An investigation of the interaural time difference threshold for speech

DAVID B. HAWKINS

Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52240

LAMAR L. YOUNG, JR.

Auditory Research Laboratory, Indiana University, Bloomington, Indiana 47401

and

CHERYL A. PARKER

Speech Pathology and Audiology Service, Veterans Administration Hospital, Washington, D.C. 20422

The interaural time difference threshold for speech has been reported to be approximately 35 μ sec (Cherry & Sayers, 1956), a value substantially larger than the 6 μ sec reported for broad-band noise signals by other experimenters (e.g., Tobias & Zerlin, 1959). In the two studies just mentioned, however, different subjects and psychoacoustical methods were employed; thus, it is unclear whether larger interaural time differences are needed to lateralize speech signals. The purpose of this experiment, therefore, was to compare lateralization performance for speech and nonspeech stimuli. Interaural time difference thresholds were obtained for speech, speech spectrum noise, speech multiplied noise, and 200-, 500-, and 1,000-Hz sinusoids for the same subjects using a 2 IFC experimental paradigm. Under these conditions, speech and speech-multiplied noise yielded essentially the same interaural time difference thresholds.

Substantial data are available regarding the interaural time difference thresholds for pure tones and noise stimuli (Hershkowitz & Durlach, 1969; Klumpp & Eady, 1956; McFadden & Pasanen, 1976; Tobias & Zerlin, 1959; Yost, 1974, 1977; Zwislowski & Feldman, 1956). In contrast, only limited results are available which describe the interaural time difference threshold for speech (Cherry & Sayers, 1956; Cherry & Taylor, 1954; Young, Parker, & Carhart, 1975). Consider Figure 1, where the interaural time difference threshold yielded by speech is compared with the thresholds obtained using different stimulus types. On the right portion of Figure 1 are the interaural time difference thresholds for pure tones as reported by Klumpp and Eady (1956), Yost (1974), and Zwislowski and Feldman (1956). Note that the differences in results between these three studies are on the order of only several microseconds. The left portion of this figure presents the available thresholds for broad-band noise (Klumpp & Eady, 1956; Tobias & Zerlin, 1959). The hexagon for the speech stimuli at 33 μ sec on the left of the figure represents the mean interaural time difference threshold for speech obtained by Young, Parker, and Carhart (1975) with 20 inexperienced normal-hearing subjects. This value is

in good agreement with the 35- μ sec interaural time difference threshold for speech reported by Cherry and Sayers (1956). For the moment, ignore the data points represented by the open circles.

It can be seen in Figure 1 that the lateralization threshold for speech is larger than that for a broad-band noise. This would suggest that in order to move

Figure 1. The interaural time difference thresholds in microseconds for several stimulus types. The threshold values shown in this figure are taken from the following studies: ● Klumpp and Eady, 1956; ■ Zwislowski and Feldman, 1956; ▼ Tobias and Zerlin, 1959; ▲ Yost, 1973; and ⬡ Young, Parker, and Carhart, 1975. The open circles are explained in the text.

The experimental portions of this study were completed while all three authors were at the Auditory Research Laboratory at Northwestern University in Evanston, Illinois 60201. Financial support for this research was provided by Grant NS 11857 from the National Institute of Neurological and Communicative Disorders and Stroke.

an intracranial image away from midline, larger interaural time differences are needed for a speech signal than for a nonspeech signal. On the other hand, this difference could be due to training effects and differences in experimental procedures between the different studies. To determine whether the interaural time difference threshold for speech is, in fact, different from the interaural time difference threshold for nonspeech stimuli, an experiment was performed in which well-trained listeners were tested with a variety of signals.

METHOD

Six stimuli were employed in this experiment, three of which were sinusoids of 250, 500, and 1,000 Hz. A fourth stimulus was a broad-band noise filtered to have the same long-term spectral characteristics of speech. The fifth stimulus was produced by multiplying white noise by a single talker reading continuous discourse. The resultant waveform was a broad-band noise which had essentially the same temporal features as speech. This signal was then shaped through a bandpass filter constructed to have a frequency response comparable to the long-term average spectral characteristics of a five-talker complex. Thus, the multiplied-filtered noise was equivalent to speech in terms of gross temporal fluctuations and also long-term spectral features. The noise, however, lacked the fine frequency structure of speech. The final stimulus was a tape recording of a single male talker reading a continuous passage of light essay.

A constant-stimulus two-interval forced-choice (2 IFC) paradigm was utilized. In one of the intervals, the stimuli presented to the two ears were identical, while, in the other interval, an interaural time difference was present. The time differences were imposed after the signals were gated, resulting in onset, ongoing, and offset disparities. The interval which contained the interaural time difference was randomly determined ($p = .5$). Each trial consisted of a warning, first observation, second observation, and response interval, each of which was marked by signal lights. The task of the listeners was to indicate the interval in which the signal was displaced from midline by pressing one of two buttons on a vote box. Trial-by-trial feedback was provided. All six stimuli had durations of 1.5 sec and rise-decay times of 25 msec; they were presented to both ears at an overall level of 75 dB (re 20 μ Pa).

Two normal-hearing listeners were used, both of whom had participated in lateralization studies over a period of 8 months and had had extensive training on each of the six stimuli. Interaural time difference thresholds were determined repeatedly for each stimulus until performance no longer improved. Three interaural time delays were then chosen for each stimulus so as to yield a three-point psychometric function centered around the 75% correct point. Two hundred and fifty trials (in blocks of 50) were run at each of these three time delays.

RESULTS AND DISCUSSION

The psychometric functions obtained with the tonal stimuli for the two listeners are shown in Figure 2. The abscissa presents the interaural time difference and the ordinate represents percent correct. The top panel gives the results obtained with Subject 1 and the lower panel represents the data yielded by Subject 2. Each point is the mean percent correct for the 250 trials. The range of percent correct scores at each time delay averaged about 10% for both listeners across the blocks of 50 trials and did not substantially vary for the differ-

ent stimuli. The lines connecting the points were determined by a method of least squares. If the 75% correct point is taken as threshold, then the following values represent the interaural time difference thresholds for the two subjects: 20 and 30 μ sec at 250 Hz; 11 and 19 μ sec at 500 Hz; and 9 and 14 μ sec at 1,000 Hz. These values are in excellent agreement with those reported by other investigators with tonal stimuli (Klumpp & Eady, 1956; Yost, 1974; Zwislack & Feldman, 1956).

The psychometric functions obtained with the speech-filtered noise (SFN), the speech-multiplied-filtered noise (SMFN), and speech for the same two listeners are shown in Figure 3. The 75% correct points for these stimuli for Subjects 1 and 2, respectively, were 8 and 9 μ sec for the SFN; 9 and 11 μ sec for the SMFN; and 11 and 12 μ sec for the speech.

The data from these two subjects for the different stimulus types have been averaged and are represented in Figure 1 by the open circles. Several observations can be made concerning the data from this experiment. First, the interaural time difference threshold for speech obtained in this experiment was not substantially different from that for the speech-multiplied-filtered noise. Such a finding indicates that the auditory system is equally sensitive in detecting interaural time differences between speech signals as it is in detecting time differences between noise signals having the same long-term spectral and temporal characteristics of speech. The data of Cherry and Sayers (1956) and Young, Parker, and Carhart (1975) suggested that the interaural time difference threshold for speech was larger than that for noise. However, the present experiment demonstrates that when speech and noise which have similar gross spectral and temporal characteristics are presented

Figure 2. Psychometric functions showing percent correct performance as a function of interaural time difference for sinusoidal signals of 250, 500, and 1,000 Hz. The results for Subject 1 are shown in the top panel, and those for Subject 2 are represented in the lower panel.

Figure 3. Psychometric functions showing percent correct performance as a function of interaural time difference for speech, speech-filtered noise (SFN), and speech-multiplied-filtered noise (SMFN). The results for Subject 1 are shown in the top panel, and those for Subject 2 are represented in the lower panel.

to well-trained listeners under the same experimental conditions, then the lateralization performance for these two stimuli is virtually identical.

A second point of interest in Figure 1 concerns the fact that smaller thresholds were obtained with the speech and noise signals than with the tonal stimuli. This finding is in agreement with results from Klumpp and Eady (1956), who found smaller interaural time difference thresholds for wide-band noise than for various pure tones. One explanation of this difference may be related to the effects of the bandwidth of a noise on a lateralization performance. McFadden and Pasanen (1976) have shown that for a noise band having a center frequency of 4,000 Hz, the lateralization threshold improves as the bandwidth of the noise is increased. This is true up to a point close to the width of the monaural critical band at 4,000 Hz. Thus, a wide-band stimulus may be filling (so to speak) a number of critical bands. This is in contrast to the tonal stimulus which activates only a narrow portion of one critical band. A second possible explanation is that with a complex wide-band stimulus, the auditory system is provided with several cues as opposed to a single cue for a tone. For example, the broad-band noise stimulus provides interaural time information for both high- and low-frequency spectral regions; in addition, there are time differences present in the envelope fluctuations. Whether the combination of these cues serves to enhance performance on a lateralization task has yet to be demonstrated.

The final observation relates to the fact that the signals in this experiment were delayed after they were

gated, thus providing three sets of potential cues: onset, offset, and ongoing time differences. Because of the duration of the intervals, and the data reported by Perott and Baars (1974), Tobias and Schubert (1959), and Yost (1977), it was felt that this arrangement would not yield different results than if only ongoing time differences were present. To determine if the presence of multiple cues did enhance performance for these signals at the durations employed, a separate experiment was conducted in which only the microstructure of a 500-Hz tone and a speech stimulus were delayed. The onset and offset times of these binaural signals were identical, and the only cue available was an ongoing time difference. Three-point psychometric functions were obtained for the 500-Hz tone and speech signal with one of the trained listeners. The interaural time difference threshold for speech obtained in this manner was identical to the one obtained in the main study, and for the 500-Hz tone it differed by only 2 μ sec. Consequently, it would appear to be safe to conclude that for the stimulus durations we employed, onset and offset differences were not altering lateralization performance.

REFERENCES

- CHERRY, E., & SAYERS, B. Human cross-correlators—A technique for measuring certain parameters of speech perception. *Journal of the Acoustical Society of America*, 1956, **28**, 889-895.
- CHERRY, E., & TAYLOR, W. Some further experiments upon the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*, 1954, **26**, 554-559.
- HERSHKOWITZ, R., & DURLACH, N. Interaural time and amplitude JND's for a 500-Hz tone. *Journal of the Acoustical Society of America*, 1969, **46**, 1464-1467.
- KLUMPP, R., & EADY, H. Some measurements of interaural time difference thresholds. *Journal of the Acoustical Society of America*, 1956, **28**, 859-860.
- McFADDEN, D., & PASANEN, E. Lateralization at high frequencies based on interaural time differences. *Journal of the Acoustical Society of America*, 1976, **59**, 634-639.
- PERROTT, D., & BAARS, B. Detection of interaural onset and offset disparities. *Journal of the Acoustical Society of America*, 1974, **55**, 1290-1292.
- TOBIAS, J., & SCHUBERT, E. Effective duration of auditory stimuli. *Journal of the Acoustical Society of America*, 1959, **31**, 1595-1605.
- TOBIAS, J., & ZERLIN, S. Lateralization threshold as a function of stimulus duration. *Journal of the Acoustical Society of America*, 1959, **31**, 1591-1594.
- YOST, W. Discrimination of interaural phase difference. *Journal of the Acoustical Society of America*, 1974, **55**, 1299-1303.
- YOST, W. Lateralization of pulsed sinusoids based on interaural onset, ongoing and offset temporal differences. *Journal of the Acoustical Society of America*, 1977, **61**, 190-194.
- YOUNG, L., PARKER, C., & CARHART, R. Coherence function for speech. *Journal of the Acoustical Society of America*, 1975, **58**, Suppl. 1.S 54.
- ZWISLOCKI, J., & FELDMAN, R. Just noticeable difference in dichotic phase. *Journal of the Acoustical Society of America*, 1956, **28**, 860-864.

(Received for publication February 5, 1978;
accepted May 3, 1978.)