THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1992, *44A* (3) 577–593

The Phonetic Integration of Speech and Non-speech Sounds: Effects of Perceived Location

Valter Ciocca, Albert S. Bregman, and Kathleen L. Capreol

McGill University, Montreal, Canada

The third-formant (F3) transition of a three-formant /da/ or /ga/ syllable was extracted and replaced by sine-wave transitions that followed the F3 centre frequency. The syllable without the F3 transition (base) was always presented at the left ear, and a /da/ (falling) or /ga/ (rising) sine-wave transition could be presented at either the left, the right, or both ears. The listeners perceived the base as a syllable, and the sine-wave transition as a non-speech whistle, which was lateralized near the left ear, the right ear, or the middle of the head, respectively. In Experiment 1, the sine-wave transition strongly influenced the identity of the syllable only when it was lateralized at the same ear as the base (left ear). Phonetic integration between the base and the transitions became weak, but was not completely eliminated, when the latter was perceived near the middle of the head or at the opposite ear as the base (right ear). The second experiment replicated these findings by using duplex stimuli in which the level of the sine-wave transitions was such that the subjects could not reliably tell whether a /da/ or a /ga/ transition was present at the same ear as the base. This condition was introduced in order to control for the possibility that the subjects could have identified the syallables by associating a rising or falling transition presented at the left ear with a /da/ or /ga/ percept. Alternative suggestions about the relation between speech and non-speech perceptual processes are discussed on the basis of these results.

The spectral changes over time of the third formant transition of a three-formant synthetic syllable can determine whether the syllable is perceived as /da/ (falling transition) or /ga/ (rising transition). When the F3

Requests for reprints should be sent to Valter Ciocca, Experimental Psychology, University of Sussex, Falmer, Brighton, BN1 9QG, U.K..

Financial support came through a grant awarded to the second author by the Natural Sciences and Engineering Council of Canada.

We are very grateful to Peter Bailey, Ray Meddis, and an anonymous reviewer for providing helpful comments on earlier drafts of the article; Rhonda Amsel for advice on data analysis; Pierre Abdel Ahad for technical support.

transition is presented in isolation to one ear (isolated transition) and the remainder of the syllable (base) is presented to the opposite ear (see Figure 1), listeners perceive a full syllable at the ear of the base, and a non-speech "chirp" at the ear of the isolated transition. For most subjects, the type of isolated transition determines whether the syllable is identified as /da/ or /ga/. Hence, the isolated transition is phonetically integrated with the base, and, at the same time, it is perceived as a distinct non-speech chirp (Liberman, 1982; Repp, Milburn, & Ashkenas, 1983). This phenomenon, known as "duplex perception", has been investigated by manipulating the F2 and/or the F3 transitions of several two- or three-formant synthetic syllables (Cutting, 1976; Liberman, Isenberg, & Rakerd, 1981; Rand, 1974; Repp & Bentin, 1984).

Duplex perception has been considered by some researchers as one of the most convincing demonstrations of the existence of independent auditory and phonetic processes (Liberman, 1982; Mattingly & Liberman, 1988). However, the results of a recent study are incompatible with the idea that the phonetic system is able to use acoustic information independently of how the auditory system may organize the same information (Ciocca & Bregman, 1989). In the latter study, the isolated transition of a duplex syllable was preceded and followed by a sequence of identical transitions (capturing transitions). The capturing transitions segregated the isolated transition into a distinct auditory event or stream, thereby weakening its phonetic integration with the contralateral base. On the

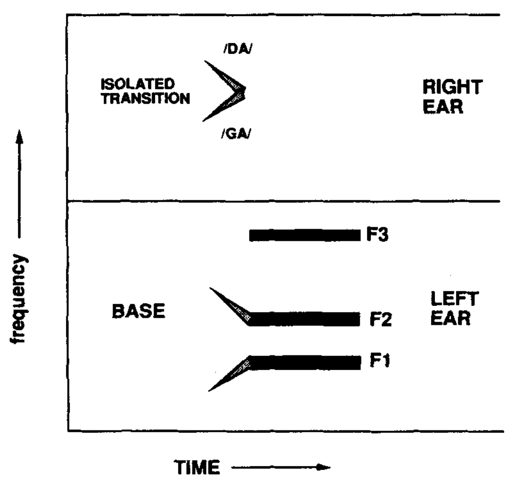

FIG. 1. A schematic representation of the type of the duplex stimuli used in the present investigation; the type of isolated transition presented at the right ear determines whether the syllable at the left ear is perceived as /ga/ or /da/.

basis of their results, Ciocca and Bregman (1989) suggested that duplex perception occurs in the presence of conflicting cues for the integration and the segregation of the isolated transition from the base ("auditory scene analysis" hypothesis). For example, the contralateral presentation of the base and the isolated transition may be taken as evidence for the presence of two distinct auditory objects. Indeed, several studies showed that phonetic integration was significantly stronger when the base and the isolated transition were presented at the same ear than when they had a different location (Ciocca & Bregman, 1989; Repp, Milburn, & Ashkenas, 1983). On the other hand, a plausible integration cue could be the common onset of the base and the isolated transition. It has been shown that phonetic integration is weakened by the presence of an onset asynchrony between the base and isolated transition (Cutting, 1976; Repp & Bentin, 1984).

The present study had the purpose of investigating the effects of the auditory organization factor of spatial location on the phonetic integration of duplex stimuli. In order to do this, the lateralization of the isolated transition was systematically varied across trials, whereas the base was always presented at the left ear. The isolated transitions employed in our study were pure tone glides (sine-wave transitions) that followed the centre frequency of periodic /da/ or /ga/ F3 transitions. Whalen and Liberman (1987) found that sine-wave transitions, when presented at the same ear as the base, could reliably be used by the listeners in the identification of the base as /da/ or /ga/. In addition, the sine-wave transitions could be heard as distinct non-speech whistles when their intensity was approximately equal or higher than that of the steady portion of F3 ("duplexity threshold"; Whalen & Liberman, 1987). Sine-wave transitions were used in the present study because we observed that the lateralization of binaural periodic transitions could not be manipulated precisely. One possible reason for this could be that, for phonetic and/or auditory reasons, periodic transitions were integrated with the base more strongly than were sine-waves. For example, phonetic processes could consider periodic transitions presented at the same ear as the base to be more appropriate as parts of a phonetic percept than sine-waves. Secondly, periodic transitions, unlike their sine-wave analogs, are harmonically related to the base, and they could, therefore, be grouped with the base on the basis of their common harmonic structure. Finally, both periodic and sine-wave transitions presented at the same ear as the base (ipsilateral transitions) were probably masked by the latter to a certain extent, due to the well-known phenomenon of the upward spread of masking, which, incidentally, was at the origin of the first duplex perception study (Rand, 1974). However, the amount of masking was likely to be larger for periodic than for sine-wave transitions whose energy is concentrated in a narrow frequency region.

EXPERIMENT 1

In Experiment 1, the /da/ or /ga/ sine-wave transitions were presented at the left ear, the right ear, or at both ears. As the base was always presented at the left ear, it was expected that the sine-wave transitions would be perceived either at the same ear as the base, at the opposite ear, or at the centre of the head, respectively. The purpose of this experiment was to determine whether and how the phonetic integration of the sine-wave transitions with the base was affected by a variation in the lateralization of the transitions. Therefore, lateralization was used as a way of capturing the sine-wave transitions that was analogous to the capturing by frequency proximity employed in the study by Ciocca and Bregman (1989).

Method

Subjects. The subjects were 16 graduate or undergraduate students at McGill University. All subjects received an audiometric test and were shown to have normal hearing. The data of 4 of these 16 subjects were discarded because they could not reliably identify the experimental speech stimuli as /da/ or /ga/.[1]

Procedure. There were two sessions in this experiment: an identification and a lateralization session. The subjects received the identification session first, and after a break of a few minutes, they participated in the lateralization session. The experimental patterns were identical for both the identification and the lateralization sessions. The stimuli consisted of the base, and either a /da/ or a /ga/ sine-wave transition. Each of the two transitions (/da/ or /ga/) could be presented either at the left, right, or to both ears. Therefore, there was a total of six duplex stimuli in this experiment. In each trial, one of the duplex stimuli was repeated twice, to allow for more accurate identification. There was a silence of 800 msec between repetitions of the stimuli within a trial; a 4-sec silence, during which time the subjects gave their responses by using a keyboard, followed the last stimulus in a trial.

[1]It should be noted that many more acoustic cues for the perception of different stop consonants are usually available in natural speech than in synthetic speech (see, for example, Lisker, 1978; cited in Liberman, 1982). In addition, as Darwin (1976) pointed out, not all listeners make use of the same phonetic cues to the same extent. For these reasons, it is not uncommon to find that about one-third to one-quarter of naive listeners cannot reliably identify synthetic syllables that differ only by the onset frequency of one or two formant transitions (see, for example, Liberman, Isenberg, & Rakerd, 1981). As the majority of naive listeners can identify synthetic syllables precisely, as is shown in the present study, it is reasonable to assume that most listeners would perform accurately if given enough training.

During the identification session, the subjects were told to ignore any whistles they might hear at the same time as the syllables. A 10-point rating scale was used to measure the subjects' identification of the syllables. The numbers 1 to 5 were chosen to indicate the perception of /da/, whereas the numbers 6 to 10 indicated the perception of /ga/. The extremes of the scale (1 and 10) were labelled "very confident"; numbers 5 and 6 indicated a "not confident" judgement. Hence, the listeners were forced to choose between /da/ and /ga/ but could rate the confidence with which they made their judgements. No feedback was given to the subjects as to whether or not their identification of the syllables was correct. Each of the six duplex stimuli was used on 10 trials in a session. The order of presentation was randomized, and it differed for every subject.

In the lateralization session, the stimuli were the same as in the identification session. The subjects were instructed to ignore the syllables and to pay attention only to the location of the whistles within their heads. A 10-point rating scale was used to measure the lateralization of the whistles. The rating scale represented the space between the subjects' ears, number 1 indicating the extreme left and number 10 the extreme right. Numbers 5 and 6 were used to indicate the perception of a centrally located whistle.

Stimuli. The base was a three-formant syllable synthesized using a parallel formant circuit. The circuit was implemented with the MITSYN digital processing software (Henke, 1987). Each formant consisted of a 50-msec formant transition and a 150-msec steady-state portion. Hence, the total duration of the syllable was 200 msec. The complete F1 and F2, but only the steady portion of F3, formed the "base stimulus" or "base". The centre frequencies of the F1 and F2 steady portions were 771 Hz and 1233 Hz, respectively. The centre frequency of the steady-state F3 was 2525 Hz. The centre frequency of the F1 transition rose from 282 to 771 Hz, and that of the F2 transition fell from 1770 to 1233 Hz. The bandwidths of the formants were similar to those typical for the /a/ vowel (Klatt, 1980). Specifically, the bandwidths were 120 Hz for F1, 70 Hz for F2, and 130 Hz for F3, and they were fixed for both the transition and the steady portions of the syllables. The fundamental frequency of the first and second formants, and of the steady portion of F3, was fixed at 100 Hz. The F3 transition was simulated by a sinusoidal glided tone (sine-wave transition), which was not harmonically related to the base and which followed the centre frequency fo the F3 transition of a normal /da/ or /ga/. A sine-wave transition falling in frequency from 2862 to 2525 Hz simulated a /da/ transition, and a tone rising from 2018 to 2525 Hz was a cue to a /ga/ percept. The sine-wave transitions were synthesized separately from the base, and they were digitally mixed with it. The intensity level of the sine-wave transition was chosen so that it could be heard as a whistle distinct from the base, even when the two stimuli were presented to the same ear. Whalen and Liberman (1987) determined the level (which they called the "duplexity threshold") at which a sine-wave /da/ or /ga/ transition could be barely audible as a whistle in the presence of the base. They found the duplexity threshold to be 0 dB for the /ga/ transition, and −6 dB for the /da/ transition, relative to the level of the steady-state F3 portion. In order to guarantee the audibility of the whistle in the light of these results, we decided to set the intensity of the sine-wave transitions to 6 dB above the level of the F3 steady-state portion. The base was presented at the left ear at a level of 70 dBA.

Apparatus. The stimuli were synthesized and played by a Compaq 386/20 computer controlling a Data Translation DT2823 16-bit D/A converter during the experiment. The stimuli had a sampling rate of 16 KHz and were low-pass filtered at 8 KHz by a TTE J1550 filter, which provided an attenuation of 60 dB at 11.2 KHz. As there was no energy above 4 KHz in the synthesized stimuli prior to filtering, this filter effectively removed noise due to aliasing from the output signal. Subjects listened individually to the stimuli over Telephonics TDH-39 headphones, in an Industrial Acoustics 1202 audiometric chamber. The intensity of the stimuli was calibrated by means of a General Radio Type 1551-C sound level meter connected to the headphones by 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

Lateralization Results. The lateralization scores represent the perceived location of the whistles within the subjects' heads. The mean rating scores, shown in Figure 2, indicate that the subjects lateralized the whistles at the expected locations within the head.

When the sine-wave transitions were presented binaurally, the perceived whistles were not lateralized exactly at the middle-of-the-head location. Instead, they were lateralized most often a little towards the right ear (opposite to the base stimulus). It is likely that this result was due to the partial masking of the ipsilateral transition by the base. This explanation is supported by the previous finding that when a monaural noise is presented at the same time as a binaural click, the click is lateralized towards the ear opposite to that of the noise (Deatherage & Hirsh, 1959; cited in Durlach & Colburn, 1978).

Identification Results. The mean rating scale scores for the identification session are shown in Figure 3. The scores were analyzed by a

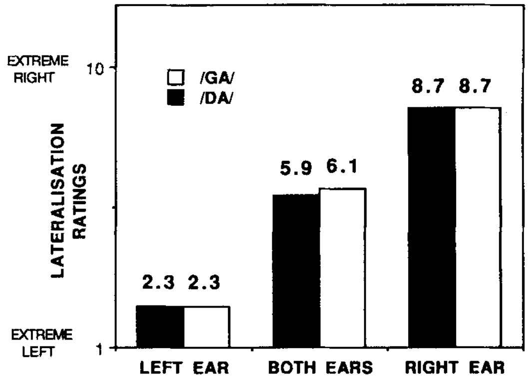

FIG. 2. Mean ratings for the lateralization session of Experiment 1.

two-way ANOVA with repeated measures. The two factors were the "transition-type", and the "ear of presentation". The main effect of "transition-type" was significant, $F(1, 11) = 31.37$, $p < 0.0002$, showing that the subjects reliably labelled the syllables as /da/ or /ga/ with the falling or rising transition, respectively. The "ear of presentation" main effect was significant, $F(2, 11) = 4.29$, $p < 0.05$. This effect shows that the number of /da/ responses changed as a function of the lateralization distance between the transitions and the base. The "ear of presentation" × "transition-type" interaction was highly significant, $F(2, 22) = 16.49$, $p < 0.0001$. This interaction indicates that the difference in rating scores between conditions with /da/ and /ga/ transitions became smaller as the perceived lateralization of the whistles differed increasingly from that of the base (see Figure 3).

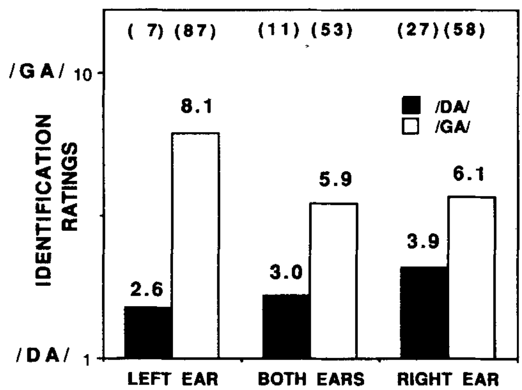

FIG. 3. Mean ratings for the identification session of Experiment 1; the percentage of /ga/ responses for each condition displayed at the top of the figure.

The results of this experiment show that, when the sine-wave transitions were perceived at a different lateralization from the base, the syllables became more ambiguous—that is, the ratings for both the /da/ and /ga/ conditions moved from the opposite extremes toward the centre of the rating score range. The difference between left and binaural, as well as that between left and right conditions was significant for each type of transition ($p < 0.01$, planned comparison). This finding suggests that the influence of the sine-wave transition on the identity of the syllable became weaker when their lateralization differed. However, the binaural and right conditions were not statistically different for either the /da/ or the /ga/ transitions ($p > 0.05$, planned comparison). Moreover, the size of the /ga/–/da/ difference, indicating the amount of phonetic integration between the base and the transitions, was not significantly smaller for the right than for the binaural conditions ($p > 0.05$, planned comparison). Hence, it is not possible to support an unambiguous conclusion that the size of phonetic integration decreased as the lateralization distance between the base and the transition increased. Finally, the results seem to indicate that some phonetic integration occurred even when the base and the sine wave transitions were presented at opposite ears. Indeed, the difference between the right /da/ and /ga/ conditions was found to be significant ($p < 0.01$, planned comparison).

Experiment 1 showed that there was some phonetic integration, though weak, even when the transition was presented at the ear opposite to the ear of presentation of the base. However, in that experiment the subjects could have learned to associate the rising transition with the perception of /ga/ and the falling transition with a /da/ syllable. Indeed, in the monaural conditions, which produced strong phonetic integration, the subjects could clearly hear a rising or falling sine-wave transition at the same ear as a /ga/ or /da/ syllable. Consequently, the identification of the syllables in the dichotic conditions could have been achieved through associative learning rather than genuine phonetic integration. The existence of weak across-ear phonetic integration of sine-wave transitions would be an important finding, because it would indicate that auditory organization factors, even when combined, might not be able to completely overcome the operation of phonetic processes. In the following experiment, associative learning was prevented by eliminating the occurrence of monaural duplex percepts.

EXPERIMENT 2

In Experiment 2, the /da/ or /ga/ sine-wave transitions were presented either monaurally or binaurally. In the monaural conditions, a sine-wave transition was presented ipsilaterally to the base at three intensity levels, which were chosen so that the subjects could not discriminate whether a rising or a falling transition had been presented at the same ear as the syllable. The highest level was close to the monaural duplexity threshold for both the /ga/ and the /da/ transitions, whereas at the lowest level the monaural transitions should not have affected the identification of the base (Whalen & Liberman, 1987). In the binaural conditions, a contralateral transition of the same type as the ipsilateral one, and exactly synchronized with it, was added to each of the monaural stimuli. The level of the contralateral transitions was chosen so that the loudness of the binaural transitions, when heard in the absence of the base, was equalized across the binaural conditions. The amount of phonetic integration due to the presence of the contralateral transitions could be determined by comparing the monaural conditions with the corresponding binaural conditions.

Method

Subjects and Pretest. The subjects were 18 undergaduate and graduate students, or staff members at McGill University. Before the beginning of the experiment, there was a training session during which binaural /da/ and /ga/ syllables were presented to the subjects. These syllables contained a periodic F3 transition, so that the subjects would not gain experience with the sine-wave transitions used in the experiment. After listening to five repetitions of each syllable, the subjects had to judge whether the members of eight pairs of syllables were identical or different. Four subjects gave two or more incorrect answers in one of three successive attempts, and, therefore, they did not participate in the experiment. Of the remaining subjects, one perceived all the syllables as /da/, and another identified all the stimuli as ambiguous. Their data were not included in the statistical analysis as they did not provide any useful information about the differences in the phonetic identity of syllables with rising and falling sine-wave transitions. In addition, one subject gave a response of 10 in all the trials of the lateralization session. It is likely that this subject misinterpreted the meaning of the scale by considering it a left–right confidence scale. For this reason, his data from the lateralization session were not included in the statistical analysis.

Procedure. The sequence and timing of stimuli and trials were identical to those of Experiment 1. In the identification session, 6 monaural stimuli (/da/ and /ga/ transitions with 3 intensity levels) and 6 binaural stimuli were presented 10 times. The subjects were told that, at times, they might hear a whistle at the same time as the syllable, and they were instructed to ignore the whistles whenever they could hear them. In the lateralization session, only the 6 binaural stimuli were presented, as the sine-wave transitions could not be heard as distinct sounds in the monaural conditions.

Stimuli. The /da/ and /ga/ sine-wave transitions were presented either monaurally or binaurally. In the monaural conditions, either a /da/ or /ga/ transition was presented at the same ear as the base. The ipsilateral transitions could have an attenuation of 6 (IH, Ipsilateral High level), 16 (IM, Ipsilateral Medium), or 26 dB (IL, Ipsilateral Low) relative to the level of the F3 steady portion. The level of the most intense (IH) transitions was well below the duplexity threshold for the /ga/ transition and close to the same threshold for the /da/ transition, as specified by Whalen and Liberman (1987). In an informal test, the first author and other listeners verified that the sine-wave transitions could not be heard as separate sounds in the monaural conditions.[2] Whalen and Liberman (1987) reported that the sine-wave transitions affected the phonetic judgement of the syllable until their level was approximately 20 dB below the duplexity threshold. On the basis of this observation, we expected that the softest (IL) transitions would not affect the identity of the syllable. Therefore, the monaural IL conditions acted as a baseline against which we could compare the across-ear phonetic integration in the binaural conditions. The binaural stimuli were created by pairing each of the monaural /da/ and /ga/ stimuli with a contralateral transition of the same type as the ipsilateral transition. The monaural stimuli in which the sine-wave transition was attenuated by 6 dB (IH), 16 dB (IM), or 26 dB (IL) were accompanied by contralateral transitions, which had attenuations of 6 (CL, Contralateral Low level), 2 (CM, Contralateral Medium), and 0 dB (CH, Contralateral High), respectively. The resulting binaural conditions will be called IH–CL (6/6 dB attenuations), IM–CM (16/2 dB attenuations), and IL–CH (26/0 dB), hereafter. The level of the contralateral transitions was set so that the loudness of the perceived whistles, resulting from the combination of the ipsilateral and contralateral transitions, was constant across the binaural conditions (Keen, 1972). When presented without the base, the binaural /da/ or /ga/ whistles were lateralized near the centre of the head (IH–CL), at the centre-right (IM–CM), and at the right ear (IL–CH).

As the level of the ipsilateral transitions was set on the basis of the results of the study by Whalen and Liberman (1987), we wanted to ensure that the masking of the sine-wave transitions by the base in this experiment

[2]An unpublished study by Bailey and Herrmann (personal communication) has shown that sine-wave transitions may be detected at levels that are considerably lower than those reported by Whalen and Liberman with similar stimuli (1987). However, the former authors also found that their subjects could not reliably discriminate which sine-wave transition (rising or falling) was present at the same time as the base when the level of the transitions was 6 dB above or below the level of the F3 steady-state portion. As the highest level of the ipsilateral transitions in Experiment 2 was 6 dB below that of the F3 steady-state portion, we can be confident that the subjects could not learn to associate a rising or falling transition with the perception of a particular syllable during the course of the experiment.

was similar to that which occurred with Whalen and Liberman's stimuli. In addition, the results of Experiment 1 showed that the base employed in that experiment was perceived as being more ambiguous than the base of all the previous investigations, which used /da/–/ga/ duplex stimuli (Repp, Milburn, & Ashkenas, 1983; Repp & Bentin, 1984; Whalen & Liberman, 1987). For these reasons, the base used in the present experiment was synthesized so that it matched the dynamic spectral characteristics of the base used in the study by Repp, Milburn, and Ashkenas (1983[3]). This was done by specifying changing bandwidths for the first 15 msec of the F1 and F2 transitions. The F1 transition bandwidth changed from 70 Hz to 120 Hz; the bandwidth of the F2 transition changed from 160 Hz to 70 Hz. The F1 and F2 bandwidths had a constant value for the remaining portion of the transitions and of the steady-state formants. The bandwidth of the F3 steady portion was fixed at 130 Hz. The F0 of all the formants was fixed at 100 Hz during the first 150 msec of the base, and then it decreased linearly to 75 Hz during the last 50 msec. The duration and intensity characteristics of the base were the same as in Experiment 1.

Apparatus. The apparatus was the same as in Experiment 1, except for the headphones. In Experiment 2, Sony MDR-V7 headphones were used. The level of the stimuli was measured through these headphones.

Results and Discussion

Lateralization Results. The mean lateralization scores are displayed in Figure 4. The data were analysed by a two-way ANOVA with repeated measures. The "relative level" (IH–CL, IM–CM, and IL–CH) and the "transition-type" (/da/ or /ga/) were the experimental factors. The "relative level" main effect was significant, $F(2, 22) = 17.67$, $p < 0.0001$. Pairwise planned comparisons showed that the IH–CL, IM–CM, and IL–CH conditions were significantly different at the 0.05 level for both rising and falling transitions. However, as can be seen in Figure 4, the actual lateralizations of the whistles in the three conditions were close to each other. The whistle in the IH–CL (intended centre) condition was moved even farther away from the base (and from the centre position) than in the analogous condition in Experiment 1. A likely explanation for this difference is that the ipsilateral transitions in the IH–CL condition were 12 dB softer and, therefore, they were probably masked by the base by a larger amount than those in the binaural conditions of Experiment 1.

[3]The stimuli used in Repp, Milburn, and Ashkenas (1983) were kindly sent to us on tape by Bruno Repp.

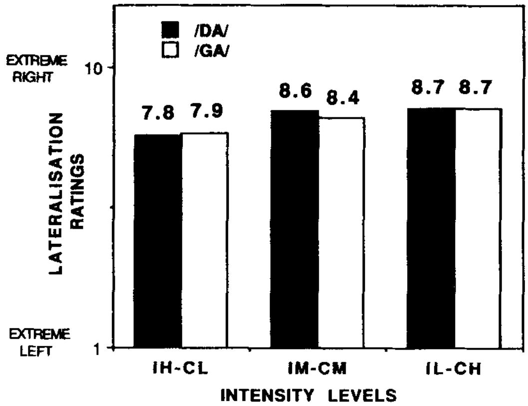

FIG. 4. Mean ratings for the lateralization session of Experiment 2; the binaural transitions had the following attenuations relative to the steady portion of F3: 6 dB at both ears for the IH–CL condition; 16 dB (left ear) and 2 dB (right ear) for the IM–CM condition; 26 dB (left ear) and 0 dB (right ear) for the IL–CH condition.

Identification Results. The mean rating scores for the identification session are displayed in Figures 5 and 6. These scores were analysed by a three-way ANOVA with repeated measures. The experimental factors were "presentation modality" (monaural versus binaural), the "transition-type" (/da/ versus /ga/), and the "intensity" of the ipsilateral transitions (IH, IM, or IL).

The subjects identified the speech percepts reliably according to the type of transition presented, as proved by the fact that the "transition-type" main effect was highly significant, $F(1, 11) = 62.95, p < 0.0001$. The difference between the /da/ and /ga/ monaural conditions was statistically significant for the IH, and IM conditions, $p < 0.01$, planned comparison, but not for the IL condition. As we expected, the phonetic integration of the monaural transitions was strong at the IH (−6-dB) level, intermediate at the IM (−16-dB) level, and virtually absent with the softest (IL) transitions (see Figure 5).

The results for the binaural stimuli replicate several findings of Experiment 1, with the important difference that, in the present experiment, the subjects could not have identified the syllables by associating a rising or falling ipsilateral transition with a /ga/ or a /da/ syllable, respectively. (1) Binaural conditions with the same transition-type obtained virtually identical ratings. As each binaural level of intensity produced small but significant differences in the lateralization of the whistles, we conclude that phonetic integration was not affected by the lateralization of the whistles in the present experiment. This finding confirms the results of Experiment 1, which indicated that phonetic integration was not different between the "centre" and "right ear" lateralizations. (2) The difference between the

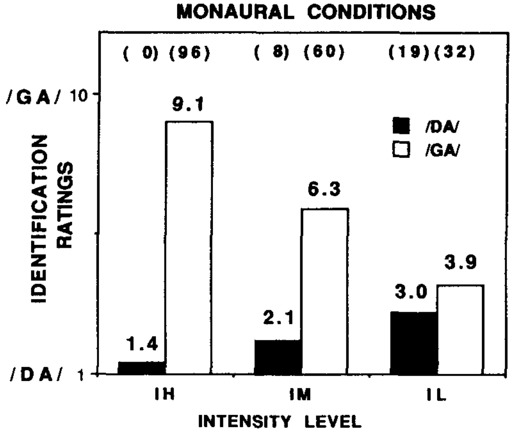

FIG. 5. Mean identification ratings for the monaural conditions of Experiment 2; percentages of /ga/ responses are displayed at the top of the figure.

/da/ and /ga/ IL–CH conditions indicates the existence of across-ear phonetic integration between the sine-wave transitions and the base. This conclusion is further supported by the fact that the IL–CH /ga/ condition obtained significantly more /ga/ ratings than the IL /ga/ condition, $p < 0.01$ level, planned comparison. (3) The IH /ga/ condition received significantly more /ga/ responses than the IH–CL /ga/ condition, $p < 0.01$, planned comparison; see Figures 5 and 6. This finding indicates that the presence of a contralateral transition produced a decrease in phonetic

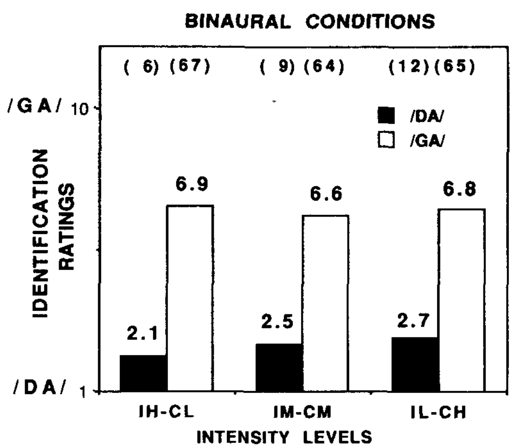

FIG. 6. Mean identification ratings for the binaural conditions of Experiment 2; percentages of /ga/ responses at the top of the figure.

integration relative to the monaural condition. This effect is analogous to the weaker phonetic integration that was observed in Experiment 1 in the "both ears" /ga/ condition, relative to the "left ear" /ga/ condition. The IM and IM–CM /ga/ conditions were not significantly different ($p > 0.05$, planned comparison).

While the levels of intensity did not affect the phonetic perception in the binaural conditions, there were significantly more /ga/ responses at higher levels in the monaural conditions (see Figure 5). This result is responsible for a significant effect of "intensity", $F(2, 22) = 4.69, p < 0.02$, and of the Intensity × Transition-type interaction, $F(2, 22) = 45.09, p < 0.0001$. In addition, the Presentation Modality × Intensity interaction was also significant, $F(2, 22) = 10.38, p < 0.001$. Finally, the three-way interaction effect was highly significant, $F(2, 22) = 19.86, p < 0.0001$. The remaining effects were not statistically significant.

GENERAL DISCUSSION

The present study showed that: (a) phonetic integration was significantly weaker when the base and the sine-wave transitions were perceived at different than at the same lateralizations; (b) sine-wave transitions were phonetically integrated with the base when presented contralaterally to the latter; (c) there was no evidence that, among transitions that were lateralized at different locations from the base, phonetic integration became weaker as the lateralization distance was increased.

The aim of this investigation was to test different hypotheses, namely the "auditory scene analysis" and the "independence" hypotheses, concerning the phenomenon of duplex perception and, more generally, the relations between auditory grouping and phonetic processes. The "independence" hypothesis, in our understanding, would suggest that the phonetic and the "general-purpose" auditory processes are "functionally" independent (Liberman, 1982; Mattingly & Liberman, in press). In other words, the operation of phonetic processes leading to the perception of speech sounds should not be affected by the concurrent processing of the same acoustic input by auditory grouping mechanisms. Accordingly, the phonetic integration of the F3 transitions with the base in duplex perception should be the result of phonetic processing only. On the other hand, the "auditory scene analysis" hypothesis suggests that phonetic processes should not be able to employ a portion of the acoustic input as "phonetic information" if there is enough evidence, based on auditory grouping principles, that the same input may belong to a co-occurring non-speech sound (Ciocca & Bregman, 1989). It follows from the latter hypothesis that duplex perception is perceived when there are conflicting auditory grouping cues for the segregation and the integration of the F3 transitions with the base.

The finding at point (a) above indicates that phonetic integration was affected by the lateralization of the sine-wave transitions, which is what one would expect to find on the basis of the auditory scene analysis hypothesis. This result replicates those obtained by presenting periodic /da/ or /ga/ transitions at the same or the opposite ear as the base (Repp, Milburn, & Ashkenas, 1983; Ciocca & Bregman, 1989). However, the decrease in phonetic integration was larger in the present than in the above-mentioned studies: for example, the difference in the percentage of /ga/ responses between /ga/ and /da/ conditions dropped from 96% to 89% when periodic transitions were moved from the same ear as the base to the opposite ear (Ciocca & Bregman, 1989). On the other hand, the /ga/–/da/ difference was 80% for the ipsilateral and 27% for the contralateral sine-wave transitions in Experiment 1. Similarly, the /ga/–/da/ difference was 96% for the ipsilateral presentation of the sine-wave transitions (−6-dB monaural condition) and 53% for the contralateral presentation (−26/0-dB binaural condition) in Experiment 2. As our base stimuli were very similar to those employed in the previous studies, our findings show that the effects of a difference in lateralization between the base and F3 transitions are noticeably increased if the transitions are also harmonically different from the base. The present results suggest that, while individual segregation cues may have little effect per se on phonetic integration, their combination seems to produce a substantial decrease in phonetic integration. This conclusion is in agreement with the auditory scene analysis, but not with the independence hypothesis.

Further support for the auditory scene analysis hypothesis comes from our finding that in Experiment 1, the /ga/–/da/ difference became smaller when the sine-wave transition was presented to both ears ("centre" condition: 42% /ga/ responses) rather than to the ear of the base only ("left" condition: 80% /ga/ responses). Similarly, in Experiment 2 the /ga/–/da/ difference was 96% in the monaural (−6-dB), and 61% in the binaural (−6/−6-dB) /ga/ conditions. It is worth noticing that, according to the independence hypothesis, the presence of additional phonetic information at the opposite ear as the monaural stimuli should have produced no change (or, possibly, an increase) instead of a decrease in phonetic integration. Our results are also at odds with the idea, originally suggested by Whalen and Liberman (1987), that the phonetic system has priority over the non-speech system in the processing of acoustic information (Bentin & Mann, 1990; see also Hall & Pastore, 1991, for further evidence against the precedence of speech over non-speech processing). The auditory scene analysis hypothesis would correctly predict a decrease in phonetic integration when the sine-wave transitions were perceived at a different lateralization from the base.

Two results of the present study—points (b) and (c) above—suggest that the auditory scene analysis hypothesis alone may not fully account for the phonetic integration of duplex stimuli, at least in its original formulation. (1) The results of both experiments showed that, as the difference in lateralization between the base and the sine-wave transitions was increased, phonetic integration did not decrease in a monotonic fashion, as one would have expected if auditory organization processes alone were responsible for the occurrence of duplex perception. For example, the phonetic integration in the "centre" and "right ear" conditions of Experiment 1 was not found to be significantly different (see Figure 3). (2) It was found that a difference in both harmonic structure and lateralization between the base and the F3 transitions was not sufficient to prevent their phonetic integration completely. These results suggest that neither the auditory scene-analysis nor the phonetic processes may have the ultimate word on how the acoustic input should be organized (Bregman, 1987; Darwin & Gardner, 1987). In other words, the auditory organization and the phonetic processes would not be functionally independent, but may compete for the same acoustic signals. This possibility has been acknowledged by Liberman and Mattingly (1989) and, more recently, by Whalen and Liberman (1991).

REFERENCES

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

Bentin, S., & Mann, V. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America*, *88*, 64–74.

Ciocca, V., & Bregman, A.S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics*, *46*, 39–48.

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

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*. Dordrecht: Martinus Nijhoff.

Deatherage, B.H., & Hirsh, I.J. (1959). Auditory localization of clicks. *Journal of the Acoustical Society of America*, *31*, 486–492.

Durlach, N.I., & Colburn, H.S. (1978). Binaural phenomena. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of perception, Vol. 4*. New York: Academic Press.

Hall, M.D., & Pastore, R.E. (1991). Musical duplex perception: Does perceptual dominance reflect general principles or specialized modules? *Journal of the Acoustical Society of America*, *89*, 4(2), 5SP10.

Henke, W.L. (1987). *MITSYN languages*. Copyright by WLH. Belmont, MA: Author.

Keen, K. (1972). Preservation of constant loudness with interaural amplitude asymmetry. *Journal of the Acoustical Society of America*, *52*, 1193–1196.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971–995.

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. *Perception & Psychophysics, 30*, 133–143.

Liberman, A.M., & Mattingly, I.G. (1989). A specialization for speech perception. *Science, 243*, 489–494.

Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, SR-54, 127–132. (ERIC Document Reproduction Service No. ED-161-096).

Mattingly, I.G., & Liberman, A.M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Functions of the auditory system*. New York: Wiley.

Mattingly, I.G., & Liberman, A.M. (in press). Speech and other auditory modules. In M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps*. New York: Wiley.

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.

Whalen, D.H., & Liberman, A.M. (1987). Speech perception takes precedence over non-speech perception. *Science, 237*, 169–171.

Whalen, D.H., & Liberman, A.M. (1991). Independence of scene analysis and the speech module. *Journal of the Acoustical Society of America, 89*, 4(2), 8SP11.

Manuscript received 10 July 1991