715

Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes

Shlomo Bentin
Department of Psychology, The Hebrew University, Jerusalem, Israel
and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Virginia Mann
Department of Cognitive Sciences, University of California, Irvine, California 92717
and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

(Received 24 April 1989; accepted for publication 28 March 1990)

Using the phenomenon of duplex perception, previous researchers have shown that certain manipulations affect the perception of formant transitions as speech but not their perception as nonspeech "chirps," a dissociation that is consistent with the hypothesized distinction between speech and nonspeech modes of perception [Liberman *et al.*, Percept. Psychophys. **30**, 133–143 (1981); Mann and Liberman, Cognition **14**, 211–235 (1983)]. The present study supports this interpretation of duplex perception by showing the existence of a "double dissociation" between the speech and chirp percepts. Five experiments compared the effects of stimulus onset asynchrony, backward masking, and transition intensity on the two sides of duplex percepts. It was found that certain manipulations penalize the chirp side but not the speech side, whereas other manipulations had the opposite effect of penalizing the speech side but not the chirp side. In addition, although effects on the speech side of duplex percepts have appeared to be much the same as in the case of normal (electronically fused) speech stimuli, the present study discovered that manipulations that impaired the chirp side of duplex percepts had considerably less effect on the perception of isolated chirps. Thus it would seem that duplex perception makes chirp perception more vulnerable to the effects of stimulus degradation. Several explanations of the data are discussed, among them, the view that speech perception may take precedence over other forms of auditory perception [Mattingly and Liberman, in *Signals and Sense: Local and Global Order in Perceptual Maps*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York, in press); Whalen and Liberman, Science **237**, 169–171 (1987)].

PACS numbers: 43.71.An, 43.71.Es, 43.66.Ba

INTRODUCTION

When we listen to speech, we tend to be unaware of the auditory signal qualities that give rise to our linguistic percepts. Careful "analytic" listening can reveal many such qualities—the hiss of fricative noises, the pops and clicks of stop consonant release bursts, etc. (Pilch, 1979; Repp, 1981). Yet, there exist certain auditory qualities in speech that remain inaccessible to even the most analytic listener. These qualities reflect energy that resides in particular regions of the spectrum such as the frequencies of individual formants and their changes over time; they can be made audible only in certain experimental situations. For example, the "chirpy" auditory quality of single formant transitions can be made audible when the transition is extracted from an utterance and presented in isolation. In that case, the formant transition is heard as a chirp, and discrimination between various pairs of transitions is a nearly continuous function of their difference in frequency modulation. Yet, when the same transitions are integrated into an acoustic speech pattern, where they cue the distinction between certain stop consonants, their chirpy quality is lost and discrimination is categorical (Mattingly *et al.*, 1971; Mann and Liberman, 1983).

A particularly appropriate method of comparing these two ways of perceiving formant transitions was devised by Rand (1974); its consequences have been dubbed "duplex perception" (Liberman, 1979). To induce duplex perception, one starts with a minimal pair of acoustic speech stimuli that differ only in a single cue, /da/ and /ga/, for example. Each stimulus is then divided into two parts. One part is a critical cue for the distinction between the two syllables, e.g., the second or third formant transition. The other is the remainder of the stimulus, that portion which is the same for /da/ and /ga/ and which is referred to as the "base." By changing the harmonic structure of the transition (see Whalen and Liberman, 1987) or by presenting the base and transition to separate ears (see Liberman *et al.*, 1981; Mann and Liberman, 1983; Nygaard and Eimas, 1990) one then introduces a discrepancy that results in a new stimulus configura-

 0001-4966/90/070064-11$00.80

tion in which listeners can infer that the two parts of the stimuli are produced by separate sources. When listeners hear stimuli in this new "discrepant" configuration, two percepts arise: a speech sound and a chirp that seems to "float" away from the speech. What is important about these two "sides" of the duplex perception is that each involves perception of the formant transition. Listeners report /da/ or /ga/ according to the nature of the transition, indicating that they have integrated the transition and the base rather than having processed the base alone. They simultaneously hear a chirp sound, which arises from a separate spatial location than the speech sound and has the auditory quality of the transition presented in isolation—a quality that is not heard in the original /da/ and /ga/ stimuli.

Duplex perception offers a controlled way of comparing perception of formant transitions as part of speech and as nonspeech, because the two percepts arise from the same physical stimulus configuration. The experimenter can hold all variables constant and selectively direct a listener's attention to either of two readily available and differently localized percepts. If the two percepts rise from the operation of different "modes" of perception, as has been suggested by several authors (i.e., Liberman *et al.*, 1981; Mann and Liberman, 1983; Repp *et al.*, 1983; Repp and Bentin, 1984), then it should be possible to separately and selectively alter each percept. If two separate modes exist, certain additions to the stimulus or changes in its structure might alter one but not the other.

Two previous studies of duplex perception have confirmed one aspect of this prediction by showing that the chirp side of duplex percepts remains unaltered by manipulations that significantly alter the speech side. Liberman *et al.* (1981) found that prefixing the base with a noise appropriate to /s/ has no effect on the listeners' ability to perceive /p/ and /t/ transitions in the other ear as rising versus falling chirps. However, the noise makes the transitions indistinguishable as cues for a speech percept, because it causes both /pa/ and /ta/ to be heard as /sa/. The effect of the /s/ noise on the speech side of the duplex percepts is exactly as it would have been had the transition and base been electronically fused. It reflects the operation of a "specifically phonetic process" (Liberman *et al.*, 1981, p. 142) in which the perception of /p/ vs /t/ is prevented by the fact that the /s/ noise has replaced the closure silence that is an important cue to the perception of stop consonants. Another study by Mann and Liberman (1983) makes a similar point. In that study, it was shown that preceding the base by natural tokens of the syllables /al/ and /ar/ has no effect on the listeners' ability to discriminate the chirp side of duplex /da/ and /ga/ stimuli, which contained transitions that varied in onset frequency. Yet the phonetic discrimination of these same transitions as cues for /d/ vs /g/ is systematically influenced by the preceding syllables because they induce a change in the location of the category boundary. Here, as well, the influences on the speech side of duplex percepts are the same as those that occur when the base and transition are electronically fused; they presumably reflect a specifically phonetic process in which listeners take account of the assimilating consequences of coarticulating /l/ or /r/ with a following /d/ or /g/ (as discussed in Mann, 1986).

Thus it seems evident that, by manipulating certain aspects of the stimulus, one can alter the speech side of duplex percepts while leaving the chirp side unchanged. However, a stronger case for the dissociation between speech and nonspeech perception requires a "double dissociation." It is desirable to show not only that the speech side of duplex percepts can be altered while leaving the chirp side unchanged, but also that the chirp side can be altered while leaving the speech side unchanged. To this end, we have examined the effects of various acoustic manipulations on each side of the duplex percept. Experiment I investigated the effect of a temporal separation between the base and the second formant transition. Experiments II through IV examined the effects of certain types of masking and experiment V examined the effect of a decrease in transition amplitude. If any of these manipulations has a greater effect on the chirp side than on the speech side, the double dissociation between two perceptual modes would be confirmed.

I. EXPERIMENT I

Experiment I examined the effect of a stimulus onset asynchrony (SOA) between the transition and the base in duplex stimuli, asking whether this manipulation will have selective effects on speech perception as compared to chirp perception. Given previous results (Cutting, 1976; Repp and Bentin, 1984), we had reason to think that speech perception would be penalized as separation of the base and transition approached 100 ms. However, previous research had not examined the effect of SOA on the chirp side of duplex percepts. It was important that we determine how each side is affected by SOA before we turned to our studies of the effect of backwards masking in experiment II.

A. Method

1. Subjects

The subjects were ten undergraduates. Two additional subjects were excused after they failed to distinguish /ba/ and /ga/ in the duplex practice series that preceded the test series.

2. Materials

The duplex stimuli comprised a base and second-formant transitions that were adapted from two-formant synthetic approximations to the syllables /ba/ and /ga/, produced on the parallel resonance synthesizer at Haskins Laboratories. The base by itself sounded vaguely like 'da'; it contained the first formant ($F1$) and the steady-state of the second formant ($F2$). Its total duration was 300 ms, with a 25-ms amplitude ramp at onset, and a 100-ms amplitude ramp at offset. Its fundamental frequency decreased linearly from 114 to 79 Hz. During the first 50 ms, $F1$ rose from 100 to 765 Hz, at which point it became steady state and was joined by a constant $F2$ at 1230 Hz. There was no energy in the $F2$ region during the first 50 ms. The $F2$ transitions were synthesized separately from the base, with pitch and amplitude contour identical to that of the first 50 ms of the base. The /ba/ transition started at 924 Hz and rose linearly to

1230 Hz; the /ga/ transition started at 2298 and fell linearly to 1230 Hz. The absolute amplitudes of the transitions were set at the values $F2$ would have had in intact syllables.

To ensure that subjects could adequately perceive the chirp and speech components of the duplex stimuli, three practice sequences were recorded. The first was designed to familiarize subjects with the /ba/ and /ga/ syllables. It presented the base electroncally fused with each of the $F2$ transitions, five times each and then five times in alternation. The second practice series was designed to familiarize the subjects with the chirps; it presented the isolated $F2$ transitions five times each and then five times in alternation. The third practice series familiarized subjects with duplex percepts; it presented the base and $F2$ on separate channels in onset synchrony, with each transition heard five times separately and then five times in alternation. In the test series designed to assess the effects of SOA, the /ba/ and /ga/ transitions preceded the base eight times at each of eight different SOAs: 0, 20, 40, 60, 70, 80, 90, and 100 ms. This yielded a total of 128 stimuli that were recorded in random sequence with intertrial intervals (ITIs) of 2.5 s and longer pauses between blocks of 16 stimuli.

3. Procedure

Each subject participated in two sessions, counterbalanced across subjects. Here, and in all the other experiments reported in this paper, the base was always heard in the left ear, the second formant transitions were always heard in the right ear (in our informal experience, ear differences tend to be negligible). One session required labeling of the speech percepts as 'ba' and 'ga'; it was preceded by the first and third practice series (i.e., the fused syllables and the duplex stimuli). The other session required labeling of the chirps as "rising" or "falling"; it was preceded by the second and third practice series (i.e., the isolated $F2$ transitions and the duplex stimuli).

B. Results and discussion

As SOA increased, the identification of the speech percepts as ba or ga became considerably less accurate. In contrast, identification of the nonspeech percepts as rising or falling chirps improved slightly. These results appear in Fig. 1, where it may be seen that speech identification declined with SOA, from a level of almost 95% correct to a level of 60% accuracy as SOA approached 100 ms. In contrast, the accuracy of chirp identification increased from a level of 85% to a level of 95% accuracy for SOAs greater than 0 ms. A repeated-measure two-way analysis of variance (ANOVA) with the factors Task (speech or nonspeech identification) and SOA confirmed these observations. There was a significant effect of Task [$F(1,9) = 12.39$, $MSe = 780$, $p < 0.005$] and of SOA [$F(7,63) = 3.39$, $MSe = 65$, $p < 0.004$]. More importantly, there was a significant Task by SOA interaction [$F(7,63) = 8.75$, $MSe = 73$, $p < 0.0001$].

These results indicate that perception on the speech side of the duplex percepts is indeed disrupted by a temporal separation between base and transition. The chirp side of duplex percepts was not impaired by the temporal separation—it was slightly facilitated—and this is consistent with the claim that chirp and speech perception in the duplex phenomenon are mediated by two different modes of perception, a claim which will be more rigorously tested in experiments II–V. The primary contribution of this experiment is its confirmation that the duplex speech percepts rest upon integration between the base and second formant transition. Although listeners have been reported to be capable of labeling isolated second formant transitions as speech (Nusbaum *et al.*, 1983), it appears that integration of the transitions and base is an important component of the speech percepts in the duplex phenomenon that concerns us here.

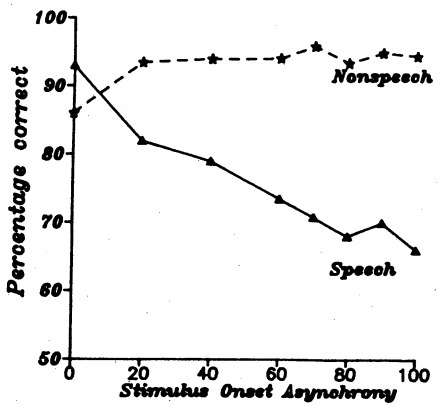

FIG. 1. Categorization of speech and nonspeech sides of duplex percepts at different SOAs (in ms) between the transition and the base.

II. EXPERIMENT II

In this experiment, we sought to further dissociate speech and chirp perception by finding a manipulation that alters chirp perception but leaves speech perception unaffected. We chose to examine the effects of a white noise presented immediately after the transition in the same ear. Our test involved a partial replication of experiment I in which we examined the effect of increasing SOA, while placing white noise following the transition. The possibility that a backwards mask might have selective effects on the chirp side of duplex percepts as suggested by the results of experiment I, where chirp perception improved slightly as presentation of the base was delayed in time. We were also prompted to consider the possibility that backward masking might have a greater effect on chirp perception than on speech perception by some of the data in Mann and Liberman (1983). Their first experiment on duplex perception revealed that the subjects' ability to discriminate chirps in the duplex condition was inferior to their ability to discriminate chirps in isolation, and it was speculated that the decrease in chirp performance in the duplex condition was a consequence of the distracting circumstance of hearing two simultaneous percepts. However, we noted that any penalizing effects of the "distraction" were unique to chirp perception, hence selective masking seemed as reasonable an account as selective attention.

A. Method

1. Subjects

The same subjects that were tested in experiment I also participated in the present experiment.

2. Materials

The stimuli were identical to those used in experiment I with the exception that a masking stimulus immediately followed each of the two transitions, presented to the same ear. The masking stimulus was a 15 ms burst of white noise with abrupt on-and off-sets; intensity was slightly above the peak amplitude of the isolated transitions. In the test series, the interval between transition and noise was fixed at 0 ms, and each transition preceded the base eight times at each of three different intervals: 0, 20, and 40 ms. This yielded a total of 48 stimuli that were recorded in random series with the same ITIs, etc. as in experiment I.

3. Procedure

The test series were presented immediately following experiment I. Thus, before being given the present test, all subjects were exposed to the three practice and two test series employed in experiment I. All subjects were tested twice: In one session they were instructed to direct attention to the speech side of the duplex and categorize the syllables as 'ba' or 'ga'. In the other, they were asked to attend to the nonspeech side and to categorize the chirps as rising or falling. The order of the tasks was counterbalanced across subjects.

B. Results and discussion

Because the same subjects and stimuli had been employed in experiments I and II, the effect of the white-noise backward mask on perception of speech and chirps at each SOA was assessed in comparison to the unmasked condition in experiment I. These results are presented in Fig. 2 where the solid lines represent data obtained in experiment I (unmasked condition) and the dotted lines represent those obtained in experiment II (masked condition). Speech perception is on the left, and chirp perception on the right.

The differential effect of the white noise mask on the perception of speech and of chirps is evident: The presence of the mask had a substantial interfering effect on chirp perception, but no effect on speech perception. A repeated-measures three-way ANOVA with the factors Task (speech or chirp), SOA, and Masking confirmed these observations. As in the previous experiment, the effects of task, of SOA, and the interaction between them were significant [$F(1,9) = 12.39$, $MSe = 780$, $p < 0.005$; $F(7,63) = 3.39$, $MSe = 65$, $p < 0.004$, and $F(7,63) = 8.75$, $MSe = 73$, $p < 0.004$, respectively]. The effect of the noise mask was significant $F(1,9) = 7.59$, $MSe = 269$, $p < 0.02$, and the Task by masking interaction was highly significant, $F(1,9) = 53.47$, $MSe = 72$, $p < 0.0001$.

The present results would seem to complement previous studies that showed selective effects on the speech percepts (Liberman *et al.*, 1981; Mann and Liberman, 1983), offering a case of selective effect on chirp perception, which supports the conclusion that the two sides of the duplex phenomenon represent the operation of two different modes of perception. However, further research is required to discover how the effects of the mask is to be explained.

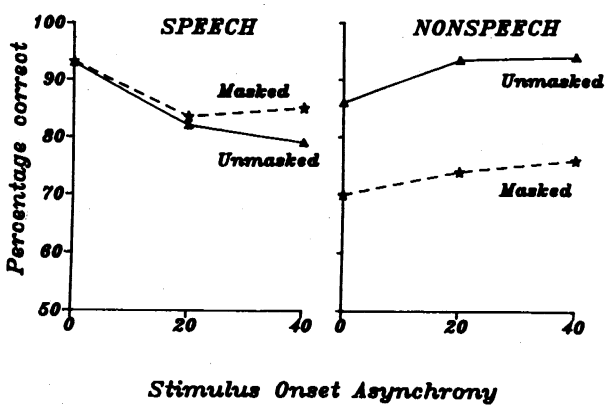

FIG. 2. Backward masking effects on speech and nonspeech sides of duplex percepts at different SOAs (in ms) between the transition and the base.

III. EXPERIMENT III

One explanation of the differential effect of the white-noise mask on chirp and speech perception is that the mask involved mechanisms that operates only within a particular auditory "stream." Drawing on the work of Bregman (1978, 1981) we might suppose that in the processing of auditory signals, "scene analysis" incorporates the transition into two separate streams. On the basis of spatial location (i.e., ear of origin) the transition and base would be assigned to separate streams whereas on the basis of a common acoustic attribute–the $F0$ contour–the base and chirp would be assigned to one and the same stream. We might then postulate that, if masking is stream-specific, masking of the chirp side of duplex percepts should decrease when the white-noise mask is presented to the ear receiving the base (i.e., there should be less masking because the mask and chirp are now perceived in separate streams). Another possibility is that speech perception is more tolerant than chirp perception of the signal degradation induced by a masking stimulus. If this is so, a contralateral mask should have the same effect as the ipsilateral mask employed in experiment II, masking the chirp side of the duplex percepts more than the speech side. This would be consistent with Massaro's (1970) report that contralateral backwards masking interferes as much with pitch (i.e., chirp) perception as does binaural backward masking. As a test of these hypotheses, experiment III studied the effects of a noise mask placed either contralateral or ipsilateral to the transition. A new set of stimuli modeled after those of Repp and Bentin (1984) was used so that we might establish the generality of experiment II to stimuli that contained third formant transitions.

A. Method

1. Subjects

The subjects were 12 undergraduates who had participated in a pilot experiment that used the same duplex

/da/-/ga/ stimuli. They had each demonstrated themselves capable of hearing the speech and chirp percepts in these stimuli.

2. Stimuli

The stimuli for this experiment were three-formant approximations of /da/ and /ga/, newly synthesized on the Haskins Laboratories equipment, and a 15 ms burst of white noise. In contrast to the earlier /ba/-/ga/ stimuli, the critical distinction between the present syllables was carried by the $F3$ transition. The base, which sounded like 'da' in isolation, was 250 ms in duration with a 50-ms amplitude ramp at onset and a constant fundamental frequency of 100 Hz for the first 100 ms followed by a linear decrease to 80 Hz at offset. $F1$ began at 279 Hz and increased linearly in frequency during the first 50 ms to a steady state of 765 Hz. $F2$ began at 1650 Hz and decreased linearly during the first 50 ms to a steady state of 1230 Hz. No steady-state $F3$ was included, as Repp and Bentin (1984) had shown that this was not critical and made the base sound more like 'da'. The /da/ and /ga/ $F3$ transitions were each 20 ms long, the short length being designed to enhance the effects of masking (the relative effectiveness of such short transitions had also been shown in Repp and Bentin, 1984). The /da/ transition began at 2800 Hz and decreased to 2745 Hz, the /ga/ transition began at 1800 and increased to 1945 Hz.

Three test sequences were prepared, one control (unmasked) and two masking conditions differing in the relative positions of the transition and the noise mask. In the ipsilateral masking condition, which replicated the masking condition of experiment II, the 15-ms noise burst immediately followed the transition and was heard in the same ear. In the contralateral masking condition, the mask occurred in the same temporal position, but was heard in the same ear as the base. SOA between the transition and base was set at 35 ms to avoid actual overlap of the mask with either the base or the transition. The control condition presented the transition and the base without the mask, at 35 ms SOA. Each sequence comprised 36 /da/ and 36 /ga/ trials, randomized into three blocks of 24 trials each. ITI was 2.5 s and there were longer pauses between blocks of stimuli.

3. Procedure

Each subject participated in a speech and nonspeech session, with order counterbalanced across subjects. Speech percepts were labeled as 'da' or 'ga', and the labels high and low were used for the chirps, following Repp and Bentin (1984).

B. Results and discussion

Consistent with the SOA effects observed in experiment I, chirp identification was superior to speech identification in all three conditions (Fig. 3). The more important result, however, was that the effect of the mask on perception of the transition depended on whether it was heard in the same ear as the transition, or the same ear as the base. When the mask was in the same ear as the transition (ipsilateral condition) it reduced the accuracy of chirp identification, but had little

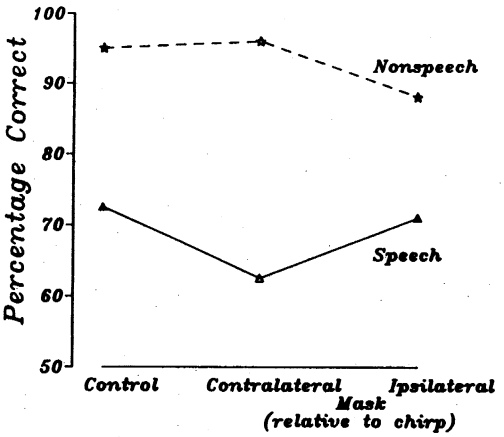

FIG. 3. The effect of ipsilateral and contralateral backward masking on the speech and nonspeech sides of duplex percepts.

effect on the accuracy of speech identification, replicating experiment II and extending that result to third-formant transitions. In contrast, when the mask was in the same ear as the base (contralateral condition), it penalized speech perception but had no effect on chirp perception.

These observations were confirmed by a two-way repeated-measures ANOVA conducted with the factors Mode (speech, nonspeech) and Masking condition (control, ipsilateral, contralateral). The effect of Mode was significant, $F(1,11) = 26.79$, $p < 0.0004$, and Mode and Condition interacted, $F(1,15) = 5.49$, $p < 0.025$. Tukey HSD post-hoc comparisons revealed that the labeling of the speech percepts was significantly worse in the contralateral masking condition than in either of the other two conditions ($p < 0.01$). In contrast, labeling of the chirps was significantly worse in the ipsilateral masking condition relative to the other two conditions ($p < 0.01$).

It appears that a noise that occurs in the same ear as the base interferes with the speech side of duplex percepts but has no particular effect on chirp side, whereas a noise in the same ear as the transition has the opposite effect. These results are consistent with an account in which masking is stream-specific and we shall defer further discussion of the role of scene analysis in duplex perception until the final discussion. The fact that the contralateral mask had a selective effect on speech perception does not support a view that speech perception is inherently more tolerant of degraded signals. The fact that the contralateral mask had less of an effect on chirp perception than the ipsilateral mask might further seem to be at odds with Massaro's (1970) report that contralateral backward masking of tonal targets by tonal masks interferes with pitch perception as much as does binaural backward masking, but it should be remembered that our stimuli are considerably different form his, as ours involved brief white-nose masks and duplex stimuli in which speech and nonspeech percepts are simultaneously present.

IV. EXPERIMENT IV

In experiments II and III we had observed that ipsilateral backward masking had a greater effect on the chirp side of duplex percepts than on the speech side. We now turn to

the question of whether this type of masking influences chirp perception in the case of isolated formant transitions. A control of this sort offers a test of a "naive" auditory masking acount in which masking should be more-or-less the same in duplex stimuli and in isolated transitions. Experiment IV addressed this possibility, using several mask intensities in order to maximize the possibility of finding masking in each condition.

A. Method

1. Subjects

The subjects were 12 undergraduates who served as paid volunteers. They were screened from a larger pool of undergraduates, using a selection criterion of > 80% correct identification of syllables in duplex presentation. The data of one subject were excluded from analysis because of very poor performance in identifying isolated transitions as chirps, which left 11 subjects.

2. Stimuli

The duplex stimuli were derived from the same synthetic /da/ and /ga/ syllables that had been used in experiment III, except that there was no SOA between the base and the transitions. The masking noise employed in this experiment was excerpted from white noise digitized at 10 kHz and low-pass filtered at 4.9 kHz. It was 10 ms in duration and at its highest amplitude the rms intensity was about 32 dB relative to that of the transitions. Four additional masks were created by digitally attenuating the noise in steps of 6 dB.

3. Procedure

The five mask intensity conditions and the no mask baseline condition were presented in a blocked design. Each block contained 48 stimuli, 24 /da/ and 24 /ga/, presented in random order. The blocks were presented in a fixed order for all subjects, the first block was the baseline condition, and the next five blocks presented masked stimuli with the relative intensity of the mask increased in equal steps of 6 from 8 dB (block 2) to 32 dB (block 6). Subjects were tested in two sesssions, one for speech and one for nonspeech perception and the order was counterbalanced across subjects. In the nonspeech condition, subjects listened to the tape twice, once when the bases were presented to the ear opposite the transitions (i.e., the normal duplex listening condition) and once when the channel presenting the bases was disconnected such that only the isolated transitions could be heard. The order of the two nonspeech conditions was also counterbalanced.

B. Results and discussion

The results for the three different conditions are compared in Fig. 4, where the mean accuracy of identification in each condition appears as a function of mask intensity. Identification of the speech and chrips in the baseline (no-mask) condition was uniformly high, but addition of the masking noise markedly reduced performance for chirps in the duplex condition.

Analysis of variance revealed significant main effects of

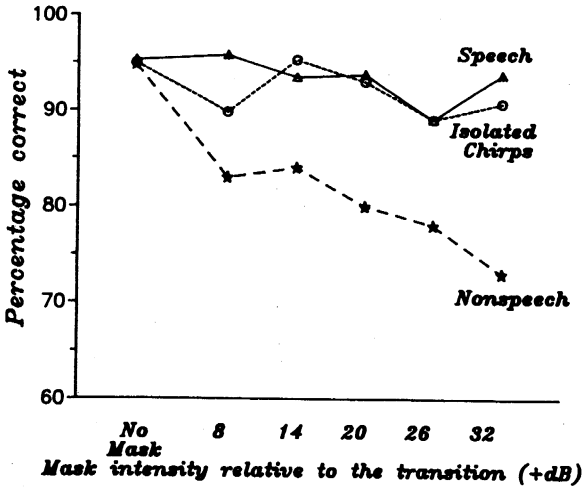

FIG. 4. The effect of backward mask intensity on the speech and nonspeech side of duplex percepts, compared with its effect on the perception of isolated transitions as chirps.

perception mode, $F(2,20) = 13.97, p < 0.0002$, and of mask intensity, $F(5,50) = 9.00, p < 0.0001$, as well as a significant two-way interaction, $F(10,100) = 3.13, p < 0.0016$. The interaction reflects the finding that perception of chirps in the duplex condition was more sensitive to mask intensity than perception of either speech or perception of chirps in the case of isolated formant transitions. However, this difference seemed to rest mainly on the decrease in performance with respect to the baseline condition. When the analysis was repeated with the baseline condition omitted, the two main effects remained reliable, but the interaction was no longer significant $F(8,80) = 1.46, p > 0.18$. In a final analysis, only the speech and isolated chirp conditions were compared. No effect reached significance, although the main effect of mask intensity was close, $F(5,50) = 2.27, p < 0.0618$.

The observation that masking has more of an effect on chirp perception in the duplex condition than on the perception of chirps heard in isolation would seem to pose a problem for a naive masking accounts. We might avoid this problem by postulating an interaction between selective attention and the effects of the noise. Further discussion of this result is deferred until the completion of experiment V, which asks about the generality of this result.

V. EXPERIMENT V

Experiment IV yielded the unexpected result that any differential effects of the ipsilateral mask on chirp vs. speech perception are peculiar to the duplex listening condition. To obtain further confirmation of a distinction between chirp perception in and out of the duplex condition, we conducted a final experiment which asked whether another type of acoustic manipulation—decreased amplitude—has a similar effect on the perception of isolated and duplex chirps. In designing this experiment we decided to improve upon our previous methodology by using duplex "foils" in addition to the duplex stimuli (Repp and Bentin, 1984). In duplex foils, the "base" is replaced by a full syllable which contains a third formant transition that is the opposite of the transition being presented of the other ear. The presence of such foils is

designed to discourage listeners from using a correlation between speech and chirp percepts to improve their performance on either task.

The literature on speech perception in the duplex condition leads us to expect a high degree of tolerance of decreased amplitude. In his initial description of the duplex phenomenon, Rand (1974) reported that a 30 dB attenuation of isolated transitions of synthetic /ba/, /da/, and /ga/ did not impair correct labeling of the speech side of duplex percepts. Even with 50-dB attenuation of the transitions, labeling performance was still above chance (see also Cutting, 1976). However, no results have been reported regarding the effect of transition intensity attenuation on the chirp side of duplex percepts.

A. Methods

1. Subjects

The subjects were drawn from a pool of 25 undergraduates who served as paid volunteers. They included nine young men and women who had passed a selection criterion of better than 96% correct identification of the syllables and chirps that comprise the duplex stimuli (see below).

2. Stimuli

The /da/ and /ga/ syllables that formed the basis for the stimuli were similar to those employed in experiment III and IV with the following differences. (1) The $F3$ transitions were 50 ms long. (2) The $F3$ transition for /ga/ began at 2018 and rose to 2527 Hz. (3) The base contained a steady-state $F3$ at 2527 Hz. Six different intensity levels of the transitions were created by digital attenuation in 6-dB steps, resulting in levels of -12 to -42 dB relative to the base. The foils consisted of full /da/ and /ga/ syllables on one channel (the same as the base) paired with transitions cuing the opposite phonetic category on the other channel, with the different stimulus intensities occurring once each.

3. Procedure

To ensure a high level of performance, subjects were screened for 96% or better speech labeling ability prior to the speech test, and for 96% or better chirp labeling ability prior to the chirp labeling tests. Each subject began the speech session by listening to a series of full /da/ and /ga/ syllables, presented to the left ear at the intensity they were to have in the duplex condition. Five alternations of the two syllables were presented first, followed by five blocks of 24 stimuli in random order. Subjects were instructed to write down 'd' or 'g' for each syllable, guessing if necessary; those who were confident of their performance were allowed to stop the pretest after two blocks. Subjects who made few or no errors on the first two blocks proceeded to the next condition, those who made an average of four or more errors per block were required to complete any remaining blocks. Of the subjects who completed all five blocks, those who averaged five or more errors per block were excused from further participation. Nine subjects had to be excused at this point. Two additional subjects passed the initial screening but were subsequently unable to label the full syllables and duplex syllables in the experiment proper; their data were also discarded.

The nonspeech session began with a screening for the ability to distinguish the two chirps when the transitions were heard in isolation. The two transitions were first presented five times in alternation to illustrate the two categories of high and low (forcing subjects to use apparent pitch as a criterion) followed by five blocks of 24 transitions presented in random order to the right ear. The subjects were instructed to write down H or L, guessing if necessary, and to complete at least two blocks. Criteria for passing subjects were the same as for the syllables, and four additional subjects had to be excused at this stage. One further subject's data were discarded because of random responding during the test phase.

In contrast to experiment IV, the effect of transition intensity was tested with a random-presentation design. The stimuli were presented in five blocks of 48 stimuli, in which each block contained 36 true duplex stimuli (18 /da/ and 18 /ga/) and 12 foils (6 /da/ transitions presented with the full /ga/ syllable in the contralateral ear and 6 /ga/ transitions presented with the full /da/ syllable in the contralateral ear). The six intensity levels were used equally often with each stimulus type in each block. Thus, at each intensity level there were six duplex trials (three /da/ and three /ga/) and two foils (one /da/ and one /ga/). Across five blocks, there were 15 /da/ and 15 /ga/ transitions in duplex trials at each intensity.

The speech test session was presented immediately after the training and selection procedure. The subjects were required to listen to the ear receiving the base and to label the stimuli as beginning with 'd' or 'g', guessing as necessary. They were told to ignore any sounds that occurred in the other ear. The isolated chirp identification followed.

In the isolated chirp identification session, subjects listened to the same duplex tapes, but with the "base" channel disconnected. Because in this condition there could be no foils, subjects heard a total of 20 /da/ transition and 20 /ga/ transitions at each intensity. Since transitions at the lowest intensities were close to inaudible, a visual indication of each trial was presented to the subjects by a small light triggered by the onset of a stimulus on the left (disconnected) channel. Subjects were required to write H or L whenever the light flashed, whether or not they had heard anything. After completing identification of the isolated transitions, subjects listened to the duplex tape for a third time, with both output channels again connected to the earphones. The task was to identify the chirps in the right ear while ignoring the syllables in the left ear, guessing if necessary.

B. Results and discussion

The accuracy of chirp and speech identification are graphed as a function of relative transition intensity in Fig. 5. The left panel contains the average percentage correct responses for speech identification, the right panel contains those for chirps. Chirp identification of isolated transitions is separated from that in the normal duplex stimuli, and in both panels, responses to duplex stimuli are separated from the responses to foil stimuli. For speech, the responses to foil

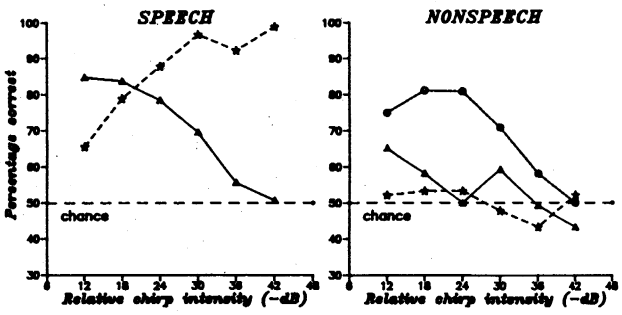

FIG. 5. The effect of transition amplitude on the speech and nonspeech side of duplex percepts, compared with its effect on the speech and nonspeech side of duplex foils and on the perception of isolated transitions as chirps. O—O, isolated chirps; *- - -*, foils; ▲—▲, duplex.

stimuli have been graphed in terms of the accuracy with which subjects labeled with the full syllable that replaced the base. For chirps, they have been graphed with respect to the accuracy with which subjects labeled the transition that was presented contralateral to the syllable.

Identification of the isolated chirps was quite good, and comparable to the identification of speech in the duplex condition. Apart from an unexplained decrease in the accuracy of chirp identification at the highest intensity, these two functions are practically identical, both reaching chance only at the lowest intensity value. In contrast, the identification of chirps in the duplex condition was adversely affected by the first step of attenuation. This pattern of results parallels that observed in experiment IV, and was confirmed by an ANOVA, which included all three conditions but omitted foil trials. That analysis revealed a main effect of condition (duplex speech versus isolated chirp versus duplex chirp), $F(2,16) = 12.61, p < 0.0005$, and an interaction with transition amplitude, $F(10,80) = 4.92$, $p < 0.0001$. Both effects were obviously due to the difference between the duplex chirp condition and the other two conditions. The main effect of transition amplitude was also significant. A separate analysis of variance, comparing identification of chirps in isolation to the identification of speech in duplex, revealed a significant main effect of transition attenuation, $F(5,40) = 59.80$, $p < 0.001$, but no main effect of stimulus condition and no interaction.

Comparison of the results obtained with the duplex stimuli and those obtained with the foil stimuli offers insight into whether subjects were using speech percepts to bolster chirp identification or vice-versa. The speech identification functions show that identification of the full syllables in the duplex foils is almost a mirror image of that of the true duplex stimuli. It is evident that, when the transition was at a very low intensity, the speech side of the foil stimuli was identified in accordance with the transition in the full-syllable "base" but that, as the transitions that were presented to the opposite ear were increased in amplitude, they competed with those in the full syllables and thereby lowered the accuracy of foil identification.

The nonspeech functions in the right panel depart from this pattern in that both the function for the duplex chirp perception and the function for the foil chirp perception hover much closer to chance. Responses to chirps in the foil trials were at chance throughout, whereas responses in the duplex trials rose above chance only at the highest level of chirp intensity. This interaction approached significance in a separate test, $F(5,40) = 2.36, p < 0.06$, but there was no significant main effect of transition amplitude for these two conditions. This lack of interaction makes it unlikely that subjects used the speech side of duplex percepts to improve their identification of the chirp side, for had this occurred, performance on the foil chirp identification trials would have been below chance at the lowest intensities. That is because the speech percepts in those cases were based on the full syllable, which contained the opposite transition of that presented to the other ear.

VI. GENERAL DISCUSSION

This series of experiments addressed the possibility that, if speech and nonspeech perception are dissociable modes of perception, it should be possible to show differential effects on each side of the duplex perception phenomenon. Since previous studies had successfully shown that phonetic manipulations affect the speech side of duplex percepts but not the chirp side, we wondered whether it would be possible to find acoustic manipulations that would affect the chirp side but not the speech side. Our venture has been successful, for we have been able to show that both ipsilateral backward masking of the transition and decreasing transition intensity had more of a disruptive effect on the chirp side than on the speech side. In contrast, both a temporal asynchrony between base and transition and a mask that is ipsilateral to the base have marked effects on the speech side, but leave the chirp side unimpaired. Thus, we obtain evidence of a double dissociation, which confirms the separability of the two forms of perception, and offers new support to previous suggestions that speech and nonspeech auditory stimuli are processed differently (Cutting and Pisoni, 1978; Liberman and Mattingly, 1985).

The result that remains to be discussed at this point is our somewhat unexpected finding that, whereas speech perception has been much the same when the base and the transition are dichotically presented in the duplex paradigm as when they are electronically fused and presented under normal binaural listening conditions (see Liberman *et al.*, 1981; Mann and Liberman, 1983), we have obtained considerable evidence that chirp perception in the duplex condition is more vulnerable than chirp perception in the case of isolated transitions. This result was anticipated by some data in Mann and Liberman (1983), which showed that the accuracy of chirp identification was considerably worse in the duplex condition than in the case of isolated transitions. It came to light in experiment IV where we found that the effects of ipsilateral masking on chirp perception were markedly greater in the duplex condition than for isolated transitions. The greater vulnerability of chirp perception was also seen in experiment V, where we found that a decrease in transition amplitude had a comsiderably greater effect on chirp perception in the duplex condition than on speech perception, and also a greater effect than on the perception of isolated chirps.

In seeking an explanation of why the duplex condition impairs chirp perception, there are several lines of reasoning to consider. We can discount a simple view that chirp perception is more vulnerable to the effects of stimulus degradation than is speech perception. Apparently, the results of experiments II, IV, and V, which showed that stimulus degradation penalized chirp perception in the duplex condition, support a view that speech perception is inherently more sensitive to marginal auditory information (possibly because the categories are better defined, more familiar, etc.). However, this account goes against the results of experiment III, which showed that speech perception was penalized when the white noise occurred in the same ear as the base, whereas chirp perception was not. In addition, it goes against the results obtained with the isolated transitions in experiments IV and V. Those results suggest that any greater susceptibility of chirp perception is somehow peculiar to the duplex condition.

Another potential account involves the concept of masking. In experiment I, we could argue that the base somehow masks the chirp. This could explain our result that, when presentation of the base is delayed, chirp perception is improved. It might further explain the finding that, in experiment V, chirp perception in the duplex condition was more vulnerable than either duplex speech perception or isolated chirp perception to the effects of decreasing transition amplitude. Could masking also explain the results of experiments II through IV? In considering a masking account, we would have to begin by postulating that the mask employed in our experiments was not a peripheral one in the traditional sense, since a peripheral mask would be expected to influence all subsequent processing—speech and chirp alike. If a masking explanation is to succeed in these cases, differences in vulnerability to masking might, for example, reflect the greater length and complexity of the base versus the isolated chirp. If length and complexity are the important factors, then one should obtain similar results when the base is replaced by some complex nonspeech sound. Note, however, that a masking account of this sort fails to acknowledge that the transition is perceived in two different manners, only one of which is masked. This latter problem might be avoided by hypothesizing that speech perception of the integrated base and transition masks chirp perception of the isolated transition, but this would pose some conceptual problems. How can a stimulus mask itself?

We had previously mentioned the possibility that a stream-specific mask might have operated in experiment III. Let us now turn to considering a scene analysis account (Bregman, 1978, 1981). Our results regarding the effects of the white-noise mask are compatible with an account in which masking is greatest when the white noise occurs in the same auditory stream as the masked stimulus: The chirp is masked by a white noise that occurs in its stream and the speech is masked by a white noise that occurs in its stream. If stream-specific masking is the correct way to interpret the results of experiment III, then we would predict that both speech and chirp perception would be released from masking when the mask occurred in both ears simultaneously. That is because binaural presentation of the mask would cause the mask to be attributed to a separate stream.

However, the scene analysis account, which we discussed above, assumes that duplex perception results from the transition being incorporated into two separate streams, one based on common location and one based on similar acoustic structure (i.e., common fundamental frequency contour). This would predict that dichotic presentation and similar acoustic structure are essential for the duplex phenomenon. Such a prediction is refuted by the demonstration of Whalen and Liberman (1987), who have shown that duplex perception is possible when neither constraint is met. In their study, listeners reported duplex percepts when a base and a sine-wave analog of a third-formant transition were presented to the same ear. There was neither a common $F0$ nor a spatial separation of chirp and transition, yet duplex percepts occurred when the transition was amplified to a level in excess of that which normally occurs.

We might also consider an account based on the differences in the ease of selective attention to the speech and chirp percepts. A selective attention account might go as follows: In the duplex phenomenon, listeners need to attend to one percept and ignore the other. Chirp perception is disrupted by the duplex condition because speech percepts are difficult to ignore, but the converse is not true, or else the duplex condition should penalize both speech and nonspeech perception. Experiments I, II, IV, and V are consistent with a view that the presence of a speech percept impairs the accuracy of chirp identification. In experiment III, we could explain the observation that the presence of the ipsilateral mask impaired chirp perception but not speech perception, and the contralateral mask had the opposite effect by postulating that it is more difficult to attend to a given percept when a white noise is heard in the same spatial location. In the end, however, although selective attention may explain many of the present results, it leaves us with the basic question of why speech should be harder to ignore.

One might ascribe the fact that speech is harder to ignore to the fact that the speech percepts reflect the louder and longer portion of the stimuli, in which case one would predict that an equally complex nonspeech sound would have the same effect on chirp perception, and this could, of course, be tested. One could also resort to the fact that speech categories are better defined, although our training and screening procedures attempted to control for this possibility, as evidenced by the high level of performance in identification of isolated chirps. One might even regard speech as "harder to ignore" because it is mediated by a higher level of processing that has some privileged access to conscious introspection. The present results, for example, might be analogous to the word superiority effect where written words are more tolerant of masking, etc. than individual letters because they involve a different level of processing (see Johnson, 1975 or 1977 for discussion of this effect). But the analogy between our results and the word superiority effect is questioned by one basic difference between the perception of letters in words and the perception of chirps in speech. Under normal reading conditions, when no mask is present and when stimulus duration is not severely limited, the letters that compose words are just as readily discernible as isolated

letters. Yet under normal listening conditions, when speech signals are presented to the two ears at a reasonable intensity, listeners perceive formant transitions as part of the speech signal but fail to perceive them as chirps. It takes some rather bizarre manipulation of the stimulus—such as the duplex condition—to make a listener hear the formant transition in speech signals as chirps.

An account that we favor for its ability to explain this and other data in the speech perception literature is prompted by the view that, in the duplex listening condition, the presence of a speech percept takes precedence over chirp perception and serves to decrease the accuracy of chirp identification. This view has been set forth in several recent papers by Liberman and his colleagues (Liberman and Mattingly, 1985; Mattingly and Liberman, in press; Whalen and Liberman, 1987), and is based on a variety of evidence that the speech processing system (the phonetic module) has priority over nonspeech systems. This priority accounts for the basic observation that began this paper, namely that, under normal listening conditions, formant transitions that are interpreted as support for one phonetic percept or another are not heard as nonspeech chirps. It further accounts for the listeners' ability to hear transitions as chirps under certain duplex conditions: when they are amplified to some level in excess of that which normally occurs (Whalen and Liberman, 1987), and when the stimulus configuration presents certain discrepancies in fundamental frequency, spatial location, etc. Such conditions allow the listener to infer that two sources have contributed to the pattern (Mattingly and Liberman, in press).

In the stimuli that this study employed, duplex perception of formant transitions was induced by separating the second or third formant transition from the remainder of the syllable and presenting it to the other ear, a maneuver that makes the chirps appear to "float" away from the speech, and to have arisen from a separate spatial location (see Nygaard and Eimas, 1990, for discussion of factors that influence the location of the chirp percept). This artificial separation of transition and base leads listeners to perceive two different events—a spoken syllable coming from one source, a chirp from another. However, there is a precedence of speech perception over nonspeech, such that the presence of the base supports a speech percept of the transitions and listeners behave as if the speech processor's interpretation of those transitions as part of a speech stimulus somehow interferes with their interpretation of the same transitions as nonspeech chirps. When the transitions are presented in isolation, there is no speech percept, precedence does not operate, and subjects achieve a higher level of performance.

Whalen and Liberman (1987) suggested that this interference exists because the speech processor operates first and subtracts energy from the signal, leaving only a residue for the nonspeech processor. However, other mechanisms (as yet unidentified) may cause this interference. What is important is that, with the precedence account of the effect of the duplex condition on chirp perception, it seems reasonable enough that manipulations that alter chirp perception in the duplex condition need not have an effect on the perception of chirps, which are heard in the case of isolated formant transitions. This was the result obtained in experiments IV and V. In experiment V, the manipulations of amplitude are a particularly appropriate test of the precedence hypothesis, for amplitude manipulations were the same means by which Whalen and Liberman (1987) were first able to demonstrate the precedence of speech in monaurally-presented stimuli. Their experiment manipulated the relative amplitude of a sinusoidal third formant transition, and showed that when the amplitude of that transition was between 0 and -7 dB relative to the third formant of a synthetic speech base, listeners heard both a whistle (i.e., chirp) and the appropriate speech percept of 'da' or 'ga'. As the amplitude of the transition was decreased, the nonspeech percept vanished, yet the speech percept continued to make use of transitions—even those that were 20 dB below the level of the base. Our stimuli are considerably different from theirs, yet we obtain a similar result: As can be seen in Fig. 5, duplex chirp perception approaches chance as transition amplitude drops more than 6 dB below that of the original transition amplitude of -12 dB, whereas speech perception is still 70% accurate at as transition amplitude that is -18 dB below the original level.

The present data are consistent with Liberman and his colleagues' hypothesis about the (biological) "specialness" of the speech processor and its priority over other forms of perception. They can even offer some insight into the conditions under which that priority is effective, if we consider the results of experiments I, II, and III more carefully. In those experiments, it appears that chirp identification is most accurate when the base and transition are asynchronous, that is, when integration of the base and transition is disrupted. There is also some indication that the effects of an ipsilateral mask are reduced when asynchrony is present, though this result is less clear. Temporal asynchrony does not prevent the listener from hearing speech, it merely disrupts the likelihood that the transition will be incorporated into the speech percept. Apparently the mere presence of speech perception is not as critical as is competition for the same acoustical information. Hence, we suggest that speech perception takes precedence when the speech and nonspeech modes are competing for interpretation of one and the same transition.

ACKNOWLEDGMENTS

The order of authorship is alphabetic, to reflect the joint contribution of the authors. The study was conducted while SB was a visiting investigator and VM was a research associate at Haskins Laboratories; it was supported by NICHD Grant HD-01994 and BRS grant RR-05596 to Haskins and by Bryn Mawr College. We thank Alvin Liberman for the valuable advice and continuous encouragement that he gave to this project, Bruno Repp for his help, and Tracy Mouk and Hwei-Bing Lin for their assistance in conducting experiments.

Bregman, A. S. (1978). "The formation of auditory streams," in *Attention and Performance VII*, edited by J. Requin (Earlbaum, Hillsdale, NJ).

Bregman, A. S. (1981). "Asking about the 'What for' question in auditory perception." In *Perceptual Organization*, edited by M. Kubovy and J. R. Pomerantz (Earlbaum, Hillsdale, NJ).

Cutting, J. E. (1976). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psycholog. Rev. **83**, 114–140.

Cutting, J. E., and Pisoni, D. B. (1978). "An information processing approach to speech perception," in *Speech and Language in the Laboratory School and Clinic*, edited by J. F. Kavanagh and W. Strange (MIT, Cambridge, MA), pp. 38–72.

Johnson, N. F. (1975). "On the function of letters in word identification: Some data and a preliminary mode," J. Verb. Learn. Verb. Behav. **14**, 17–29.

Johnson, N. F. (1977). "A pattern unit model of word identification," in *Basic Processes in Reading: Perception and Comprehension*, edited by D. LaBerge and S. J. Samuels (Earlbaum, Hillsdale, NJ).

Liberman, A. M. (1979). "Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language," in *Proceedings of the Ninth International Congress of Phonetic Sciences*, edited by E. Fischer-Jergensen, J. Rischel, and N. Thorsen (University of Copenhagen, Copenhagen), Vol. 2, pp. 468–473.

Liberman, A. M., Isenberg, D. Rakerd, B. (1981). "Duplex perception of cues for stop consonants: Evidence for phonetic mode," Percept. Psychophys. **30**, 133–143.

Liberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Mann, V. A. (1986). "Distinguishing universal and language specific levels of speech perception: Evidence from Japanese perception of /l/ and /r/," Cognition **15**, 169–196.

Mann, V. A., and Liberman, A. M. (1983). "Some differences between phonetic and auditory modes of perception," Cognition **14**, 211–235.

Massaro, D. W. (1970). "Pre-perceptual auditory images," J. Exp. Psychol. **85**, 411–417.

Mattingly, I. G., and Liberman, A. M. (**in press**). "Speech and other auditory modules," in *Signal and Sense: Local and Global Order in Perceptual Maps*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York).

Mattingly, I. G., Liberman, A. M. Syrdal, A. M., and Halwes, T. (1971). "discrimination in speech and nonspeech modes," Cognitive Psychol. **2**, 131–157.

Nygaard, L. C., And Eimas, P. D. (1990). "A new version of duplex perception: Evidence for phonetic and nonphonetic fusion," J. Acoust. Soc. Am. **88**, 75–86.

Nusbaum, H. C., Schwab, E. C., and Sawusch, J. R. (1983). "The role of 'chirp' identification in duplex perception," Percept. Psychophys. **33**, 469–474.

Pilch, H. (1979). "Auditory phonetics," Word **29**, 148–160.

Rand, T. C. (1974). "Dichotic release from masking for speech," J. Acoust. Soc. Am. **55**, 678–680.

Repp, B. H. (1981). "Two strategies in fricative discrimination," Percept. Psychophys. **30**, 217–227.

Repp, B. H., and Bentin, S. (1984). "Parameters of spectral/temporal fusion in speech perception," Percept. Psychophys. **36**, 523–530.

Repp, B. H., Milburn, C., and Ashkenas, J. (1983). "Duplex perception: confirmation of fusion," Percept. Psychophys. **33**, 333–337.

Whalen, D., and Liberman, A. M. (1987). "Speech perception takes precedence over nonspeech," Science **237**, 169–171.