

April 01, 1963

Foreign accent and speech distortion

Harlan Lane

University of Michigan - Ann Arbor

Recommended Citation

Lane, Harlan, "Foreign accent and speech distortion" (1963). *Psychology Faculty Publications*. Paper 10. <http://hdl.handle.net/2047/d20000857>

This work is available open access, hosted by Northeastern University.

Foreign Accent and Speech Distortion

HARLAN LANE

Department of Psychology, University of Michigan, Ann Arbor, Michigan

(Received 7 August 1962)

Speech distortion is defined broadly as any operation that evokes inappropriate behavior by a listener in response to speech. Two categories of distorting operations are distinguished: (1) response-independent, in which the transfer function applied to the original speech signal is not determined by the probable response of the listener (e.g., masking, filtering); and (2) response-dependent, in which the distorting operation is related to the probable response of the listener during undistorted transmission and therefore may be specified in linguistic terms (e.g., foreign accent). Two experiments examine the effects and interactions of these two types of distortion. Twenty-four Midwest Americans listened to recorded articulation lists rendered by one American and three foreign-born speakers under eight conditions of masking and filtering. Reducing the speech to noise ratio to 20 dB or the transmission bandwidth to 500 cps yields approximately 50% reduction in word articulation for both native and foreign accent speech. The latter was approximately 40% less intelligible than native speech under all experimental conditions.

THE transmission of speech is bounded at both ends by behavior, that of the speaker and the listener. Speech distortion may be broadly defined as any operation that evokes inappropriate behavior by the listener in response to speech. Two broad categories of speech-distorting operations may then be distinguished: response-independent and response-dependent. The former category has received the lion's share of research and includes such operations as filtering, masking, time sampling, etc. These operations are termed response-independent because the parameters of the transfer function applied to a given speech signal are not determined by the probable response of the listener to that signal. Experimental findings show that "vocal communication is highly resistant to distortion" of this kind.¹ The situation is otherwise with response-dependent distortion, however. In this category belong those distorting operations that are based on the probable response of the listener during undistorted transmission. The serial transmission of rumor is one example of such distortion.² The transmission of an original message by an aphasic or dysarthric speaker is another.³ The manipulation of acoustic cues for speech recognition

by such devices as PAT⁴ and Pattern Playback⁵ provides a third example. In all these instances, the nature of the distorting operation is most effectively specified in linguistic terms, that is, with reference to the behavior of a standard listener, although, of course, an acoustic transfer function may be written for each speech signal.

The experiments reported in the present article describe some effects of these two kinds of speech distortion and their interaction. Masking and filtering of speech were selected as representative of response-independent distorting operations. Foreign accent was selected as a response-dependent type of speech distortion, and also in view of its practical importance in vocal communication.

METHOD

Speakers

The four speakers had the following national origins: United States (S₁), Yugoslavia (S₂), India (S₃), Japan (S₄). Thus the Indo-European (Germanic, S₁; Serbian, S₂; Punjabi, S₃) and Japanese (S₄) language groups

¹J. C. R. Licklider and G. A. Miller, "The Perception of Speech," in *Handbook of Experimental Psychology*, edited by S. S. Stevens (John Wiley & Sons, Inc., New York, 1951).

²G. W. Allport and L. Postman, *The Psychology of Rumor* (Henry Holt & Company, Inc., New York, 1947).

³R. S. Tikofsky, Ilse Lehiste, and Rita Tikofsky, "A Study of the Intelligibility of Dysarthric Speech," paper read at the 62nd Meeting of the Acoustical Society of America, Cincinnati (1961).

⁴W. Lawrence, "The Synthesis of Speech from Signals Which Have a Low Information Rate," in *Communication Theory*, edited by W. Jackson (Academic Press, Inc., New York, 1953), pp. 460-471.

⁵F. S. Cooper, A. M. Liberman, and J. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proc. Natl. Acad. Sci.* 37, 318-325 (1951).

were represented.⁶ Each speaker was male and between 25 and 35 years old. Speaker 1 spoke "General American"; speakers 2-4 had little or no training in spoken English prior to coming to the United States, three months before the experiment. They had a "strong" foreign accent (according to listener ratings) and an inadequate command of English for university study according to the University of Michigan's English Proficiency Test.

Word Lists

Five "PB" lists of 50 words each were constructed from phonetically balanced sets compiled by the Harvard Psycho-Acoustic Laboratory.^{7,8} The five typewritten lists were presented to each speaker, and S_1 read the lists aloud while the group was seated in an audiometric room. Each of the four speakers then read the five PB lists in a different order at the rate of one word every five seconds; 30 seconds elapsed between lists. The spoken lists were recorded on one channel of a four-channel tape recorder (Ampex 300-4) and then copied onto a second channel while the record level of each word was adjusted in order to maintain a constant peak amplitude (10 dB below 0 VU or approximately 50 dB SPL with the TDH-39 earphones employed for listening).⁹

Listeners

Twelve Midwest American undergraduates (six male, six female) served in groups of three in the experiment on speech distortion by masking and foreign accent, and a like number in that on distortion by filtering and foreign accent. The subjects wore binaural calibrated headsets while seated in an audiometric room. Sixty-four stimulus conditions (4 speakers \times 4 lists \times 4 levels of masking or filtering) were presented in four different hyper-Graco-Latin square designs, one to each group of three subjects. These designs were orthogonal with respect to masking or filtering. Each listener never heard a speaker read the same list twice, nor was the same list ever heard twice at the same level of masking or filtering. The listeners were told that they would hear common English monosyllables and that they were to write them down.

Masking and Filtering

Masking noise was introduced in the first experiment by mixing the tape-recorded speech signals with

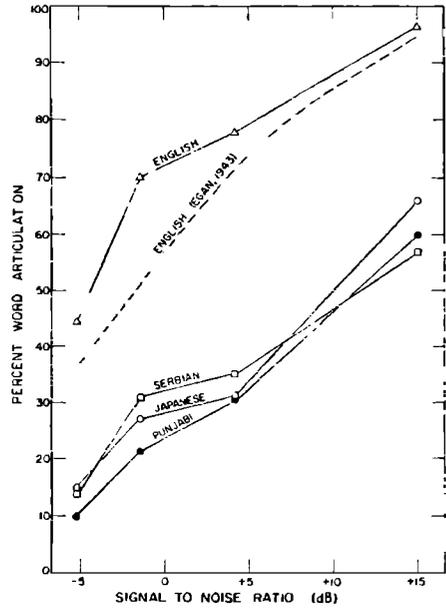


FIG. 1. Percent word articulation as a function of speech-to-noise ratio for native and foreign-born speakers reading English monosyllables. The dotted curve is from an experiment by Egan *et al.*,¹⁰ in which four natives read 400 to 800 words at each of five S/N ratios to six practiced listeners; the level of received speech was 115 dB (SPL). In the present study, speakers of English (Δ), Japanese (\circ), Punjabi (\bullet), and Serbian (\square) each read 200 words at four S/N ratios to 12 listeners; the level of received speech was 50 dB (SPL).

the output of an equal-excitation source (Grason-Stadler, Model 901A) at one of four levels (measured with a Ballantine rms VTVM) to give four signal to noise ratios (S/N): 15, 4, -1.5, -5 dB. This range of S/N ratios was selected in the light of findings reported by Egan *et al.*¹⁰ "When white noise is used . . . articulation is affected only slightly by S/N ratios greater than +15 dB. Further increase in the level of the noise results in a very rapid decrease in articulation until, with a S/N ratio of -10 dB, articulation is practically zero." The output of the mixer was applied to a low-pass filter (Krohn-Hite 310 AB), with cutoff frequency 8000 cps (see below) and thence to a low-noise, high-fidelity earphone amplifier that supplied three binaural headsets in parallel.

Frequency selection was employed for response-independent distortion in the second study. The cutoff frequency of the transmission system was set at one of four values: 600, 1200, 2400, or 8000 cps. To obtain these cutoff frequencies, a sweep-frequency tone was recorded in place of the speech signals; the response of the earphone terminating the system was measured in a 6-cc coupler with a condenser microphone (Western Electric 644A) and graphic level recorder (General Radio, 1521-A); and a suitable adjustment in the nomi-

¹⁰ J. P. Egan, J. Miller, M. I. Stein, G. G. Thompson, and T. H. Waterman, "Studies on the Effect of Noise on Speech Communication," Psycho-Acoustic Lab., Harvard Univ. OSRD No. 2038 (1943).

⁶ H. A. Gleason, *An Introduction to Descriptive Linguistics* (Henry Holt & Company, Inc., New York, 1955).

⁷ S. S. Stevens and L. L. Beranek, "Word Lists for Articulation Testing," Psycho-Acoustic Lab. Harvard Univ. Rept. No. 1C-26 (1942).

⁸ J. P. Egan, "Articulation Testing Methods II," Psycho-Acoustic Lab., Harvard Univ. OSRD No. 3802 (1944).

⁹ The level of received speech was measured by impressing an equivalent sinewave voltage (measured on a Ballantine rms VTVM) on the listener's headphones and measuring the resultant sound-pressure level in a 6-cc coupler with a calibrated microphone and VTVM.

nal filter setting was made. The system had an essentially flat frequency response from the lowest speech fundamental up to the indicated cutoff frequency, whereafter signals were attenuated at the rate of 24 dB per octave.

RESULTS AND DISCUSSION

Figure 1 shows the effect of signal-to-noise ratio and of foreign accent on the percent word articulation. The intelligibility of the English speaker decreases at a rate which is comparable to that obtained by Egan *et al.*¹⁰ for much larger samples of speakers, words, and listeners. The disparity in the ordinate position of these two articulation curves may be attributed largely to a difference in the received level of speech: 50 and 115 dB (SPL), respectively. Egan *et al.*¹⁰ have shown that the percent word articulation is inversely related to the level of received speech at high intensities. Comparable rates of speech distortion as a function of masking were obtained for the foreign speakers, although their articulation scores are, at all points, about 36% below those for the English speaker. It is particularly interesting to note that there is no appreciable interaction effect due to the two types of distortion operating in concert. Furthermore, individual differences among the foreign speakers, especially with respect to national origin, had no marked effect on articulation scores.

Essentially the same relations among response-independent and response-dependent speech distortion are revealed in Fig. 2, which shows the effect of low-pass filtering and of foreign accent on percent word articulation. Once again, the articulation curves for the foreign speakers do not differ appreciably from each other, but lie, in general, about 36% below that curve for the American speaker. The articulation scores obtained with cutoff frequency 8000 cps constitute a replication of the first experiment with a second group of 12 listeners; corresponding means do not differ by more than 5%. As observed in Fig. 1, the two types of

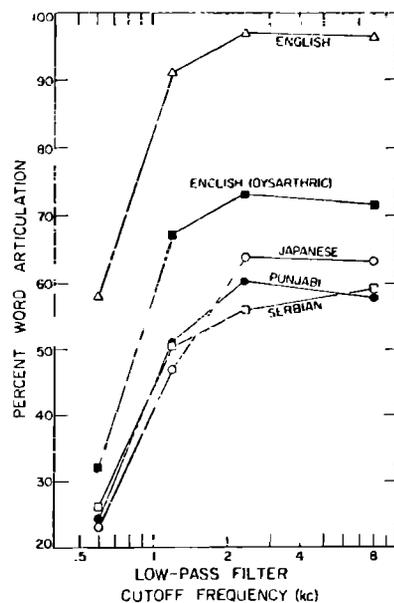


FIG. 2. Percent word articulation as a function of the high cutoff frequency of the transmission channel. Twelve listeners heard one native, one native dysarthric, and three foreign speakers read English monosyllables. Each point is the mean percent correct transcription for 2400 responses.

speech distortion do not interact in their effects on intelligibility. Articulation scores for a dysarthric speaker are presented for comparison (c.f. Tikofsky *et al.*³). Tape recordings of this speaker's rendering of English monosyllables were prepared and presented to the same group of listeners under comparable conditions.

ACKNOWLEDGMENTS

This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education. The assistance of Mr. K. Anderson, Mr. W. Watrous, and Miss A. Crabbs is gratefully acknowledged.