A STUDY OF RELATIVE EFFICIENCY OF ACOUSTIC PARAMETERS IN THE INTONATIONAL SIGNALS OF AMERICANS ENGLISH

YUKIO TAKEFUTA*

The theory of signal detectability has provided an interesting model for the study of psychophysical phenomena. The difference between the means of the two density functions expressed in terms of their standard deviation is defined as the index of signal detectability in this theory. Utilizing this index, one can quantify relative effectiveness of any measurable acoustic parameters as elements of certain phonetic or linguistic signals by plotting a frequency distribution of "the stimuli with a phonetic signal to be studied" and another distribution of "the stimuli without such a signal" on a scale of the measured acoustic parameter. In this study, the indices have been calculated to assess the relative effectiveness of 27 arbitrarily chosen measures of five acoustic parameters as possible elements of the phonetic signal of American intonation.

Sixteen native speakers of American English listened to 400 pairs of English sentences recorded by 20 native American speakers, and reported whether the paired sentences were read with the same or different intonations. The difference in the intonations perceived by native listeners between the paired sentences was considered the phonetic signal of the intonation in this study. Specifically, a pair of sentences which was heard by 12 or more of the 16 listeners to be different in intonations was treated as "a stimulus with an intonational signal." Conversely, a pair of sentences which was heard by 12 or more of the listeners to be the same was treated as "a stimulus without an intonational signal." From the initial 400 stimuli, two sets of 94 stimuli, the "SET SN" having an intonational signal and the "SET N" not having such a signal, were randomly selected and subjected to acoustic analyses of (1) pitch pattern, (2) pitch level, (3) intensity, (4) rate, and (5) the extent of pitch change. A measure of the signal strength of each stimulus in an acoustic parameter was made by calculating the average of the differences of the measure obtained by comparing the paired sentences syllable by syllable.[1]

Several problems had to be resolved in the process of data reduction. The two

*The Ohio State University, Columbus, Ohio, U.S.A.

[1] Syllables were defined in this study as the equal time segments obtained by dividing the time spent in reading the total sentence by the number of the actual syllables defined phonetically.

rules proposed by Witting[2] (two "dimensional rules" concerning duration and frequency) and the data on the threshold of pitch modulation[3] expressed in terms of the direction, duration, and the extent of frequency modulation were used to simplify and schematize the pitch pattern. Quantification of the schematized pitch pattern was made by assigning a number 3, 2, and 1 to a rising, level, and falling pattern respectively.

Data reduction with regard to the measurement of the pitch level and the intensity was attempted from two directions: first, by sampling only one representative value

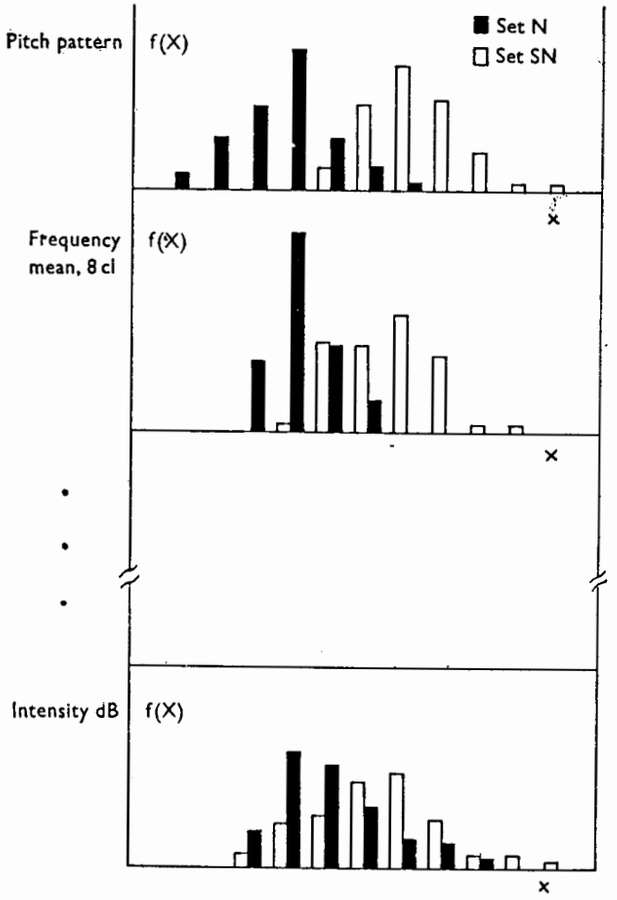

Fig. 1. Frequency distributions of the two sets of stimuli: SET SN and SET N, plotted on the scale of each physical measurement.

either by measuring the center of each syllable or by taking the average of all the raw data of the syllable; second, by making the units of contrastive scale more gross by dividing the total range of pitch or intensity of one speaker into equal contrastive scale units of 8, 4, or 2. The second method of data reduction was used also for the reduction of the data of duration and the extent of frequency change. By testing several different methods of data reduction in measuring each acoustic parameter 27 different physical measures of the signal strengths were obtained of the two sets of stimuli: SET SN and SET N. The frequency distributions (histograms) of the signal strengths for the SET SN and for the SET N were plotted on the scale of each of these 27 physical measurements (Figure 1). Assuming the normality in these distributions, 27 indices of signal detectability, in turn, were obtained from the information of the difference in the means and the standard deviation of the paired distributions. The indices varied from 0.1 to 2.1 (Table 1). No individual measures were as good as the average signal detectability calculated from the signal detection of an individual native American listener.[4] However, the combined infor-

Parameter	Fundamental Frequency (mean)					Fundamental Frequency (center)				
Scale	Hz	8 cu	4 cu	2 cu		Hz	8 cu	4 cu	2 cu	
Index of Detectability	1.1	1.4	1.2	0.7		0.8	1.1	1.1	0.9	
Parameter	Extent of Frequency Change					Rate of Speaking				
Scale	Hz	8 cu	4 cu	2 cu		msec.	8 cu	4 cu	2 cu	
Index of Detectability	0.5	0.7	0.6	0.6		0.8	0.4	0.1	0.2	
Parameter	Intensity (mean)					Intensity (center)				
Scale	Ratio	8 cu	4 cu	2 cu	dB	Ratio	8 cu	4 cu	2 cu	dB
Index of Detectability	0.3	0.2	0.1	0.1	0.5	0.7	0.2	0.1	0.1	0.2
Parameter	Pitch Pattern					Pitch Pattern and Fundamental Frequency				
Scale	3 cu					3 cu	and		8 cu	
Index of Detectability	1.2					2.5				

Table. 1. Summary of 28 indices of signal detectability obtained from 27 measures of five physical parameters and from a combined data of pitch pattern and pitch level (*CU* stands for the contrastive units).

[2] C. Witting, "A Method of Evaluating Listeners Transcriptions of Intonation on the Basis of Instrumental Data," *Language and Speech*, 5, 1962, 138—150.

[3] Y. Takefuta, "Perception of Frequency Modulation," Paper prepared for the 74th Meeting of the Acoustical Society of America, Miami Beach, Florida, U.S.A., 1967.

[4] Y. Takefuta, "A Study of Relative Efficiency of Acoustic Parameters in the Intonational Signal of American English," unpublished Ph. D. dissertation, The Ohio State University, Columbus, Ohio, U.S.A., 1966. The average signal detectability of a native American listener in the detection of an intonational signal in American English was 2.4.

mation of the most efficient physical measure (pitch pattern) and the second most efficient measure (pitch level) elicited a detectability index of 2.5 which was about the same as the signal detectability of an average native listener.

From the results of the present study, it was found that the order of relative efficiency of five acoustic parameters as possible intonational signals were (1) pitch pattern, (2) pitch level, (3) extent of pitch change, (4) rate of speaking, and (5) intensity. Efficiency of any acoustic parameter changed considerably by using different methods or units of measurement. The combined information of pitch pattern schematized according to the rules based on the findings of psychophysical experiments and the pitch level measured in eight contrastive units seems to comprise the primary phonetic signal of intonation, perceived by average American listeners.

REFERENCES

Cowan, J. M. "Graphical Representation of Perceived Pitch in Speech," *Proc. 4th Int. Cong. Phon. Sci.*, Helsinki, 1961, 567—570.

Sidowski, J. B. (ed.) *Experimental Methods and Instrumentation in Psychology.* New York: McGraw-Hill, 1966.

Swets, J. A. (ed.) *Signal Detection and Recognition by Human Observers.* New York: John Wiley & Sons, 1964.

Takefuta, Y. "A Study of Relative Efficiency of Acoustic Parameters in the Intonational Signal of American English." Unpublished Ph.D. dissertation, The Ohio State University, Columbus, Ohio, U.S.A., 1966.

Takefuta, Y. "Perception of Frequency Modulation" (Paper prepared for the 74th Meeting of the Acoustical Society of America, Miami Beach, Florida, U.S.A., 1967).

Witting, C. "A Method of Evaluating Listeners Transcriptions of Intonation on the Basis of Instrumental Data," *Language and Speech*, 5, 1962, 138—150.

ACKNOWLEDGMENT

This research was supported by a contract between the Office of Naval Research and The Ohio State University Research Foundation, Nonr 495(18).

DISCUSSION

Black:

These are very interesting results. I would appreciate your telling us how the interspeaker differences in pitch were controlled.

Takefuta:

ad Black: The interspeaker differences in pitch were controlled by subtracting the median pitch of each speaker from all the measurements before comparing them.