctively phonetic mode.

REFERENCE NOTES

1. Darwin, C. J., Howell, P., & Brady, S. A. *Laterality and localization: A right ear advantage for speech heard on the left* (Status Report on Speech Research, SR-48, 257-277). New Haven, Conn: Haskins Laboratories, 1976.
2. Bastian, J. *Silent intervals as closure cues in the perception of stops* (Speech Research Instrumentation, 9, Appendix F). New Haven, Conn: Haskins Laboratories, 1962.

REFERENCES

BAILEY, P., & SUMMERFIELD, A. Q. Information for speech; Observations on the perception of [s]-stop clusters. *Journal of*

Experimental Psychology: Human Perception and Performance, 1980, 6, 536-563.

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

BROADBENT, D. E., & LADEFOGED, P. On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, 1957, 29, 708-710.

DARWIN, C. J. The perception of speech. In E. Carterette & M. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic Press, 1976.

DORMAN, M. F., RAPHAEL, L. J., LIBERMAN, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, 65, 1518-1532.

FITCH, H. L., HALWES, T., ERICKSON, D. M., & LIBERMAN, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 1980, 27, 343-350.

HALWES, T. G. *Effects of dichotic fusion on the perception of speech*. Unpublished doctoral dissertation, University of Minnesota, 1969. (Issued as Supplement to Haskins Laboratories Status Report on Speech Research, September 1969)

ISENBERG, D. S., & LIBERMAN, A. M. Speech and nonspeech percepts from the same sound. *Journal of the Acoustical Society of America*, 1978, 64(S1), S20. (Abstract)

KUHL, D. S., & MILLER, J. D. Speech perception by the chin-chilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, 190, 69-72.

LADEFOGED, P. *Preliminaries to a linguistic phonetics*. Chicago: University of Chicago Press, 1971.

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

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. S., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8) *Perception*. Heidelberg: Springer-Verlag, 1978.

MILLER, J. D. Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory systems. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. Berlin: Dahlem Konferenzen, 1977.

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

REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 621-637.

STUDDERT-KENNEDY, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

TUREK, S. V., DORMAN, M. F., FRANKS, J. R., & SUMMERFIELD, Q. Identification of synthetic /bdg/ by hearing-impaired listeners under monotonic and dichotic formant presentation. *Journal of the Acoustical Society of America*, 1980, 67, 1031-1040.

NOTE

1. Just how discriminable patterns of this sort will be depends, in our experience, on several factors. When silence is removed from a pattern containing a "t" transition, the resulting percept is not likely to be very different from a perfectly normal [sa], if only because the places of production (hence, the second- and third-formant transitions) for "t" and "s" are virtually the same (alveolar). The "p" transitions, on the other hand, are appropriate to a different place of production (bilabial); hence, they are not so readily "absorbed" into the fricative percept when, in the ab-

sence of silence, perception of the stop vanishes. If the "p" transitions are of very low intensity, it is possible that the listener will simply perceive [sa]. But, if perception is affected by the "p" transitions, then we can expect any one of the following consequences: (1) The perceived fricative takes on the place of production of the "p" transitions, in which case the percept becomes [fa]; (2) a semivowel appropriate to the place of the "p" transitions is introduced, in which case the percept becomes [swa]; or (3) the tran-

sitions are rejected as speech yet remain audible, in which case the listener is aware of a nonspeech "chirp" or "thump." At all events, we do not expect—at least not in all cases—that the "t" and "p" transitions will be perfectly indiscriminable when they are heard as speech in the no-silence condition.

(Manuscript received February 13, 1981;
accepted for publication May 11, 1981.)