Dichotic release from masking for speech

Timothy C. Rand

Haskins Laboratories, New Haven, Connecticut 06510

The University of Connecticut, Storrs, Connecticut 06268

(Received 2 July 1973; revised 17 December 1973)

A dichotic method of listening to synthetic CV speech syllables was investigated, where the first formant ($F1$) is presented to one ear and the second and third formants ($F2$, $F3$) to the other ear. As a test parameter, the intensity of $F2$, $F3$ was varied. Compared to the binaural condition, dichotic presentation permits $F2$, $F3$ to remain effective as speech cues at significantly lower intensity levels. The results are given the following interpretation: peripheral masking, which occurs for ordinary binaural listening, is avoided when speech is heard dichotically. A similar release from masking was found in a second, related experiment where only the initial $F2$, $F3$ transitions were presented to one ear and the remainder of the syllable to the other.

Subject Classification: 70.30; 65.58, 65.75.

Several investigators have observed that presenting the first formant (F_1) of synthetic speech to one ear and the second formant ($F2$) to the other ear results in fusion (Broadbent, 1955; Broadbent and Ladefoged, 1957; Halwes, 1969). Two experiments are reported here that take advantage of this dichotic phenomenon in exploring the perceptual interaction between different formants and formant transitions at varying amplitudes.

It is well known in auditory psychophysics that a low-frequency tone is an effective masker of a higher-frequency tone. Relatively little has been done along these lines with speech, but one might suspect, along with Flanagan and Saslow (1958), that an analogous statement could be made for speech stimuli. Whereas strong low-frequency speech components may mask higher-frequency components under binaural listening conditions, the present results indicate that dichotic presentation produces a release from this masking.

I. EXPERIMENT I

In the first of the two experiments, perception of synthetic speech syllables presented dichotically with the first formant ($F1$) on one channel and the second and third formants ($F2$, $F3$) on the other channel was compared with perception of the same syllables presented binaurally. The syllables [ba, da, ga] were produced with the Haskins Laboratories parallel resonance synthesizer and digitized. A revised version of the Haskins Laboratories pulse code modulation (PCM) system (Cooper and Mattingly, 1969) was used to prepare audio tapes. Separate tapes were recorded for the dichotic experimental condition and the binaural condition. On the dichotic tape, $F2$, $F3$ of each syllable was recorded at six levels of attenuation, ranging from 0 to 50 dB in 10-dB steps. On half of the trials, $F1$ was recorded on channel 1 and $F2$, $F3$ on channel 2; on the remaining trials, this relationship was reversed. With five repetitions of each trial, there were 180 trials: (3 consonants) $\times$(2 ear/formant relationships)$\times$(6 $F2$, $F3$ intensity levels)$\times$(5 repetitions). The trials were recorded in a randomized order with four seconds between trials.

The binaural control tape used six levels of $F2$, $F3$ attenuation ranging from 0 to 30 dB in 6-dB steps. The attenuated $F2$, $F3$ signal was mixed with $F1$ and the composite was recorded on both channels. With six repetitions of each trial there were 108 trials: (3 consonants)$\times$(6 $F2$, $F3$ intensity levels)$\times$(6 repetitions). Since the magnitude of the masking phenomenon had been estimated during pilot work, the unequal attenuation step size between the dichotic and binaural tapes was used in an attempt to cover the intensity ranges most effectively.

The experimental dichotic condition is illustrated in Fig. 1(a); Fig. 1(b) shows the corresponding binaural control condition. The attenuator in the path leading from $F2$, $F3$ permits control over the relative intensities of these formants. Since the three syllables used in this experiment (Fig. 2) differ acoustically only in the $F2$, $F3$ region, attenuating these formants tends to obscure the distinctiveness of the syllables.

Calibration signals were also recorded on both channels of both tapes. The calibration signal was a sustained [a], the vowel used in the syllables. This calibration signal served a number of purposes: first, it enabled the channels to be equalized for level when the tapes were played to listeners; second, it permitted presentation levels to be equated between the dichotic and binaural tapes; and third, it was used to measure the absolute presentation level at the earphones. In this way, the playback level was adjusted so that the

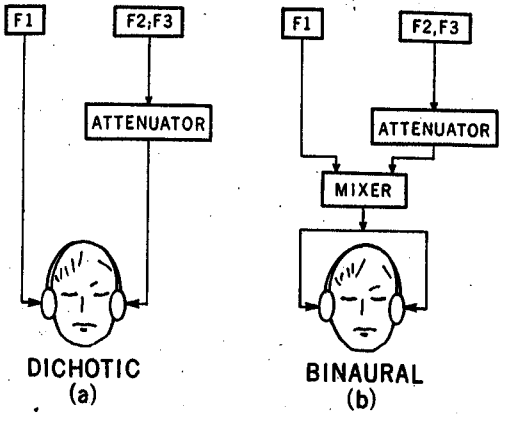

FIG. 1. Dichotic and binaural presentation techniques.

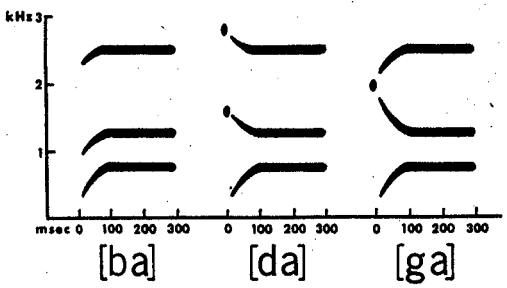

FIG. 2. Syllables from which stimuli were derived.

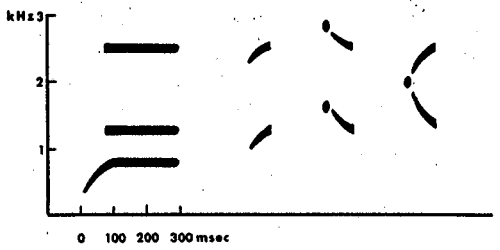

FIG. 4. Stimuli used in Experiment II.

presentation level of the vowel [a], and hence of the vocalic portion of the syllables, was 80 dB SPL. *F*1 in isolation produces a level of 79.5 dB SPL. *F*2, *F*3 (unattenuated) produce 75.5 dB SPL. Thus, for the dichotic condition, the presentation levels on the *F*2, *F*3 channel occurred at decade steps below this level.

For the binaural condition, the presentation level varied less than $\frac{1}{2}$ dB as *F*2, *F*3 ranged from 0 to 30 dB attenuation. This arises from the fact that rms levels for complex signals are not linearly additive with respect to the levels of the signal's components.

Four subjects (young college adults with normal hearing) heard both tapes at one session. Their task was to write down "b", "d", or "g" for each trial.

The results are shown in Fig. 3, where percent correct responses are plotted against the various *F*2, *F*3 attenuation levels for both the binaural control condition (broken line) and the dichotic experimental condition (solid line). In both cases, performance is near 100% for small attenuations and decreases at higher attenuation levels. The high degree of overall identification performance indicates that fusion of the signals in the two channels took place. For the range of attenuations used, performance remains always above the chance level. Inspection of Fig. 3 reveals a rather dramatic release from masking. Approximately 20 dB of attenuation separates the two curves for equal performance levels of 90% or less, the dichotic condition permitting greater attenuation. This may be interpreted as evidence that the normal binaural mode involves a certain degree of masking of the higher formants and that presenting the stimuli dichotically results in a release from this masking. Thus the masking level difference (MLD) is on the order of 20 dB.

II. EXPERIMENT II

The second experiment was similar to the first, the only difference being the way in which the syllables were formed from complementary pieces. Rather than simply separating the formants and leading them to separate channels, the *F*2, *F*3 transition that is the minimal acoustic segment differentiating the syllables was separated from the remainder. This remainder, all of *F*1 plus the steady-state portions of *F*2 and *F*3, was constant across all three syllables. These acoustic segments are displayed in Fig. 4.

Tapes were prepared and the experiment was run as described for Experiment I, with the same group of listeners. The results are plotted in Fig. 5. From a comparison of Fig. 5 with Fig. 3, it is apparent that a similar release from masking took place. The amount of this release, the MLD, is somewhat less for Experiment II (approximately 15 dB).

Closer inspection of the two figures reveals that the primary dissimilarity is between the binaural conditions (broken lines). This says, in effect, that binaural presentation of the Experiment I stimuli, where *F*2, *F*3 were attenuated over the entire syllable, produces roughly 5 dB greater masking than binaural presentation of Experiment II stimuli, where only the *F*2, *F*3 transitions were attenuated. For both experiments, dichotic presentation reduces this masking to a uniform minimum.

It is difficult, in general, to produce a set of synthetic stimuli that are equally identifiable under varying conditions, and the present case is no exception. The overall percentages for correct identification for [ba, da, ga] were 85%, 97%, and 86%, respectively, indicating

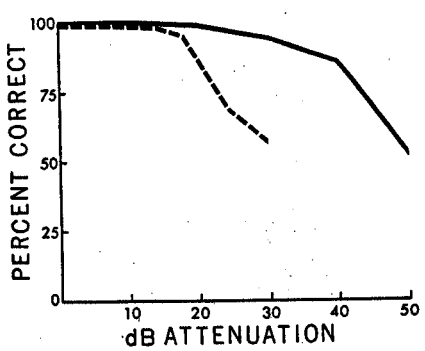

FIG. 3. Experiment I results. Dichotic ——; binaural – – –.

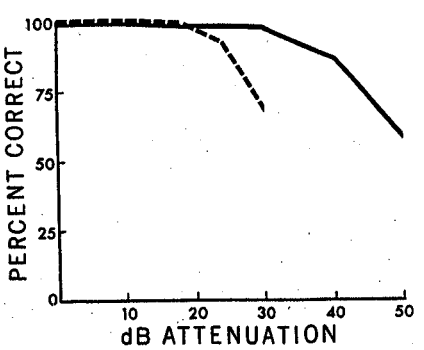

FIG. 5. Experiment II results. Dichotic ——; binaural – – –.

680

that [ba] and [ga] were weaker stimuli.

The individual subjects varied somewhat in overall level of performance, but with respect to the dichotic/binaural differential their results are quite uniform. That is, the pooled results of Figs. 3 and 5, illustrating the dichotic release from masking, are representative of each subject's individual performance.

The dichotic presentations were balanced with respect to which ear received which formant(s). There was no significant difference in performance between these two symmetric conditions, and the dichotic results reported here are combined data for both ear/formant conditions.

III. DISCUSSION

There is a certain practical significance to these results, particularly in the case of Experiment I. The release from masking suggests that dichotic presentation could be beneficial for improving intelligibility under poor signal-to-noise conditions. Investigation along these lines is currently under way at Haskins Laboratories.

The stimuli of Experiment II produce an interesting phenomenon that bears on the question of speech/nonspeech processing. When presented with an $F2$, $F3$ transition in one ear and the remainder of a syllable in the other, listeners report hearing the syllable as well as a nonspeech sound. The nonspeech sound is heard at the ear to which, in fact, the $F2$, $F3$ transition is presented, and the syllable is heard as entering the other ear. Transitions of this type are generally heard as speech events or auditory events depending upon the context in which they occur (Mattingly, Liberman, Syrdal, and Halwes, 1971). The Experiment II stimuli, representing a situation in which a speech "context" is presented to the contralateral ear, produce the experience of hearing both kinds of event at the same time. Liberman, Mattingly, and Turvey (1972) advanced this phenomenon as support for the notion that the perceptual distinction between speech and nonspeech is not made at some early stage on the basis of broad acoustic characteristics.

Broadbent, D. E. (1955). "A Note on Binaural Fusion," Quart. J. Exp. Psychol. 7, 46–47.

Broadbent, D. E., and Ladefoged, P. (1957). "On the Fusion of Sounds Reaching Different Sense Organs," J. Acoust. Soc. Am. 29, 708–710.

Cooper, F. S., and Mattingly, I. G. (1969). "Computer-Controlled PCM System for Investigation of Dichotic Speech Perception," Haskins Lab. Status Rep. Speech Research SR-17/18, 17–21.

Flanagan, J. L., and Saslow, M. G. (1958). "Pitch Discrimination for Synthetic Vowels," J. Acoust. Soc. Am. 30, 435–442.

Halwes, T. G. (1969). "Effects of Dichotic Fusion on the Perception of Speech," Unpublished Ph. D. thesis, Univ. Minnesota (issued as Supplement to Haskins Laboratories Status Report on Speech Research).

Liberman, A. M., Mattingly, I. G., and Turvey, M. T. (1972). "Language Codes and Memory Codes," in *Coding Processes in Human Memory*, edited by A. W. Melton and E. Martin. (V. H. Winston, Washington, D.C.), pp. 307–334.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., and Halwes, T. (1971). "Discrimination in Speech and Nonspeech Modes," Cog. Psychol. 2, 131–157.