The effects of auditory streaming
on duplex perception

VALTER CIOCCA and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

When a formant transition and the remainder of a syllable are presented to subjects' opposite ears, most subjects perceive two simultaneous sounds: a syllable and a nonspeech chirp. It has been demonstrated that, when the remainder of the syllable (base) is kept unchanged, the identity of the perceived syllable will depend on the kind of transition presented at the opposite ear. This phenomenon, called *duplex perception*, has been interpreted as the result of the independent operation of two perceptual systems or modes, the phonetic and the auditory mode. In the present experiments, listeners were required to identify and discriminate such duplex syllables. In some conditions, the isolated transition was embedded in a temporal sequence of capturing transitions sent to the same ear. This streaming procedure significantly weakened the contribution of the transition to the perceived phonetic identity of the syllable. It is likely that the sequential integration of the isolated transition into a sequence of capturing transitions affected its fusion with the contralateral base. This finding contrasts with the idea that the auditory and phonetic processes are operating independently of each other. The capturing effect seems to be more consistent with the hypothesis that duplex perception occurs in the presence of conflicting cues for the segregation and the integration of the isolated transition with the base.

When the second formant transition of a synthetic two-formant syllable is presented to one ear (isolated transition), and the remaining part of the syllable is presented to the other ear (base stimulus), most listeners perceive a complete syllable at the ear of the base stimulus and, at the same time, a brief nonspeech sound at the opposite ear. Yet even though the isolated sound is perceived as a nonspeech "chirp," its spectral changes over time (for example, a rising vs. a falling transition) determine the perceptual identity of the syllable at the contralateral ear. This phenomenon, known as *duplex perception*, was discovered by Rand in 1974. Duplex perception has also been shown to occur with three-formant syllables in which the third (or the second and third) formant transitions are isolated from the rest of the syllable and played to the contralateral ear (Liberman, 1982; Liberman, Isenberg, & Rakerd, 1981).

In duplex perception, while the isolated transition is integrated into the contralateral base, it is perceived as a nonspeech sound at the same time. This has been interpreted as the result of the operation of two independent systems that extract information from the isolated transition: a phonetic and an auditory mode (Liberman, 1982). This hypothesis, which we call the *independence* hypothesis, implies that the phonetic mode takes from the acoustic signal the information that it needs, independently of how the same signal may be processed by auditory organization processes. In other words, the independence hypothesis suggests that the double allocation of the isolated transition to two percepts is due to the independent operations of a speech system and a general-purpose, auditory-recognition system.

The present study was made with the purpose of testing the independence hypothesis. To do this, the isolated formant transition of a duplex syllable was embedded in a sequence of formant transitions that preceded and followed it in the same ear (see Figure 1). This procedure was expected to produce the sequential streaming of the isolated transition into a sequence of transitions, based on the Gestalt principles of spatial and frequency proximity (Wertheimer, 1923). Indeed, Steiger and Bregman (1982) had found that the dichotic fusion of two contralateral, simultaneous, complex tones was weakened by embedding one of them (the target tone) into a sequence of captor tones at the same ear; this is called the *capturing effect*. In that study, the captor tone had a spectral content similar to that of the target tone, and both tones were presented at the same ear. We thought that the capturing effect produced by such a streaming procedure might also be obtained with speech stimuli such as the duplex syllables. The presence of a sequence of identical transitions in the same ear might be interpreted as a cue for the assignment of the isolated transition to the sequence of streaming transitions.

The occurrence of a capturing effect with duplex stimuli would constitute direct evidence against the independence explanation of the duplex phenomenon. Because the independence hypothesis states that the double allocation
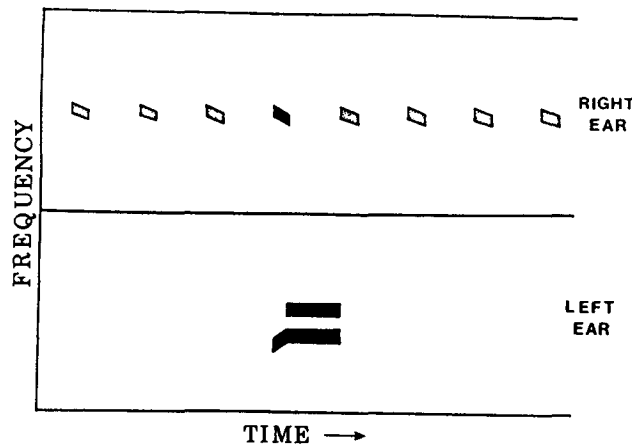
Figure 1. The experimental pattern used in the streaming conditions of Experiments 1 and 2. The duplex stimulus (black color) is presented with the base stimulus to the left ear and the isolated transition to the right ear. The capturing transitions (lighter color) are presented to the right ear before and after each of the isolated transitions of a duplex stimulus.

of the transition is the result of the independent processing of the speech and auditory systems, the integration of the isolated transition with the base by the speech system should not be significantly affected by the auditory organization of the same information into a sequence of streaming transitions. In other words, according to the independence hypothesis, the streaming procedure should not produce a significant weakening of the integration of the isolated transition with the base.

Four experiments were conducted. In Experiment 1, subjects were asked to identify duplex syllables either in isolation or under the streaming condition. Two sets of syllables were employed, two-formant /ba/-/da/ and three-formant /da/-/ga/ syllables. In Experiment 2, each of the stimuli in the previous experiment was followed by two monaural syllables, which were presented for the purposes of comparison. The subjects had to judge whether the target stimuli were more similar to the first or second comparison syllable.

EXPERIMENT 1

In this experiment, the identification of the monaural and duplex syllables was measured under various conditions. Two sets of synthetic stimuli (/ba/-/da/ and /da/-/ga/ syllables) were used in different parts of the experiment. The experiment comprised two sessions. In the first session, the experimental stimuli were identified in isolation (isolation conditions). In the second session, the isolated transition of duplex syllables was always preceded and followed by a sequence of identical or different frequency transitions (streaming conditions). We tested whether it was possible to weaken the integration of the isolated transition with the base by means of capturing the former into a sequence of identical transitions (see Figure 1). This was called the *capturing* condition. The occurrence of a capturing effect should have caused the

perception of the speech percept to become more as it would have been, had the isolated transition not been presented at all. In a control (*no-capturing*) condition, the transitions that preceded and followed the target transition were outside the frequency region of the target. This condition was intended to demonstrate that the possible occurrence of a capturing effect was not due to the mere presence of a sequence of transitions at the ear at which the isolated transition was presented.

Method

Subjects. Different pools of subjects were employed for the two sets of syllables. All were undergraduate students, graduate students, or staff members in the Psychology Department at McGill University, who were paid for their services. All the subjects received an audiometric test. A criterion was defined in order to establish whether each subject properly identified the speech percept of the duplex stimuli. The criterion was defined as 60% correct identification with the duplex stimuli or a rating-scale score of 6.0 or more with the duplex /da/ or /ga/ stimuli.

Eighteen subjects were tested with the /ba/-/da/ syllables. All the subjects except 1 had normal hearing. This subject had a 40-dB loss in the left ear at 8 kHz. His data and the data of another subject were discarded, because both subjects consistently misperceived the base stimulus as /da/ (0% and 10%, respectively). The data of 2 more subjects were discarded because these subjects did not reach the criterion of correct identification of the speech percept.

Seventeen subjects were tested with the /da/-/ga/ stimuli. The data of 2 subjects were discarded because these subjects could not consistently identify the /da/-/ga/ stimuli. The data of one other subject were not considered in the analysis, because she failed to reach criterion for the speech percept of the duplex /ga/ stimulus.

Stimuli. The /ba/-/da/ stimuli were two-formant syllables digitally synthesized on a computer with a parallel formant circuit. The syllables had a duration of 200 msec. The formant transitions, 40 msec long, were followed by 160-msec steady-state formants. The steady portions of F1 and F2 were centered at 649 and 1240 Hz, respectively. F1 was identical for both /ba/ and /da/ and rose from 310 to 649 Hz. F2 rose from 740 to 1240 Hz for /ba/ and fell from 1540 to 1240 Hz for /da/. The durations of the rise and decay of amplitude were 1 and 150 msec, respectively. These amplitude changes followed a quarter of a sine wave function. The duplex syllables were generated by presenting the second formant transition (isolated transition) to the right ear. The remaining part of F2 and the whole F1 (base) were presented to the left ear. In the duplex stimuli, the isolated transition had a 30-msec decay appended at its end, and F2 in the opposite ear had a 30-msec rise time. This was done in order to improve the fusion between the isolated transition and the base. The bandwidths of the formant filters were 60 Hz for F1 and 100 Hz for F2. The fundamental frequency (F0) was steady at 120 Hz. The base stimuli and the monaural syllables were presented at a level of 72 dB SPL.

In the capturing condition (see Figure 1), the streaming transitions were identical to those of the duplex syllables. In the no-capturing condition, the added sequence of transitions had a center frequency varying from 1740 to 2240 Hz when presented with the duplex /ba/ and from 2540 to 2240 Hz with the duplex /da/. These values were chosen to be 1000 Hz higher than the center frequency of the isolated transitions. This was done to create a condition in which the streaming transitions did not group strongly with the isolated transitions. If this condition did not affect phonetic integration, then the possible occurrence of an effect in the capturing condition could not be accounted for by the mere presence of a sequence of transitions in the unattended ear. The intensity and duration characteristics of the streaming transitions were identical to those of the transitions of the duplex syllables.

The /da/-/ga/ stimuli were three-formant syllables synthesized on the Haskins Laboratories parallel resonance synthesizer. The formant transitions were 50 msec long, followed by 200-msec steady-state formants. The center frequency of the steady-state formants was 771 Hz for the first, 1233 Hz for the second, and 2525 Hz for the third formant. The onset frequencies were 285 Hz for the first formant and 1770 Hz for the second formant. The third formant transition rose from 2018 Hz for the /ga/ syllable and fell from 2862 Hz for the /da/ syllable. The base stimulus did not have the third formant transition. These stimuli were recorded on an audio tape and sent to our laboratory.[1] We digitized the syllables at a 10-kHz sampling rate. We deleted the last 50 msec of the signal by means of digital editing, so that the total duration of the syllables became 200 msec. A decay (quarter of a sine wave) was imposed on the final 50 msec of the shortened syllables.

In both the capturing and the no-capturing conditions, the isolated transitions were those of the original Haskins stimuli. However, the streaming transitions were synthesized at our laboratory by means of our trying to match them with the original Haskins Laboratories' transitions. This was done because we needed some transitions that were 1000 Hz higher than the isolated transitions provided on the Haskins tape, to be used in the no-capturing conditions. Moreover, we did not want to use the original Haskins transitions in the capturing conditions, because we wanted to use similar acoustic parameters for the synthesis of both the capturing and the no-capturing transitions. Matches to the Haskins transitions were made by ear and by sight (i.e., the spectra were looked at, until the isolated—the original—and the capturing transitions could not be distinguished). The no-capturing transitions were then synthesized with the same parameters, apart from the frequency range. The center frequency of the no-capturing transitions ascended from 3018 to 3525 Hz for the duplex /ga/ syllable and descended from 3862 to 3525 Hz for the /da/ syllable.

Procedure. Each set of stimuli was tested in two separate sessions. In the first session, one of the two sets of stimuli (either the /ba/-/da/ or the /da/-/ga/) was presented in the isolation condition. The experimental pattern was begun with a warning tone, followed by 1 sec of silence. After the silence, one of the five stimulus syllables was presented. It was repeated three times with a 1-sec silence between repetitions. The subjects had to identify the stimuli during 5 sec of silence after the presentation of the syllables. The full syllables and the base were always presented to the left ear; the isolated transitions were always presented to the right ear. The subjects were informed that they might hear brief sounds ("something like chirps") in the right ear, but they were told to ignore them and pay full attention to the syllables in the left ear. Each of the five experimental stimuli was presented 10 times.

The second session followed the first one, after a pause of approximately 5 min. In the second session, the streaming conditions were presented. The duplex syllables from the previous session were repeated six times in a cycle within each trial. The offset-to-onset duration between successive syllables was 1.2 sec. In the unattended ear, the transitions were repeated at a rate of one every 200 msec. Therefore, there was a pause of silence of 160 msec between successive transitions in the right ear. Six transitions were presented during the time interval between successive duplex syllables within each trial. Six transitions followed the last syllable of a trial. Each six-repetition pattern had a duration of 9.2 sec, which was expected to be sufficient for the occurrence of capturing. Indeed, it had been demonstrated that streaming mechanisms need to accumulate evidence for at least 4 sec in order to form an auditory object (Bregman, 1978). In the same study, it was found that at least 4 sec of silence were necessary in order to disrupt a capturing or segregating effect in the processing of auditory sequences. For this reason, a pause of 5 sec was introduced in the present experiment after the last syllable in a pattern. In the second session, there were four streaming conditions, two duplex syllables (/ba/-/da/ or /da/-/ga/) presented

in the capturing or the no-capturing conditions. Each streaming condition was presented 10 times within a session. The order of presentation of the experimental stimuli was randomized and differed for every subject. In both sessions, a 10-point rating scale was used to measure the subjects' identification of the speech stimuli. If subjects identified a syllable as /ba/ (/ba/-/da/ stimuli) or /da/ (/da/-/ga/ stimuli), they had to choose a number from 1 through 5; the numbers 6 through 10 indicated the perception of the other phonetic category. The extremes of the scale (1 and 10) indicated a "very confident" judgment, and numbers in the center of the scale indicated a "not confident" judgment.

Apparatus. The stimuli were digitally synthesized at a 15-kHz sampling rate with the MITSYN software package (Henke, 1975), and played by the computer during the experiment. The stimuli were low-pass filtered at 5 kHz by a Rockland Dual Filter Model 852. The subjects listened individually to the stimuli over TDH-39 headphones, in an Industrial Acoustics 1202 audiometer chamber. The intensities of the stimuli were measured with a General Radio Type 1551-C sound level meter set at A weighting and connected to the headphones with a flat-plate coupler. A Maico Hearing Instruments MA27 audiometer was used to test the subjects' hearing. A keyboard was used by the subjects to record their responses.

Results and Discussion

Isolation conditions. The subjects were quite consistent in identifying the monaural and the duplex stimuli. Data were analyzed in two ways: as percentages and as rating scores. The rating-scale scores for the /ba/-/da/ stimuli are shown in Figure 2; the mean percentages of /da/ responses are displayed in parentheses at the top of the figure. A one-way repeated-measures ANOVA was applied to the rating scores, with the five stimuli considered as the levels of a single factor. The main effect of syllable type was significant at the .0001 level [$F_{(4,52)} = 128.72$]. The results clearly show that subjects identified the syllables reliably and that they were identifying the duplex syllables according to the kind of isolated transition presented to the unattended ear. The difference in confidence ratings between the full /da/ (9.6) and the duplex /da/ (7.5) was found to be significant at the .01 level by a planned comparison.
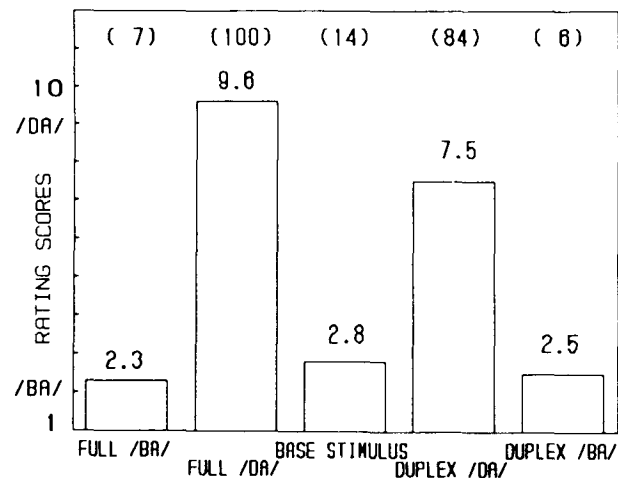

Figure 2. Mean rating scores for the /ba/-/da/ stimuli of Experiment 1 in the isolation condition. The mean percentages of /da/ responses are displayed in parentheses at the top of the figure.
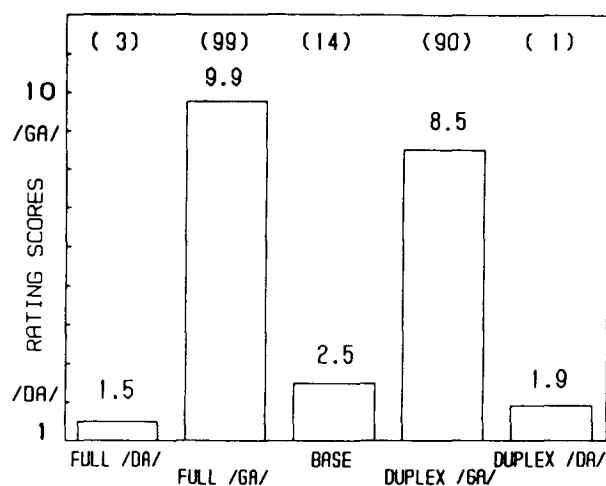
Figure 3. Mean rating scores for the isolation condition of the /da/-/ga/ stimuli in Experiment 1. The mean percentages of /ga/ responses are presented at the top of the figure.

Similar results were obtained for the /da/-/ga/ stimuli. We wanted to test whether the Haskins syllables could be reliably identified after the removal of 50 msec of the signal at the syllable offset. The average confidence ratings are shown in Figure 3. A one-way repeated-measures ANOVA was applied to the mean ratings, with syllable type as a factor. This effect was significant at the .0001 level [$F(4,52) = 206.04$]. The difference between the full and duplex /ga/ was significant at the .01 level (planned comparison), indicating that the full /ga/ was correctly identified significantly better than was the duplex /ga/. This result, as well as the above-mentioned difference between full and duplex /da/ stimuli, replicated Repp, Milburn, and Ashkenas's (1983) findings. These authors found that their duplex /ga/ stimulus, even if labelled reliably, was not correctly identified as often as the monaural /ga/.

Streaming conditions. The average rating scores and percentages of /da/ responses for the /ba/-/da/ stimuli are displayed in Figure 4. They show that there was a difference in the identification of the duplex /da/ syllable in the capturing and no-capturing conditions. A two-way ANOVA with repeated measures was applied to the mean ratings. The main effect of syllable type (/da/ vs. /ba/) was highly significant [$F(1,13) = 67.05, p < .0001$]. The main effect of capturing was also highly significant [$F(1,13) = 21.64, p < .0005$]. The interaction between capturing and syllable type was highly significant [$F(1,13) = 22.06, p < .0004$]. The duplex /da/ stimulus was significantly identified as /ba/ more often in the capturing than in the no-capturing condition. This finding shows that the integration of the isolated transition with the base was significantly weakened in the capturing condition of the experiment. The capturing factor had no significant effect on the /ba/ duplex syllable, because the base stimulus itself was identified as /ba/. Therefore, removing the effect of the /ba/ transition would not be expected to affect the subjects' choices.

The mean rating-scale scores for the /da/-/ga/ stimuli, shown in Figure 5, were also analyzed by a two-way ANOVA with repeated measures. The syllable-type (/da/ vs. /ga/) main effect was highly significant [$F(1,13) = 169.4, p < .0001$]. The capturing main effect was significant at the .0007 level [$F(1,13) = 19.41$]. The capturing × syllable interaction was also highly significant [$F(1,13) = 24.9, p < .0002$]. These results fully replicated those obtained with the /ba/-/da/ stimuli, and they show the existence of a significant capturing effect. These findings are not consistent with the hypothesis that the double allocation of the isolated transition in the duplex phenomenon is due to the independent operation of speech and auditory processes.

EXPERIMENT 2

Experiment 2 was an attempt to replicate the capturing effect obtained in the previous experiment, this time with
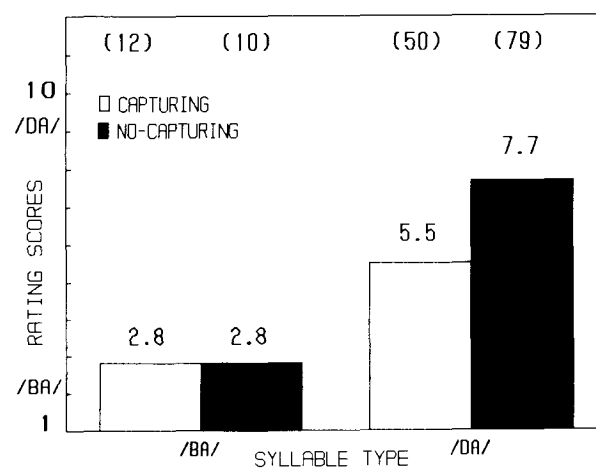

Figure 4. Mean rating scores for the streaming condition of Experiment 1 (/ba/-/da/ stimuli). The mean percentages of /da/ responses are in parentheses at the top of the figure.
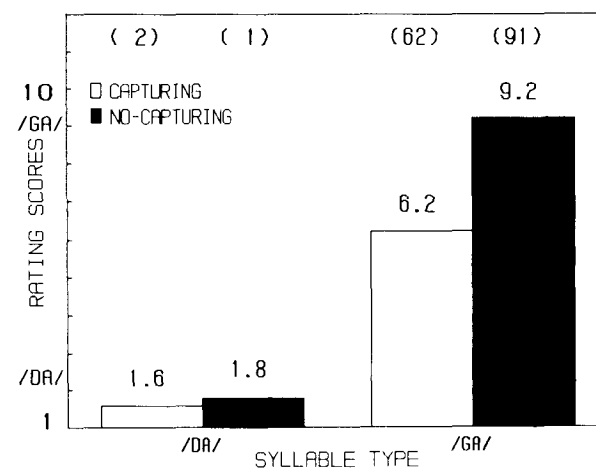

Figure 5. Mean rating scores and percentages of /ga/ responses for the streaming condition of Experiment 1 (/da/-/ga/ stimuli).

a discrimination procedure. Discrimination was measured under both the isolation and the streaming conditions. The sets of syllables (/ba/-/da/ or /da/-/ga/) used in Experiment 1 were again employed. The duplex and base stimuli were followed by two of three possible monaural comparisons: the full syllables (/ba/-/da/ or /da/-/ga/) and the base stimulus. The base stimuli were always presented to the left ear. The isolated transitions of the duplex stimuli were presented to the right ear. The comparison stimuli were always presented to the same ear as was the base.

Method

Subjects. Different pools of subjects were used for the /ba/-/da/ and the /da/-/ga/ stimuli. They were all psychology graduate and undergraduate students at McGill University, who were paid for their services. Eighteen subjects were tested with the /ba/-/da/ stimuli. The data of 2 subjects were discarded because they failed to perceive the duplex percept. The data of 2 other subjects were discarded because they consistently identified the base stimulus as more similar to the /da/ comparison.

Fifteen subjects were tested with the /da/-/ga/ stimuli. The data of 1 of them were discarded because she did not integrate the isolated transitions with the base.

Stimuli. The target and comparison stimuli were those employed in Experiment 1. As in the previous experiment, one group of subjects was tested with the /ba/-/da/ stimuli; the other group was presented with the /da/-/ga/ stimuli. The target stimuli for the /ba/-/da/ subjects were the /ba/-/da/ duplex stimuli and the respective base stimulus. The comparison stimuli were chosen from among the /ba/-/da/ monaural syllables and the respective base stimulus. For the /da/-/ga/ subjects, the target stimuli were the /da/-/ga/ duplex syllables and the relative base. The comparisons were the monaural /da/-/ga/ syllables and the base. For each set of stimuli, six pairs of comparison stimuli were obtained through taking all the pairwise combinations of the three comparisons. Therefore, a total of 18 conditions (three targets × six comparison pairs) were presented for each of the two stimulus sets.

Procedure. In the isolation condition, an XXXAB paradigm was employed. The target syllable (Sound X) was repeated three times as in Experiment 1; each time it was followed by .5 sec of silence. Next, the first comparison (Sound A) was presented, followed by 1 sec of silence and the second comparison stimulus (Sound B). During a 5-sec interval of silence following Sound B, the subjects had to judge whether Sound X was like Sound A or Sound B. A 10-point rating scale was used, as in Experiment 1. The numbers 1 through 5 indicated that the target syllable was perceived as being the same as the first comparison; the numbers 6 through 10 meant that the target syllable was more like the second comparison. The extremes of the scale (1 and 10) were labelled "very confident." Numbers 5 and 6 indicated that subjects were "not confident" in their choices. It should be stressed that this rating scale did not involve a judgment of the similarity of the target to one of the comparison stimuli. The subjects rated how confident they were in judging that the target stimuli sounded like the first or the second comparison. Each of the 18 experimental patterns was presented two times in the experiment. The order of presentation of the 36 trials in a session was randomized, and it differed for every subject.

In the streaming conditions, the experimental patterns were the same as those used in the analogous conditions of Experiment 1. However, within each trial, a pair of comparison stimuli followed the last presentation of the target syllable. The comparison stimuli were the same as in the isolation conditions. There were 24 conditions in the experiment: capturing/no-capturing conditions × syllable type (/ba/-/da/ or /da/-/ga/) × six comparison pairs. Each condition was presented twice in the experimental session. The order

of presentation of the stimuli was randomized; it differed for every subject. As in Experiment 1, the isolation conditions were presented in a first session. A second session contained the streaming conditions for the same stimulus set, and followed the first session after a pause of about 5 min.

Apparatus. This was the same as in Experiment 1

Results and Discussion

Isolation conditions. The rating scores were converted so that Ratings 1 through 5 indicated that the target syllable was like the full /ba/ or the base stimulus. Ratings 6 through 10 indicated that subjects perceived the target as similar to the full /da/ syllables. The mean rating-scores responses for the /ba/-/da/ stimuli are displayed in Figure 6. The numbers in parentheses at the top of the figure are the percentages of /da/ (/ba/-/da/ and base-/da/) comparisons) or base responses (base-/ba/ comparison). A two-way ANOVA with repeated measures was performed on the data obtained with the /ba/-/da/ and base-/da/ comparisons. The three targets were considered as levels of the syllable-type factor. The second factor was the comparison type (/ba/-/da/ and base-/da/ comparisons). The comparison-type main effect and the syllable × comparison interaction were not significant. This indicates that the comparison pairs employed did not make a difference in the judgments. The syllable-type factor was significant at the .0001 level [$F(2,26) = 59.64$]: this effect shows that the subjects discriminated the syllables reliably. Moreover, it can be seen from Figure 4 that the base stimulus was consistently judged as more similar to the full /ba/ than to the full /da/ stimulus. The base-/ba/ comparison pair was included in the experiment because we wanted to test whether the subjects could discriminate these two stimuli. This was done because these stimuli were not identified as different syllables in the previous experiment. It can be seen in Figure 6 that the subjects could not discriminate the base-/ba/ comparisons reliably when they followed the base and the /ba/ targets. This
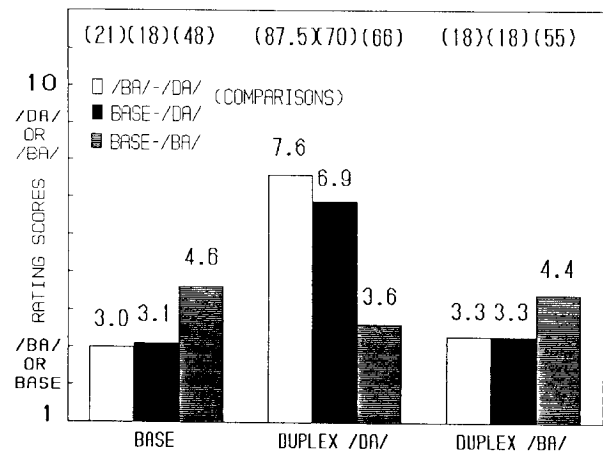

Figure 6. Mean rating scores for the /ba/-/da/ target stimuli in the isolation conditions of Experiment 2. For the /ba/-/da/ and base-/da/ comparisons, the numbers at the top of the figure represent the mean percentages of /da/ responses. For the base-/ba/ comparisons, the numbers indicate the mean percentages of base responses.

result parallels the subjects' reports of not being able to distinguish the comparison stimuli of some of the pairs. No statistical analysis was undertaken for the data for the base-/ba/ conditions, since these were not the critical ones for the purpose of the experiment.

The results for the /da/-/ga/ stimuli are illustrated in in Figure 7 as rating scores and as percentages of /ga/ or base (base-/da/ comparisons) responses. Rating scores of 1 through 5 indicate that the target syllable was perceptually similar to the full /da/ or base comparisons. A two-way repeated-measures ANOVA was performed on the rating scores for the /da/-/ga/ and base-/ga/ comparison pairs. Syllable type (/da/, /ga/, or base) and comparison type (/da/-/ga/ and base-/ga/) were the experimental factors. The syllable-type main effect was the only significant effect [$F(2,26) = 78.18, p < .0001$]. These results fully replicated those obtained with the /ba/-/da/ stimuli. The data for the base-/da/ comparisons showed that these stimuli sounded alike. For this reason, the target stimuli received similar ratings when paired with these comparison stimuli.

Streaming conditions. The data for the /ba/ stimuli are displayed in Figure 8a, and those for the /da/ stimuli in Figure 8. The /ba/ stimuli were included for the purpose of determining whether the streaming procedure caused any generalized impairment in the subjects' reliability in discriminating the speech percepts. Figure 8a shows that subjects consistently discriminated the /ba/ stimuli with the base-/ba/ and /ba/-/da/ comparison stimuli. The base-/ba/ stimuli did not seem to be perceived as different syllables. Indeed, the targets presented with them received similar, middle-scale ratings. This result was expected, since the base and /ba/ stimuli could not be discriminated reliably in the isolation conditions of this experiment. A repeated-measures two-way ANOVA was performed on the /da/ syllable with the base-/da/ and /ba/-/da/ comparison stimuli. The comparison-type
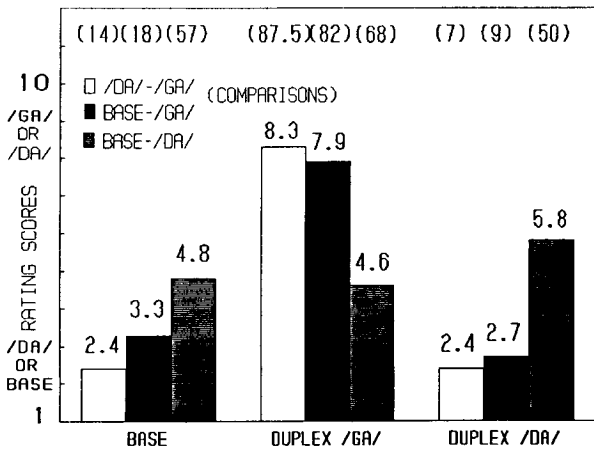

Figure 8. The mean data for the /ba/-/da/ stimuli of Experiment 2 in the streaming conditions. (a) The mean ratings percentages of base or /ba/ (/ba/-/da/ comparisons) responses for the /ba/ duplex stimuli. (b) The mean percentages of /da/ or base (base-/ba/ comparisons) responses for the /da/ duplex stimuli.

and capturing conditions were the experimental factors. This analysis showed a significant effect of capturing [$F(1,13) = 16.65, p < .001$], indicating that in that condition the duplex effect was significantly weakened. This replicates the capturing effect found in Experiment 1 with an identification task. The comparison-type and interaction effects were not found to be significant.

The mean rating scores for the /da/ stimuli are shown in Figure 9a. The /da/ duplex stimuli were consistently perceived as sounding like the base or /da/ comparison stimuli, unless they were followed by the base-/da/ pairs. This finding replicated the analogous one obtained with the /ba/-/da/ stimuli. The scores for the /ga/ stimuli are displayed in Figure 9b. A two-way ANOVA with repeated measures was applied to the data for the /ga/ stimuli with the /da/-/ga/ and base-/ga/ comparison stimuli. The experimental factors were capturing and comparison type (/da/-/ga/ vs. base-/ga/). The comparison-type main effect was significant at the .0003 level [$F(1,13) = 24.82$].

Figure 7. Mean rating-scale scores for the /da/-/ga/ target stimuli of Experiment 2 (isolation conditions). The mean percentages of /ga/ responses (in the case of the /da/-/ga/ and base-/ga/ comparisons) or base responses (in the case of the base-/da/ comparisons) are displayed at the top of the figure.
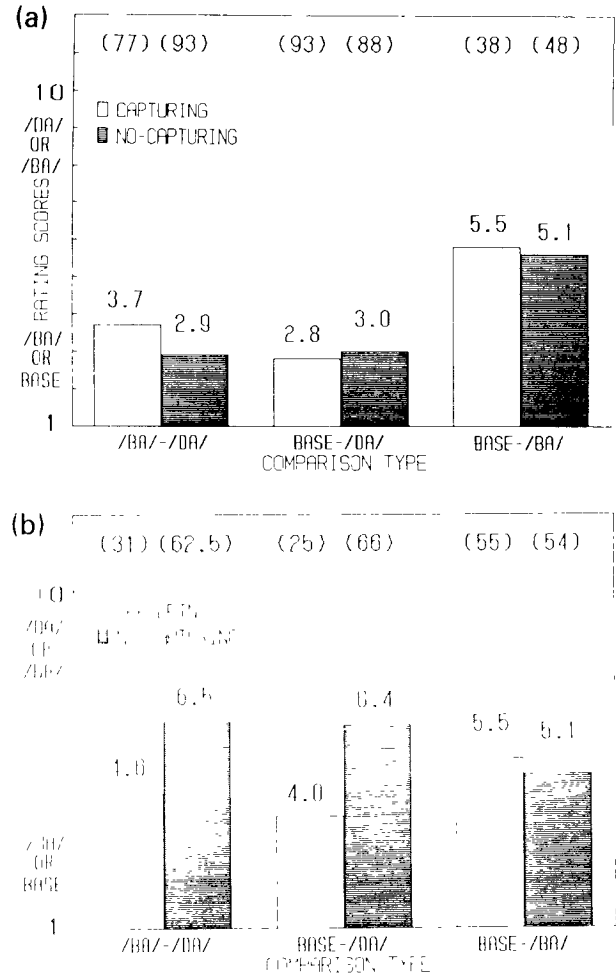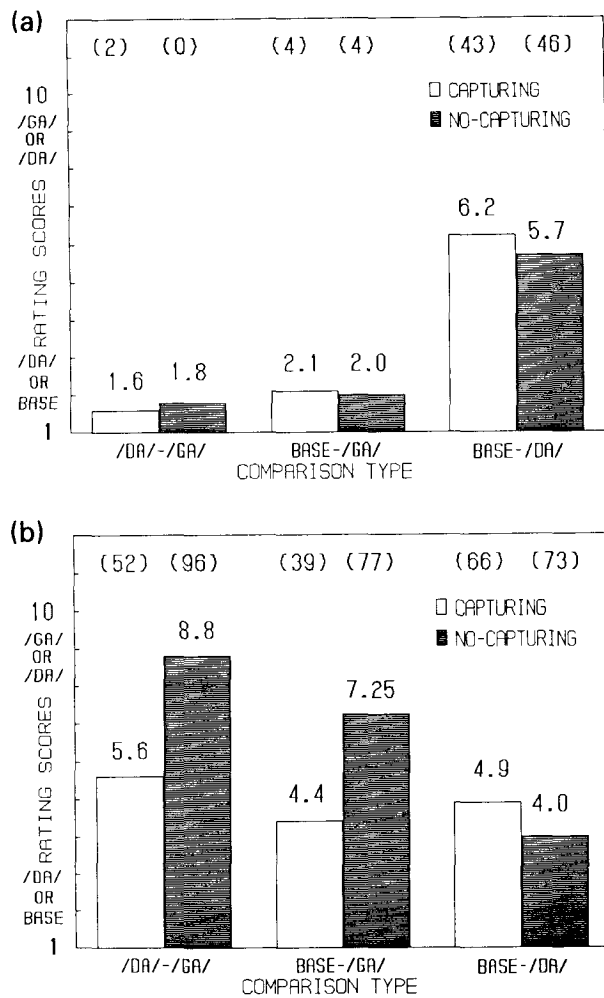
Figure 9. The mean data for the /da/-/ga/ stimuli of Experiment 2 in the streaming conditions. (a) The mean rating scores and percentages of /ga/ or base (base-/da/ comparisons) responses for the /da/ duplex syllables. (b) The mean ratings and percentages of /ga/ or base (base-/da/ comparisons) responses for the /ga/ duplex syllables.

The capturing main effect was also highly significant [$F(1,13) = 30.68$, $p < .0001$]. The interaction effect was not significant. These results show the presence of a significant capturing effect. The captured /ga/ sounded more like the /da/ or the base comparison stimuli than the noncaptured /ga/ did. The comparison-type effect shows that subjects perceived the /ga/ stimuli as more similar to the base than to the /da/ comparison overall.

EXPERIMENT 3

In the streaming conditions of the previous experiments, the capturing transitions were always identical to the critical one. Therefore, the repeated presentation of the critical transition might be suspected of having had its effect through the habituation of the phonetic detectors sensitive to it. Therefore, the observed weakening of duplex perception could have been the result of the habituation

of phonetic processes rather than the streaming of the critical transition by the auditory scene-analysis processes. It is possible that habituation might have decreased the effectiveness of the isolated transition as a cue for a stop consonant.

The occurrence of habituation could be thought of as equivalent to a decrease in the level of the perceived intensity of the isolated transition. Rand (1974) systematically varied the intensity of the isolated transition of a duplex syllable, and found that duplex perception was not weakened considerably until the level of the isolated transition was attenuated by 40 dB. At this attenuation, the isolated transition is not clearly audible as a separate sound. Therefore, the habituation explanation of our results is not really very plausible. Furthermore, in the streaming experiments, the target transition was always clearly audible. None of the subjects reported any gap in the sequence of the streaming transitions.

The preceding evidence argues against a habituation explanation of the capturing effect. Nevertheless, we wanted to collect direct evidence to make the auditory scene-analysis explanation of our results more likely. In order to do this, in Experiments 3 and 4 we included capturing conditions, in which the stream-inducing transitions were aligned along the same trajectory as the critical one on a frequency-by-time scale. In this way, the stream-inducing transitions were always in a different frequency region from that of the critical transition (see Figures 10a and 10b). Therefore, the occurrence of a significant capturing effect under such capturing conditions would be strong evidence against the habituation explanation of the capturing effect. In Experiment 3, the capturing effect was measured with the use of an identification paradigm.

Method

Subjects. Eighteen graduate and undergraduate students at McGill University were tested. All were paid for their services. The data from 1 subject were discarded, because she could not identify the syllables consistently. Another subject identified all the syllables in the experiment as /da/, because he expected that /ga/ percepts had a release burst—which was never present. His data were therefore discarded too. After this experiment was over, he was told that he should try to decide which stimuli could better fit into the judgment categories. He was then tested in the next experiment, and he reliably discriminated the speech side of the duplex percepts.

Procedure. In Experiment 3, the isolated transition (critical transition) of the duplex stimuli was preceded and followed by three transitions at the same ear (see Figures 10a and 10b). The transitions that followed the critical one were the same in both the capturing and the no-capturing conditions, and they were aligned along the same trajectory as was the critical transition. In the capturing condition, the transitions that preceded the critical one were aligned on the same trajectory as were the transitions that followed it. In the no-capturing condition, the three transitions that preceded the critical one were not aligned on the same trajectory as was the critical transition. The base stimulus was presented at the opposite ear. An experimental pattern was formed by seven transitions at the right ear and one occurrence of the base at the left ear. There was a silence of 10 msec between successive transitions. Each pattern was repeated six times in a trial, each time followed by a 1-sec silence. The offset-to-onset time of successive base stimuli was 1,210 msec. After the last pattern was presented, the subjects had 5 sec to give
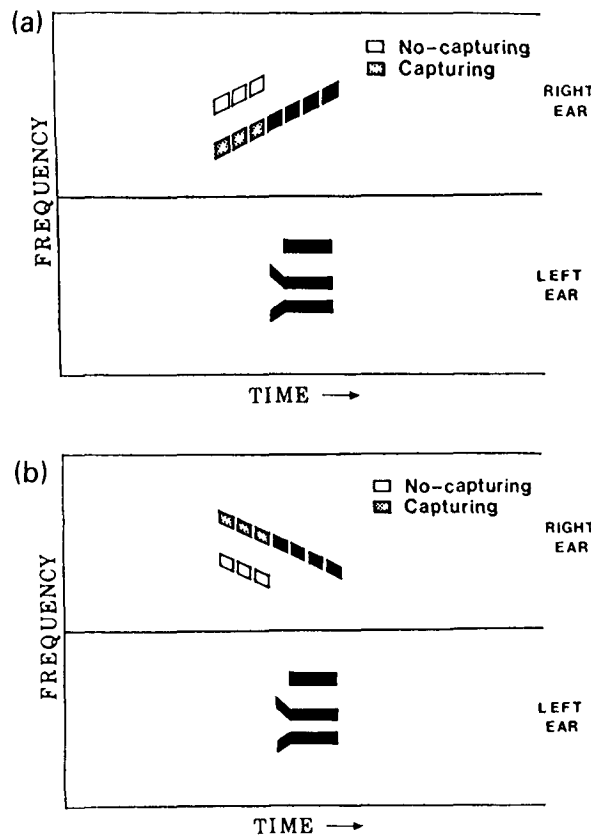
Figure 10. The experimental patterns used in Experiments 3 and 4. (a) The /ga/ capturing and no-capturing patterns. (b) The /da/ capturing and no-capturing patterns.

a response. The procedure for the identification of the syllables was the same as that of Experiment 1.

Stimuli. The duplex stimuli were the /da/-/ga/ syllables employed in Experiments 1 and 2. All the stimuli had a sampling frequency of 10 kHz and were low-pass filtered at 4.8 kHz. The capturing and no-capturing transitions were synthesized by a procedure analogous to that employed in Experiment 1 for the synthesis of the no-capturing transitions. In the capturing /ga/ condition, the center frequency of the critical transition rose from 2018 to 2525 Hz. The three transitions that preceded the critical one (capturing transitions) rose from 192 to 698 Hz, 799 to 1307 Hz, and 1409 to 1916 Hz, in that order (Figure 10a). In the no-capturing /ga/ condition, the sequence of three preceding (no-capturing) transitions ascended in a frequency region 2000 Hz higher than that of the capturing transitions. Their actual frequencies rose from 2192 to 2698 Hz, 2799 to 3307 Hz, and 3409 to 3916 Hz, respectively. The three transitions that followed the critical transition were aligned on the same trajectory as were the critical transitions in the capturing and no-capturing conditions. They rose from 2626 to 3133 Hz, 3235 to 3742 Hz, and 3843 to 4350 Hz, respectively. The reason for using the same poststimulus transitions in both experimental conditions was that all the stimuli were low-pass filtered at 4.8 kHz. If these transitions had been aligned with the no-capturing transitions in the corresponding condition, most of them would have varied in a frequency region higher than the filtering frequency.

The center frequency of the critical /da/ transition fell from 2862 to 2525 Hz. In the capturing /da/ condition, the three capturing transitions fell from 4075 to 3738 Hz, 3671 to 3334 Hz, and 3266 to

2929 Hz, respectively (Figure 10b). In the no-capturing /da/ condition, the three no-capturing transitions were lower than the critical transition in frequency, falling from 2075 to 1738 Hz, 1671 to 1334 Hz, and 1266 to 929 Hz, respectively. The three transitions that followed the critical one were aligned on the same trajectory as were the capturing /da/ transitions. They fell from 2457 to 2121 Hz, 2053 to 1716 Hz, and 1649 to 1312 Hz, respectively. Each transition had a duration of 50 msec. The intensity of the base was 72 dB SPL.

Apparatus. The apparatus was the same as in Experiment 1.

Results and Discussion

A two-way ANOVA with repeated measures was used to analyze the mean rating-scale scores (see Figure 11). The main effect of syllable type (/da/ vs. /ga/) was found to be significant at the .0001 level [$F(1,15) = 104.58$]. The main effect of capturing was also highly significant [$F(1,15) = 19.76, p < .0005$]. The syllable × capturing interaction was significant at the .002 level [$F(1,15) = 14.37$]. These results replicate the previous finding that the integration of the critical transition with the base was significantly weakened in the capturing condition. This result is incompatible with the idea that habituation caused the capturing effect observed in the earlier experiments.

EXPERIMENT 4

In Experiment 4, the capturing effect of transitions aligned along the trajectory of the critical transition was tested with the use of a discrimination task. This time, the experimental patterns used in Experiment 3 were followed by two monaural comparison stimuli.

Method

Subjects. The subjects were the same as those who participated in Experiment 3. The present experiment followed the previous one after a pause of about 5 min.

Procedure. The discrimination procedure was the same as in Experiment 2. A pair of monaural syllables chosen among the /da/,
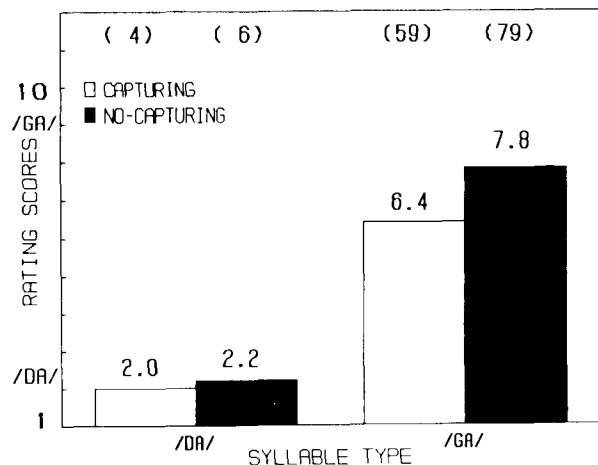

Figure 11. Mean rating scores and percentages of /ga/ responses in Experiment 3.
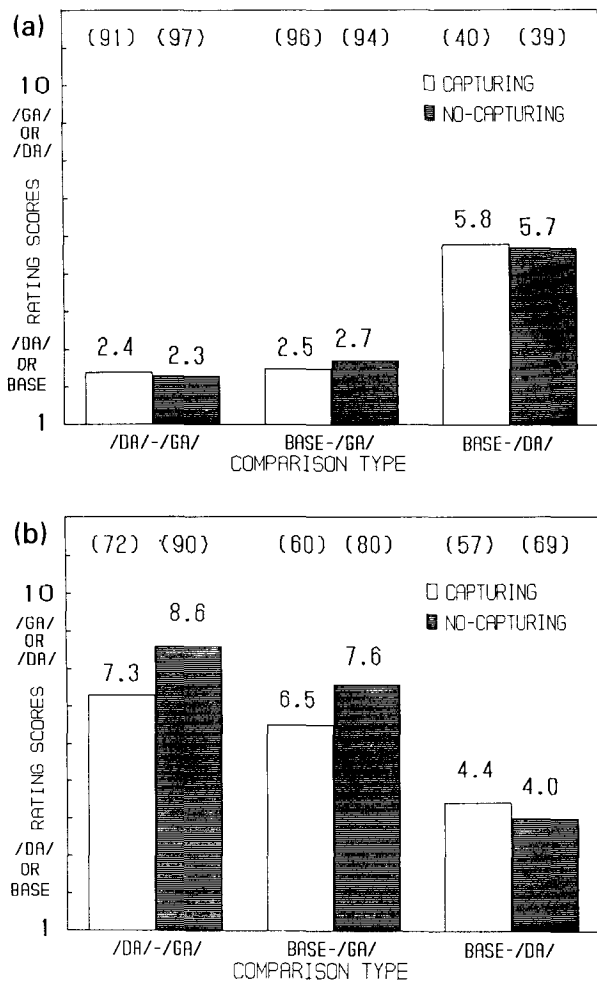
Figure 12. Data for the stimuli of Experiment 4. (a) The mean rating scores and percentages of base or /da/ (/da/-/ga/ comparisons) responses for the /da/ stimuli. (b) The results for the /ga/ stimuli. The mean percentages of /ga/ or base (base-/da/ comparisons) responses are presented at the top of the figure.

/ga/, and base were presented after the experimental patterns. The capturing and no-capturing /da/-/ga/ patterns were those used in the previous experiment.

Stimuli. The stimuli were the same as in the previous experiment. The comparison syllables were those employed in Experiment 2 for the /da/-/ga/ stimuli.

Apparatus. The apparatus was the same as in Experiment 1.

Results and Discussion

The means of the rating scores for the /da/ stimuli are shown in Figure 12a and those for the /ga/ stimuli are displayed in Figure 12b. A two-way ANOVA with repeated measures was applied to the rating scores of the /ga/ stimuli for the /da/-/ga/ and base-/ga/ comparisons. The comparison-type main effect was significant at the .004 level [$F(1,16) = 11.49$]. This effect indicates that the /ga/ stimuli were judged as more similar to the base than to the /da/ comparison. This finding replicated that of Experiment 2. Most importantly, the capturing effect was found to be significant [$F(1,16) = 12.31, p < .003$].

This effect is consistent with the capturing effect found in Experiment 2 with the same discrimination task. The capturing × comparison-type interaction was not statistically significant. Together with the results of Experiment 3, the findings from Experiment 4 support the idea that the weakening of phonetic fusion is likely to be due to the streaming of the critical transition rather than to habituation. This conclusion further supports the idea that duplex perception is not likely to be due to the independent operation of speech and auditory systems.

GENERAL DISCUSSION

The results of these experiments show that duplex perception can be weakened by capturing the isolated transition of a duplex stimulus into a sequence of identical transitions. This has been shown to occur when both the identification and the discrimination tasks are employed, as well as with two different pairs of synthetic syllables. The trajectory-based capturing of Experiments 3 and 4 was weaker than that obtained with streaming transitions identical to the critical one (in Experiments 1 and 2). This finding agrees with the results of Darwin and Gardner's (1987) study. Darwin and Gardner found that a sequence of tones had a weaker capturing effect on a harmonic of a vowel when ascending or descending series of tones were employed than when a steady-state series of tones was employed. They suggested that these results may indicate the absence of trajectory-based capturing. One might think that the remaining capturing effect could be due to the frequency proximity of the streaming transitions immediately preceding and following the target transition. In any case, Experiments 3 and 4 showed that a significant capturing effect was found even when habituation was not likely to have occurred. This evidence contradicts the hypothesis that the auditory and the phonetic modes are independently extracting whatever information they need from the acoustical signal. Our results are also incompatible with the "preemptiveness" idea suggested recently by Mattingly and Liberman (in press) and Whalen and Liberman (1987). According to this theory, the speech system has priority over the auditory system in extracting information from the acoustic signal.

In the duplex phenomenon, the acoustic energy of the isolated transition is employed for the formation of two simultaneous auditory objects. The results of the present study indicate that the integration between the isolated transition and the base was strongest if there was no clear evidence for the assignment of the isolated transition to a separate stream. In the capturing conditions of the previous experiments, where such evidence was present, the integration of the transition with the base stimulus was not very strong. This finding suggests that the independence hypothesis is not an adequate explanation of the duplex perception phenomenon.

There is an alternative explanation of the duplex perception phenomenon, which agrees with the results of the present study. This alternative, which can be called the

auditory scene-analysis hypothesis, suggests that the double assignment of the isolated transition to two co-occurring sounds is due to the presence of conflicting cues for the integration and the segregation of the isolated transition and the base. The difference in the ear of presentation and the offset asynchrony may be cues favoring the segregation between the isolated transition and the base. For example, Repp et al. (1983) reported a significant difference in the identification of diotic and duplex syllables, as we have done in the present study. On the other hand, acoustic cues such as the simultaneous presence of acoustic energy, the onset asynchrony, and the similarity in fundamental frequency may favor the integration of the isolated transition with the base. Previous studies have shown that such integration can occur even with relatively large onset asynchronies and F0 differences between the contralateral stimuli (Cutting, 1976; Repp & Bentin, 1984). However, in none of those studies was it conclusively excluded that the possibility that a combination of acoustic cues may cause a strong integration, even though each cue may prove relatively unimportant by itself. Moreover, the integration of base and isolated transition *was* found to be significantly affected by sufficiently large F0 differences and onset asynchronies (Repp & Bentin, 1984; Repp et al., 1983). Therefore, a duplex stimulus may produce a situation of ambiguity in which the transition is equally likely to belong to the syllable and to be a distinct nonspeech sound. To summarize, the auditory scene-analysis hypothesis states that the primitive auditory grouping processes create the duplexity of the percept before the sounds are recognized as a syllable and a nonspeech chirp. According to this idea, the speech and auditory recognition processes would operate after the double allocation of the isolated transition to two auditory objects by the primitive grouping processes. The description of the contralateral sounds as a CV syllable and a chirp would be accomplished at this later stage of processing.

The occurrence of a significant capturing effect in our streaming experiments can be explained according to the auditory scene-analysis hypothesis. When a duplex stimulus is presented, the auditory scene-analysis processes must determine where the isolated transition belongs. In the capturing conditions of the present study, preattentive processes of organization (Neisser, 1967) would make it more likely that the transition is part of a nonspeech event, the sequence of chirps. This is because the presence of capturing transitions is an additional segregation cue with respect to both the duplex stimuli in isolation and the no-capturing conditions. For this reason, one would expect the integration of the transition with the base stimulus to be weaker. Such an explanation is similar to that formulated by Bregman (1987). It is possible that attention processes operate after the assignment of the isolated transition to two contralateral organizations of the sensory data. This could explain the failure to disrupt the duplex phenomenon by selective attention strategies (Repp et al., 1983).

REFERENCES

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception & Performance, 4*, 380-387.

Bregman, A. S. (1987). The meaning of duplex perception: Sounds as transparent objects. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 95-111). Dordrecht, The Netherlands: Martinus Nijhoff.

Bregman, A. S. & Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception & Performance, 1*, 262-267.

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

Darwin, C. J., & Gardner, R. B. (1987). Perceptual separation of speech from concurrent sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 112-124). NATO ASI Series, Nijhoff.

Henke, W. L. (1975). *MITSYN: An interactive dialogue language for time signal processing.* Cambridge, MA: MIT Research Laboratory of Electronics.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist, 37*, 301-323.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics, 30*, 133-143.

Mattingly, I. G., & Liberman, A. M. (in press). In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Functions of the auditory system.* New York: Wiley.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton.

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

Repp, B. H., & Bentin, S. (1984). Parameters of spectral/temporal fusion in speech perception. *Perception & Psychophysics, 36*, 523-530.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics, 33*, 333-337.

Steiger, H. & Bregman, A. S. (1982). Competition among auditory streaming, dichotic fusion, and diotic fusion. *Perception & Psychophysics, 32*, 153-162.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Eindhoven, Holland: Institute for Perception Research.

Wertheimer, M. (1923). Untersuchungen zur Lehre der Gestalt, *Psychologische Forschung, 4*, 301-350.

Whalen, D. H. & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 273*, 169-171.

NOTES

1. The /da/-/ga/ syllables employed in Experiments 1-4 were synthesized on the Haskins Laboratories parallel resonance synthesizer. These syllables had been used in the Repp, Milburn, and Ashkenas (1983) study and were kindly sent to us on tape by Bruno Repp.