63-4-1

AFCRL-63-119

SPEAKER RECOGNITION

Gary L. Holmgren

TEXAS INSTRUMENTS INCORPORATED
6000 Lemmon Avenue
P. O. Box 6015
Dallas 22, Texas

FINAL REPORT

Contract No. AF19(628)-345
14-73801-14
Project No. 4610
Task No. 461002

May 1963

Prepared for
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

Requests for additional copies by agencies of the Department of Defense, their contractors, and other Government agencies should be directed to the:

DEFENSE DOCUMENTATION CENTER (DDC)
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA.

Department of Defense contractors must be established for ASTIA services or have their "need to know" certified by the cognizant military agency of their project or contract. All other persons and organizations should apply to the:

US DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
WASHINGTON 25, D. C.

AFCRL-63-119

SPEAKER RECOGNITION

Gary L. Holmgren

TEXAS INSTRUMENTS INCORPORATED
6000 Lemmon Avenue
P. O. Box 6015
Dallas 22, Texas

FINAL REPORT

Contract No. AF19(628)-345
14-73801-14
Project No. 4610
Task No. 461002

May 1963

Prepared for
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

ABSTRACT

This research is concerned with defining a perceptual space within which listeners locate voices, to the end that the effects of manipulating speaker, hardware, and listener characteristics can be measured, and eventually, that specifications for elements of the communication system can be prepared to produce the desired system characteristics. In the experiments taped speech samples were rated by listeners using Osgood's semantic differential method. Previous study indicated only four basic dimensions were required to account for ratings given speakers on a large number of characteristics. In a second experiment, a reduced number of characteristics, selected from the original list as best representing the four necessary factors, was used by listeners to rate speakers from AFCRL's speaker library. The experimental design allowed examination of the effects on ratings due to differences between listeners, due to repetition of the rating task, and to order of speaker presentation. Results of these examinations and the following are presented:

The adequacy of original factors to account for listeners' ratings

The differentiation between speakers.

The reliability of ratings

The familiarity of previously unheard voices.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

SPEAKER RECOGNITION

by

Gary L. Holmgren

SECTION I

INTRODUCTION

This report summarizes research conducted at Texas Instruments, Apparatus Division, under Contract AF19(628)-345, Item II. The objective of this portion of the contract was to investigate methods of classifying and categorizing speech samples in terms of subjective factors.

Recent work in speech processing devices emphasizes the need for accurate and reliable definitions of the minimum signal requirements for adequate listener responses, not only in terms of the intelligibility of the speech material, but also with respect to the recognizability of the speakers' voices. This emphasis has arisen out of interest in developing speech processing systems that meet a given fidelity criterion (e. g., intelligibility or recognizability of the speaker) while permitting increases in bandwidth compression by sacrificing some faithfulness in reproduction of of speakers' inputs.

Perhaps the more obvious characteristics that make voices recognizable are larynx frequency, accent, rate of speaking (Goldman and Eisler[1]) and speech idiosyncrasies such as hesitancy. However, even among speakers in whom most of these features are similar, recognition is still often possible on the basis of the property identified as "voice quality."

Ladefoged and Broadbent[2] have reported that some samples of synthetic speech, differing only in frequency range, "sounded like the same sentence pronounced by people who had the same accent, but differing in their personal characteristics." There are, conceivably, several other features of the speech spectral envelope which might affect "voice quality," such as bandwidths of formants, relative amplitudes of formants, and possibly the shape of the spectrum between the main formant peaks.

The present research is concerned with investigating the ability of listeners to discriminate among speakers on the basis of how they perceive the speakers' voices.

---

[1] See bibliography, Section VI.

SECTION II

BACKGROUND

Although the ability to recognize or identify a speaker on the basis of hearing his voice is generally acknowledged, the accuracy of the identifications and the factors upon which they are based have seldom been investigated. Of the generally accepted variables in speech (pitch, volume, duration, quality, and articulation), quality has been considered the most influential by a number of writers in the speech field.

As early as 1922, Miller[3] stated that "quality is, by psychological definition, the distinguishing characteristic of vocal and instrumental sounds of identical pitch, loudness, and duration." The distinguishing characteristics are thought to be determined by the harmonic composition of the initial vocal fold tone and by the modifications effected by resonance.

It has been stated or implied in several speech texts that it is this distinguishing characteristic, quality, which enables a listener to recognize or identify a speaker's voice. According to Anderson,[4] "It is . . . quality that enables us to distinguish the voice of Jim from the voice of Fred, even though both may speak with similar pitch and inflectional patterns." Curry[5] also states that ". . . the factor of quality . . . forms the basis of the recognizable meaning of the words and conveys the individuality of the speaker." Gray and Wise write:

> "No two voices sound exactly alike. Even when we cannot see the faces of our friends we are usually able to identify their voices, much as we are able to distinguish the tones of mucical instruments. Voices are different primarily because of difference in timbre—and differences of timbre result from differences in the blend of overtones."[6]

Similarly, Judson and Weaver[7] state that, "voice quality varies so greatly among individuals that we may rely upon it as a means of identification when we are unable to see the person who is talking."

However, most of these statements are assumptions not verified experimentally. In spite of the fact that there seems to be considerable agreement that there are qualities of speakers' voices that listeners perceive and rely on to differentiate among speakers, relatively few studies have attempted to determine just how recognition takes place and what cues listeners employ to differentiate among speakers' voices.

Perhaps the earliest study of speaker recognition reported in the psychological literature was that of McGehee.[8] She was primarily concerned with the listener's ability to identify a voice he had heard once before when presented with four unfamiliar voices. She concluded, tentatively, that recognition was reduced, not only by lengthening the time intervals between the

judgements and the original presentation of the voices, but also by increasing the number of voices in the series and by disguising the pitch of the speakers' voices. It was concluded that men surpassed women in voice recognition ability.

In a followup study, McGehee[9] secured 30 judgements pertaining to the "unlikeness" and "agreeableness" of five recorded male voices. She found no general agreement among the judges, on uniqueness, from which she inferred that many factors, including pitch, rate, and quality, are involved in voice recognition.

The effects of five factors of speaker recognition were investigated by Pollack, Pickett, and Sumby.[10] The factors examined were:

The size of the class of possible voices

The duration of the speech signal

The frequency range of the speech signal

The voicing and nonvoicing speech characteristics

The simultaneous presentation of several voices.

As in McGehee's study, male voices with no pronounced speech defect or accent were used. However, unlike her study, the 16 speakers' voices were familiar to the seven listeners. Also, a list of phonetically balanced (PB) words, rather than connected speech, were used to minimize inflectional and rate cues. Volume was controlled through the recording process, and pitch cues through the use of whispered speech. The results indicated that duration was the most influential factor. This was true, however, "only insofar as it admits a smaller or larger statistical sampling of the speaker's speech repertoire."

In terms of information transmission measures, Pollack, et al., found that the information transmitted increased with the size of the class of possible voices. Identification was resistant to selective frequency emphasis using both high- and low-pass filters. Approximately 75-percent correct identification was obtained for eight voices, with the high-pass and low-pass filters set at 500 cps, indicating that the frequency spectrum of the voices may not have been as important an identification factor as some have supposed. They further found that a whispered sample three times the duration of the voiced samples was necessary for comparable identification. However, the duration required for approximately 95-percent information transmission was only 3.4 seconds.

In an effort to study more closely the ability of listeners to identify a speaker by voice, Peters[11] studied the effects of certain restrictions imposed on the voice signal. These restrictions included

High-pass, low-pass, and octave-bandpass filtering of the voice signal

The altering of the relative sound pressure level of the voice signal

The masking of the voice signal by noise.

Peters found that a decrement occurred in correct identifications of the speakers with increasing amounts of the signal rejected through progressive high-pass or low-pass frequency filtering. For the octave bands considered, maximum correct identification of voices occurred when the voice signal was presented at a relatively low signal level. Correct identification of the speaker by the listener decreased as the signal-to-noise ratio of the masking noise was decreased in the range from +8 to −8 signal-to-noise ratio.

In a second study, Peters[12] made the additional evaluations of the effects of interruption of the signal at known rates and the addition of octave frequency bands of the signal to the original voice signal upon the ability of listeners to correctly identify speakers' voices.

The effect of the alterations of the original voice along with the alterations included in his original study led to the following conclusions.

The relative level of the voice signal affects the listener's ability to identify speakers. A 6-decibel change in level, either an increase or decrease from a standard level, is effective in lowering identification. This finding suggests the importance of a perceived dimension of loudness.

Short time interruptions of the voice signal decrease the listener's ability to identify speakers' voices correctly. This suggests possible importance of a perceived dimension of rhythm.

The addition of octave frequency bands to the original signal, especially the octave band that contains the fundamental of the voice, significantly aids the listener in identifying the speakers' voices.

The limiting of the voice signal through high-pass or low-pass frequency filtering reduces the listener's ability to correctly identify speakers' voices. This finding conflicts with the Pollack, et al., finding using high-pass and low-pass filters centered at 500 cps; however, Peters did not use the filters simultaneously and unlike Pollack, et al., Peters used sentences rather than PB words as the speech material. These last two conclusions suggest the importance of a perceived dimension of pitch.

Skalbeck[13] has investigated the relative influence of several factors in speaker recognition. The factors she studied were pitch and inflectional patterns, articulation and pronunciation characteristics, and voice quality.

Six experimental conditions were designed to control or distort these factors. The control condition consisted of a normal reading of a prose passage, in which no distortion was imposed. Pitch and inflection were distorted by having the passage read in a monotone. Articulation and pronunciation were distorted by playing the recording backward. "Voice quality" was distorted by low-pass filtering. One fact that tends to confound the results is that the ten speakers were familiar to the listeners through previous daily contact, and eight of the speakers served also as listeners. The results were reported as follows.

> There was a low correlation between the speakers' predicted recognition rank (based on prior listener ratings) and their experimental score.
>
> The eight speakers who were also listeners made significantly more errors in identifying their own voices than in identifying the voices of the other speakers.
>
> Male listeners made fewer identification errors than did female listeners, although the difference was not statistically significant.

One interesting finding was that recognition was impaired more by the filtering than by the backward reproduction.

Black and Dreher[14] have been concerned with messages other than those carried by the definition of words; i. e., they sought to determine whether the listener could identify the speaker by his voice, recognize the voice as a man's or woman's, ascribe an emotional state to the voice, etc. The problem confronting them was that these "extra" messages may be restricted to personal interpretations varying from listener to listener, or they may have a similar meaning among most listeners, as is generally supposed. If general meanings are to be interpreted from voice, listeners would need to agree on a "normal voice" from which deviations would denote special meanings.

They found that when recorded voices were distorted by altered turntable speed, listeners were able to return the voices to the original speed with standard deviations of 1.4 rpm. The judgemental responses of the subjects to the readings of untrained speakers indicated that intended characterizations (e. g., certainty-uncertainty) were identified through vocal characteristics apart from the verbal content of the messages.

Howell,[15] in connection with a vocoder development study, has made an attempt to determine the extent to which speaker recognition varies as the speech is presented via telephone and various vocoder conditions. The method employed is that discussed by Surgent and Yost[16] at the sixty-first meeting of the Acoustical Society of America. The two methods discussed are the transfer method and the recognition method. The latter was used in Howell's study. The subjects were trained on a reference telephone circuit, then they heard the same voices speaking a previously unheard test sentence over each of the

various vocoder systems. Five previously unheard voices were interspersed among the previously identified speakers (i.e., those on which the listeners were trained) in the rating task. The standard paired-associate learning method was employed in the training task, whereas in the testing tasks the listeners identified the speakers' voices by name and checked (✓) if the voice was familiar (i.e., one of the voices on which they were trained) or (o) if the voice was not familiar (i.e., one of the previously unheard voices interspersed in the testing task). The results of the recognition task were reported as follows as speaker recognition test scores expressed as percentages of maximum possible scores.

	Subjects			
	Group I	Group II	Group III	Three-Group Average
Reference telephone*	75.0	77.1	70.0	73.9
Hybrid vocoder	60.0	71.4	52.5	60.9
27-channel vocoder	25.0	54.3	32.5	36.5
22-channel vocoder	37.5	51.4	22.5	36.5
17-channel vocoder	32.5	60.0	35.0	41.7

*The reference telephone circuit was band-limited only by the attenuation characteristics of the nonloaded cable.

An analysis of variance of these data after an arc-sine transformation revealed that all comparisons for the transmission systems are significant except for the differences between the three channel vocoders. The three listener groups were found to be significantly different, which indicates non-homogeneity of listener variance. Since the three vocoder systems were not significantly different, one concludes that the recognition test employed lacked the sensitivity required to show differentiations among vocoder processing effects on speaker recognition.

Further tests, however, revealed that there is no simple effect on recognition test scores due to the training task (i.e., either telephone or vocoder) and the testing task (i.e., either vocoder or telephone). Out of the possible score of 40, the following scores were obtained on the various telephone-vocoder combinations.

	Telephone Tested	Vocoder Tested
Telephone trained	37/40	24/40
Vocoder trained	27/40	20/40

Relative to the particular vocoders employed and the recognition method the following conclusions were made

> Listeners apparently learn to recognize speakers on a vocoder; however, the learning process takes approximately three times longer. Thirty-six or more sentences may be required for vocoder learning as compared to 12 or fewer for telephone learning.
>
> When listeners are trained using one system but tested on the alternate, scores are lower.

A possible refinement that would probably contribute to a method such as Howell's, would be to provide more complete counterbalancing of the voice processing techniques in both the training and testing tasks. A covariance analysis could then be employed to assess the extent to which differences obtained in the testing task are not attributable to differences in the training task. Such an approach can indicate the extent to which recognition varies with processing methods, but it does not yield much information as to how recognition takes place or what aspects of the speakers' voices are influenced by voice processing

Recently, Shearme and Holmes[17] have studied speaker recognition by employing short recorded passages of disconnected discourse. In this study the speech signal was treated in two ways:

> Simple passage through a vocoder to equalize basic speech frequency
>
> Displacement of the relative position of the formants.

Samples of the discourse were matched in various combinations of the same and different speakers, of the two types of treatment, and recorded in pairs on two tracks of tape. The listener was required to judge the two tracks as the same or different speakers.

The authors reported contrary to the observations of Howell[15] that "simple passage of the speech through a vocoder did not affect the recognition of speakers. The second treatment destroyed recognizability though it left intelligibility intact."

McGee[18] has recently investigated the possibility of determining perceptual spaces for the quality of filtered speech. As he used only one speaker under several conditions, his study is limited in the extent to which the results can be generalized. However, he did find that perceived quality depends on judgements of "naturalness" and intelligibility, and that the "naturalness factor" is most significantly related to the presence or absence of the fundamental frequency of the speaker's voice.

Williamson[19] has investigated several factors that affect the ability of listeners to identify speakers' voices as the same as or different from preceding voices. She concludes that there is much variation among listeners in the

ability to judge short speech samples as being spoken by the same or different persons and that training in phonetics does not appear to influence the listeners' recognition of speakers' voices.

Meeker and Nelson, [20] in a performance evaluation of vocoder systems, have compared various vocoder processings on the basis of intelligibility, voice quality, and speaker recognition. The voice processing equipment evaluated were as follows:

Identification Label	Type of Processing Equipment
A	Filtered 400-20000 cps (no sharp cutoffs)
B	Telephone (transmitted over a 30-mile loop)
C	18-channel vocoder, analog connection, normal pitch
D	18-channel vocoder, analog connection, lowered pitch
E	18-channel vocoder, digital connection, lowered pitch
F	Eight-channel vocoder

The recorder messages from the speakers were processed by AFCRL.

The rating tasks had two basic objectives: one was directed toward the rank-ordering of the system, the other was to provide an estimate of the adequacy of each type of processing. For our purpose we are concerned with the part of the tasks that dealt with the assessment of quality and speaker recognition. In the first task, the listeners rated the speakers' voices on the various systems using the following format.

Speech Quality	Talker (Speaker) Recognition
a. Voices sounded natural.	a. Had no difficulty distinguishing between talkers
b. Voices were noticeably distorted but distortion was not objectionable.	b. Could not determine immediately who was talking but might with careful listening
c. Voices sounded unnatural and distorted.	c. Believe I would have difficulty recognizing who was talking even after extended use.

The subjects rated the speakers' voices on the various systems by simply writing a, b, or c under Speech Quality and a, b, or c under Talker (Speaker) Recognition. In the task intended to provide an estimate of the adequacy of each type of processing, the judgements were as follows.

Speech Quality	Talker (Speaker) Recognition
a. Better than needed	a. Better than needed
b. Suitable for normal use	b. Suitable for normal use
c. Usable but not entirely satisfactory	c. Usable but not entirely satisfactory
d. Unsatisfactory	d. Unsatisfactory

Again the listeners indicated their impressions by marking the appropriate letter under Speech Quality and Talker (Speaker) Recognition.

The following ratings were obtained using 58 listeners' average weighted judgements for the first task on the various systems (A, B, C, D, E, F) over all talkers.

Speech Processing System	Ratings	
	Speech Quality	Talker Recognition
A	2.97	2.87
B	2.68	2.83
C	2.42	2.70
D	1.97	2.10
E	1.70	2.22
F	1.46	1.86

In scoring the ratings, a, b, and c equaled 3, 2, and 1 respectively.

The results of the second task are as follows. The scores were obtained by assigning the values 4, 3, 2, and 1 to responses a, b, c, and d respectively.

Speech Processing System	Ratings	
	Speech Quality	Talker Recognition
A	3.48	3.49
B	2.88	2.95
C	2.51	2.48
D	2.23	2.27
E	1.92	2.12
F	1.43	1.49

These results indicate, generally, that judgements on quality and recognizability were inversely related to the amount of processing. It may be seen that the telephone Sample B is generally considered "satisfactory for normal

use" and that the best vocoder sample is rated between "satisfactory for normal use" and "usable but not entirely satisfactory." It is considered appropriate to ask whether a rating of the extent to which the listener believes the voice is recognizable would correspond directly to an actual task of recognizing the speaker's voice (i.e., there may be a difference between thinking a voice is recognizable and actually recognizing who is speaking by name).

This survey of the literature indicates the ways investigators have attempted to study speaker recognition. From this information we can make the the observation that in speaker recognition (or identification) the listener is capable of selecting from a given speech sample various combinations of cues upon which he bases his judgements. If enough cues are in the speech sample, the listener not only can understand the content of the text (intelligibility) but also can recognize or identify the speaker. However, if cues are progressively reduced (i.e., degradation of the speech signal through filtering or digitizing operations), the listener is unable to recognize the speaker, and the speech sample soon becomes unintelligible, which results in a breakdown of communication. In effect, we observe an inverse relationship between recognizability and the extent to which the speech sample has been processed (degraded).

If this is true, as the above studies indicate, then it follows that more must be known about how the listener perceives various speakers' voices. Our purpose in this research was to determine if it is possible to develop a technique to determine how listeners differentiate among speakers' voices on the basis of perceived voice characteristics. To determine how listeners perceive the voices, a semantic differential rating form was employed (Osgood[21]).

It was hypothesized that this method would permit catagorization of the speech samples (speakers' voices) and measurement of difference between samples in terms of subjective factors (perceived voice characteristics). If the method is successful it will provide the objective measurements needed for evaluations of

Speech processing devices (i.e., in making specifications for a fidelity criterion)

Relations between intelligibility and physical characteristics of the speech samples

The effects of such variables as training, dialect, and procedures on the intelligibility and recognizability of speakers.

SECTION III

DEVELOPMENT OF SEMANTIC DIFFERENTIAL RATING FORM EMPLOYED IN PRESENT EXPERIMENT

A series of experiments was conducted to develop a technique to measure the extent to which listeners differentiate among speakers on the basis of perceived voice characteristics. In these experiments, taped speech samples were rated by listeners using Osgood's[21] semantic differential methods. Several investigators have used the semantic differential and fount it well suited for stimulus classification (Elliot and Tannenbaum,[22] Lichte,[23] Peters,[24] and Uldall[25]).

A semantic differential is a set of adjectives specially selected to represent a perceptual domain. The adjectives are arranged in pairs of words having opposite meaning (e.g., hot-cold) with a seven-point scale between. A subject (listener) marks this scale to indicate correspondence between his perception and the descriptive terms (items). In the preliminary experiments, we found that the ratings, first obtained on a large number of characteristics (49 items in Form II and then 20 items in Form III) could be accounted for or described by only four factors. These factors were identified by a factor analysis of the item correlations obtained from both Forms II and III. In the present experiment, a reduced number of characteristics (Form IV), selected from the two original lists of adjuctive pairs as best representing the four factors, was used by ten listeners to rate ten speakers from AFCRL's speaker library.

The factors isolated from the earlier Forms II and III are found in Tables I and III respectively. Table II contains the item correlation matrix on which the square-root factor analysis was conducted, the summary of which comprises Table III. The data in Table I was arrived at by the same method, using the item correlation matrix based on Form II. The factors isolated from these two forms in Tables I and III were then compared as to their similarity. This evaluation was conducted to arrive at a selected set of items, capable of efficient measurement of the principal dimensions, to be included in Form IV. The criteria for selecting the items for Form IV, relative to the analysis of Forms II and III, were

The items having the highest factor loading on a single factor

The purity of the item factor loadings

The extent to which the factor loadings on the items were similar in the analysis of both forms (Form II and Form III)

The communality (h^2) of the item.

Table I. Summary of Rotated Orthogonal Factor Analysis of Initial Item Correlation on Form II*

	Item	I	II	III	IV	h^2
1	Loud–Soft	8701	-1252	2699	-2528	9098
2	Heavy–Light	3357	-1677	8979	-0748	9528
3	Beautiful–Ugly	-0422	9185	0802	-0730	8572
4	Clear–Hazy	7795	5425	0591	-2059	9479
5	Belligerent–Friendly	8261	-4036	-1342	0403	8651
6	Tense–Relaxed	8446	-3390	-2803	2252	9577
7	Familiar–Strange	-3226	6006	-0605	-2140	5144
8	Colorful–Colorless	6977	4681	4628	1861	9549
9	Cool–Warm	6317	-4027	-3842	0584	7124
10	Rising–Falling	7754	-0872	-3813	3175	8552
11	Large–Small	4070	2164	8632	-0357	9590
12	Pleasant–Unpleasant	-3213	7864	4362	0736	9174
13	Definite–Uncertain	7564	4377	3254	0543	8726
14	Violent–Gentle	9258	-2352	2364	-0012	9685
15	Tight–Loose	8246	-1972	-3572	3374	9603
16	Wet–Dry	-3181	1944	2262	3343	3020
17	Rich–Thin	0663	4296	8875	0368	9781
18	Sharp–Dull	9438	1515	-0166	-0021	9141
19	Masculine–Feminine	4199	-1470	8641	-0494	9473
20	Rumbling–Whining	0997	-0754	9429	0873	9125
21	Good–Bad	0471	8750	4336	0805	9625
22	Uneven–Even	7470	-5496	0343	-0063	8614
23	Exciting–Calm	8992	-2592	-0506	2420	9370
24	Hard–Soft	9713	-1515	0944	-1072	9869
25	Active–Passive	9476	1050	1865	1818	9770
26	Happy–Sad	8910	2566	0825	2057	9090
27	Rugged–Delicate	6555	-0891	7213	-1326	9756
28	Fast–Slow	7911	-0462	-0698	5008	8838
29	Wide–Narrow	-3023	2674	8311	0811	8604
30	Pleasing–Annoying	-3783	7257	4675	1357	9068
31	Concentrated–Diffused	9009	0238	2339	-1585	8921
32	Reassuring–Disturbing	-5443	6366	3629	-3165	9336
33	Agitated–Serene	8775	-4001	-0832	2215	9862
34	Steady–Fluttering	-1891	7936	1047	-3769	8189
35	Deliberate–Careless	5691	6699	0506	1874	8104
36	Gliding–Scraping	-6754	6593	-0492	0459	8956
37	Easy–Labored	-4805	7838	1229	0623	8643
38	Low–High	-3721	0314	8819	0432	9193
39	Smooth–Rough	-6709	6150	-2988	2180	9653
40	Obvious–Subtle	7374	-1930	1944	-2213	6679
41	Complex–Simple	8110	1946	2795	3739	9137
42	Intense–Mild	9823	-0000	1315	-0727	9876
43	Foreign–Native	7555	0600	4077	1909	7771
44	Full–Empty	3717	3256	8323	0509	9397
45	Powerful–Weak	6591	1734	7186	-0470	9832
46	Deep–Shallow	3066	1263	9130	-0030	9437
47	Busy–Resting	9142	-0702	-0522	3425	9608
48	Varied–Repeated	8363	3331	0773	0385	8178
49	Clean–Dirty	1668	8401	-1681	-1167	7755

*Decimals omitted.

Table II. Correlation Matrix* of Ten Items Over Each of 16 Speakers Using Semantic Differential Form III
(Dependent variable—average of five listeners' judgements over three trials)

Item	1	2	3	4	5	6	7	8	9	10
1	10000	-2538	6769	-5195	-7518	6903	5164	1482	2203	6599
2		10000	-5686	-0999	4944	-4618	0976	7173	-6454	-6437
3			10000	-3726	-6430	8437	3459	1217	4643	8420
4				10000	5308	-6359	-7997	-2351	1022	-2456
5					10000	-6216	-4659	2377	-3283	-7376
6						10000	4690	-1318	3357	7682
7							10000	1890	-0471	1022
8								10000	-6568	-3527
9									10000	5204
10										10000

*Decimals omitted.

Table III. Summary of Square-Root Factor Analysis* of Ten Items Over Each of 16 Speakers Using Semantic Differential Form III
(Pivot variables chosen in order of results of factor analysis of Form II; e.g., variable 2, 5, 4, and 1 respectively.)

	Item	I	II	III	IV	h^2
1	Fast—Slow Active—Passive	2538	7205	0866	4372	7823
2	Loud—Soft Intense—Mild	10000	0000	0000	0000	10000
3	Clear—Hazy Sharp—Blurred	5686	4163	2053	2297	5916
4	High-Pitched—Low-Pitched Shallow—Deep	-0999	6675	7379	0000	10000
5	Fluttering—Steady Uneven—Even	4944	8692	0000	0000	9999
6	Shrill—Muffled Bright—Dark	4618	4524	5150	2165	7301
7	Thin—Rich Nasal—Resonant	-0976	5915	5355	0733	6515
8	Rough—Smooth Harsh—Mellow	7173	-1345	-0999	-2402	6003
9	Rigid—Limp Unyielding—Yielding	6454	0106	-0607	0579	4237
10	Colorful—Colorless Dynamic—Monotonous	6437	4824	-0164	1607	6732

*Decimals omitted.

Table IV. Items and Factors Represented Selected for Form IV

Item	Factor Represented
Intense—Mild	I
Hard—Soft	I
Sharp—Dull	I
Beautiful—Ugly	II
Good—Bad	II
Clean—Dirty	II

Item	Factor Represented
Whining—Rumbling	III
Shallow—Deep	III
High—Low	III
Fast—Slow	IV
Complex—Simple	IV
Busy—Resting	IV

Items that satisfied these criteria were designated as marker variables. Three marker variables (items) were selected for each of the four factors. The items selected and the factors they represent are in Table IV.

The items selected were randomly assigned numbers from 1 to 12 to determine their positions on the new form. The polarity of descriptive terms in each of the items was then randomly determined. By randomly assigning the items and item polarity, the need for additional item orders (forms) was obviated. Since each factor was represented by three items, the effect of item position relative to listener judgements was minimized. Inspection of the new Form IV (Figure 1) indicates that the marker variables for each of the factors are evenly distributed throughout the form.

Speaker Number_____

Listener_____

VOICE RATING FORM IV

Instructions: Place an "X" in the box which corresponds to your perception of the speakers voice you are rating.

1.	Simple	___	___	___	___	___	___	___	Complex
2.	Slow	___	___	___	___	___	___	___	Fast
3.	Beautiful	___	___	___	___	___	___	___	Ugly
4.	Low	___	___	___	___	___	___	___	High
5.	Shallow	___	___	___	___	___	___	___	Deep
6.	Dirty	___	___	___	___	___	___	___	Clean
7.	Dull	___	___	___	___	___	___	___	Sharp
8.	Good	___	___	___	___	___	___	___	Bad
9.	Rumbling	___	___	___	___	___	___	___	Whining
10.	Hard	___	___	___	___	___	___	___	Soft
11.	Resting	___	___	___	___	___	___	___	Busy
12.	Intense	___	___	___	___	___	___	___	Mild

7868

Figure 1. Voice Characteristic Rating Form IV

SECTION IV

METHOD

The following method was employed in evaluating the new Form IV. Form IV was used by ten listeners to rate ten speakers from AFCRL's speaker library. The experimental design provided for replication over three days, the order of speaker presentation was counterbalanced, and a familiarization rating task was employed before each of the voice characteristic rating tasks.

A. SPEAKERS

Ten male speakers with no obvious speech defects, were selected from the AFCRL speaker library. The selection was based on controlling for sex and place of residence. The speakers were randomly assigned numbers from one to ten to determine the order of presentation for the two rating tasks. The two orders of presentation were as follows.

Order	Speakers
B	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
B'	10, 9, 8, 7, 6, 5, 4, 3, 2, 1

1. Familiarization Rating Task

In this task the speakers' voices were presented in Order B. The presentation sequence went as follows. First, the speaker's number was announced: "Speaker number one." This was followed by a five-second silent period, after which the speaker began the familiarization rating text, which lasted about 45 seconds. This was followed by a ten-second silent period (during which the listeners rated the speaker's voice on familiarity) followed by the announcement of the next speaker: "Speaker number two," etc.

2. Voice Characteristic Rating Task

Both orders of presentation were used in this task in an effort to control for the effect of speaker position on the listeners' ratings. On the first trial of each testing day, half of the subjects (Group I) heard and rated Order B and the other half (Group II) heard and rated Order B'. On the second trial the listeners heard and rated the reverse presentation. Thus, Group I heard Order B for the first trial and Order B' for the second trial while Group II heard Order B' first and then Order B for the second trial. This was repeated for three days. The presentation sequence for the voice characteristic rating task for both orders (B and B') was as follows. First the speaker's number was announced "Speaker number ____;" this was followed by a ten-second silent period, after which the speaker began the voice characteristic rating text, which lasted about 75 seconds. Then there was a 20-second

silent period (during which the listeners finished their ratings and prepared for rating the next speaker's voice) followed by the announcement of the next speaker: "Speaker number _____," etc.

The speakers employed in this experiment are cataloged as follows in the AFCRL's speaker library.

Experimental Number	AFCRL Identification
1	V0002
2	V0014
3	V0019
4	T0101
5	V0048
6	V0030
7	V0038
8	V0037
9	V0046
10	T0104

B. SPOKEN MATERIAL

There were two sets of spoken material for each of the ten speakers. The first set was used in familiarizing the listeners with the various speakers' voices (familiarization rating text); the second set was employed so that the listeners could rate the various speakers on the basis of perceived voice characteristics (voice characteristic rating text).

1. Familiarization Rating Text

While this text was being read, the listeners rated each of the ten voices on a familiarity scale. This rating task will be described in detail in a following section. The text selected for the familiarization ratings was AFCRL "Selection IX" which is as follows.

> "I am going to describe briefly for you an emergency that arose in one of our large cities a few years ago. It is the kind of situation which when it occurs, generates much comment and calls for thoughtful, critical listening in order to describe what position one wants to take.

"The emergency arose in Detroit. The bus drivers went on strike and the public transportation system broke down almost completely. Many factories and businesses had to close because employees coule not get to work. Even the social life of the city was disrupted. Under these conditions, feeling ran high and there was much argument pro and con as to whether such strikes should be allowed."

2. Voice Characteristic Rating Text

The second set of spoken material selected for this experiment was employed so that the listeners could rate the various speakers on the basis of perceived voice characteristics. The text selected was a portion of AFCRL "Selection VI" which is as follows:

"What I wish to do today is illustrate the semantic changes which occur in language—to make you more aware of the ambiguities which can arise when we use words. There is the story of the American girl visiting in England. She was engaged and so was the daughter her hostess. The two girls began to exchange confidences. In the course of their remarks, the American girl said, with respect to the English girl's fiance, "I suppose he must see you every day." The English girl was insulted. Where the American girl had wished to stress the idea that wild horses couldn't keep him away, the English girl got the suggestion that her fiance had to be dragged in by the collar to visit her.

"When we talk about a semantic change in language, we are referring to a change which occurs in the meaning of words. Words have a meaning today; in Shakespeare's day they may have had another; and yet a third in Chaucer's. As a matter of fact, they may have different meanings today as they are used by different people."

C. RATING FORMS

Two rating forms were used by the listeners to evaluate the various speakers' voices in terms of familiarity (Familiarization Rating Form I) and perceived voice characteristics (Voice Characteristic Rating Form IV).

1. Familiarization Rating Form I (Figure 2)

This form was used by the listeners to indicate how the speakers sounded familiar on the first hearing and subsequently with various amounts of exposure to the voices. The subjects rated each of the speakers individually on each of the four experimental days (Table V). The taped instructions for the familiarization rating task were as follows.

LISTENER_____

FAMILIARIZATION RATING FORM I

Instructions: Each voice must be rated according to the number of times you have experienced it in the past. Place an "X" in the box which corresponds to your perception of the speaker's voice you are rating.

SPEAKER

1.	Never	Rarely	Sometimes	Often	Very Often
2.	Never	Rarely	Sometimes	Often	Very Often
3.	Never	Rarely	Sometimes	Often	Very Often
4.	Never	Rarely	Sometimes	Often	Very Often
5.	Never	Rarely	Sometimes	Often	Very Often
6.	Never	Rarely	Sometimes	Often	Very Often
7.	Never	Rarely	Sometimes	Often	Very Often
8.	Never	Rarely	Sometimes	Often	Very Often
9.	Never	Rarely	Sometimes	Often	Very Often
10.	Never	Rarely	Sometimes	Often	Very Often

7869

Figure 2. Familiarization Rating Form I

Table V. Order of Assignment of Groups to Task and Speaker Order

Group (Subjects)	Number of (Subjects (n)	Experimental Days	Task Description	Trial	Speaker Order
I and II	10	1, 2, 3, and 4	Familiarity Ratings	1	B
I	5	1, 2, and 3	Voice Characteristic Ratings	1	B
I	5	1, 2, and 3	Voice Characteristic Ratings	2	B'
II	5	1, 2, and 3	Voice Characteristic Ratings	1	B'
II	5	1, 2, and 3	Voice Characteristic Ratings	2	B

"Now I am going to play some speakers' voices for you.

"Each speaker's voice will have a number which will be announced before it is played. After you hear the voice, I want you to place an X on the proper line on your rating form. Each voice must be rated according to the number of times you have experienced it previously. The five possible ratings are

Never—this means you have never heard that speaker's voice before

Rarely—this means you have heard it before at least once, but only rarely.

Sometimes—this means you have heard the voice a few times, but not often.

Often—this means you have heard the voice more than just a few times, but not very often.

Very often—this means you have heard the voice very frequently.

"If you have never heard the voice before, place and X on the line at the extreme left. If you have heard it very frequently, place an X on the line at the extreme right. If your experience falls somewhere in between, place an X on the line over the word that best describes how often you have heard the voice in the past.

"Each line on the page is for a different speaker. The number of the speaker announced will correspond to the speaker numbers on your rating form.

"Don't try to work out a system for making your responses; just put down your first impression of how familiar you are with the voice. A snap judgment is generally better than one you stop and worry about. You will have about 60 seconds to make your response on each of the speakers, and that should be more than you will need.

"Does everyone understandwhat he is to do?

"First you will hear 'Speaker one' announced. Then you will hear him speak the familiarization text. After this speaker stops you will rate him on how familiar his voice sounds. After 15 seconds you will hear the next speaker number announced, here him speak, and then rate him on the familiarity of his voice. This sequence will be continued until you have heard and rated each of the ten speakers' voices.

"Any questions?

"Ready?"

(This was followed by each of the ten speakers' voices.)

2. Voice Characteristic Rating Form IV

This form was employed to obtain the subjects' ratings of the various speakers' voices in terms of perceived voice characteristics. The subjects rated the speakers' voices for two trials on each of the initial three experimental days (Table V). The taped instructions given the subjects for this task were as follows.

"Now we will begin the second task. In this part of the experiment you are going to hear a series of speakers speaking the voice characteristic rating text. Please read silently your copy of this text as you hear me read it.

"As you can see, on each of the rating forms there are 12 pairs of adjectives. Each of the pairs consists of two adjectives having opposite meaning separated by a seven-point scale. This scale is what you will use to make your responses.

"The way this form will be used is as follows.

"First you will hear the speaker's number announced. You will then write this number in the space indicated at the top of the page above your name. Then, you will hear the speaker begin the selection.

"Now look at item 1 on the first voice rating form (Simple-Complex). While you are listening to the speaker's voice, you will determine whether his voice sounds simple or complex. If you think the voice

sounds very simple, place an X on the line nearest simple. However, if you think the voice sounds very complex, place an X on the line nearest complex. If you think the voice sounds like something between simple and complex, place an X on the line which best indicates your experience.

"You will make your responses on the remaining items of the form as you continue to hear the voice.

"Try to make your response on each item quickly. We are interested in your first impression for each of the items as you are listening to the voice. You will hear each speaker for almost two minutes so you have plenty of time to make your response on each of the items.

"There will be a 20-second delay after the speaker stops and the next speaker starts. During this time you will finish any items you did not complete and turn to the next rating form. You will then hear the next speaker announced. Write his number at the top of this form. As the speaker begins the selection, you start making your responses.

"This sequence will be repeated until you have heard and rated all of the speakers. Are there any questions?"

D. SUBJECTS (LISTENERS)

Ten male subjects were selected at random from Texas Instruments Research and Development Department, personnel who had not participated in any of the earlier experiments, either as a listener or a speaker. Their hearing was reported to be normal. The subjects were then randomly assigned numbers from 1 to 10 and divided into two groups:

Group	Subjects
I	1, 3, 5, 7, 9
II	2, 4, 6, 8, 10

Two groups of subjects were required to control for speaker presentation. The order of assignment of subjects to groups was the same for all testing days (Table V).

E. APPARATUS

Two Ampex 601 tape recorders were employed to present the two sets of spoken material. Recorder I was used to present Order B and Recorder II was used to present Order B'. The outputs of the recorders were fed into a selector box which allowed the experimenter to select Recorder I or Recorder II for each of the ten subjects (Figure 3). In the familiarization rating task, only Recorder I was employed to present the speakers' voices, as there was only

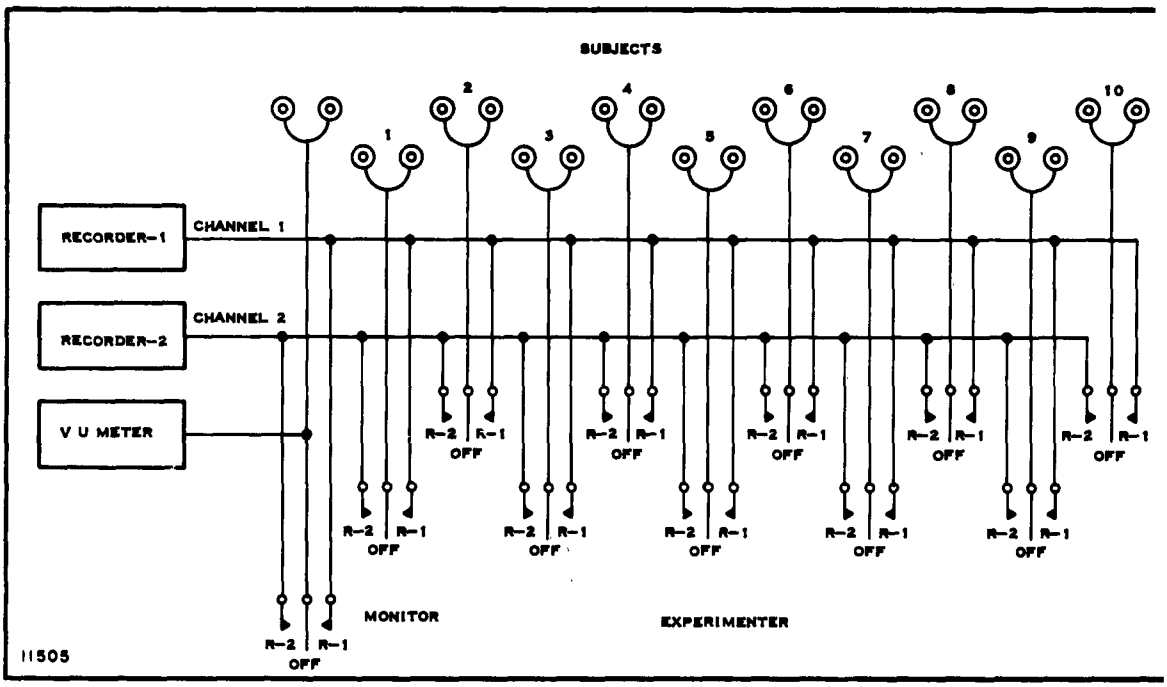


Figure 3. Schematic of Apparatus Used in Experiments

one order of presentation (Order B); whereas, in the voice characteristic rating task, both recorders were required to present the two orders of presentation, B and B'.

Ten matched pairs of Brush Crystal Headphones were employed to present the speakers' voices to the subjects. Each subject used the same pair of headphones for all rating tasks to control for continuity of stimulus presentation.

F. PROCEDURE

The subjects were seated in a small, sound-treated auditorium and furnished a pair of headphones and a test booklet. The subjects listened to the tape recording of the experimenter reading the familiarization rating text. This was done to minimize the effect of text content on the subjects' judgment as to the familiarity of the speakers' voices. The experimenter then presented taped instructions for the familiarization task. This was repeated for four consecutive days.

After completing the familiarization task, the subjects listened to a tape recording of the voice characteristic rating text read by the experimenter. This was to control for the effects of text content on the characteristic ratings. The subjects then heard and rated each speaker's voice for two

trials with a different order of speaker presentation for each trial. This was repeated on three consecutive days. On the fourth day the listeners only judged the speakers' voices on familiarity (see Table V for a summary of the procedures). A summary of the experimental design related to the evaluation of Form IV is presented in Figure 4.

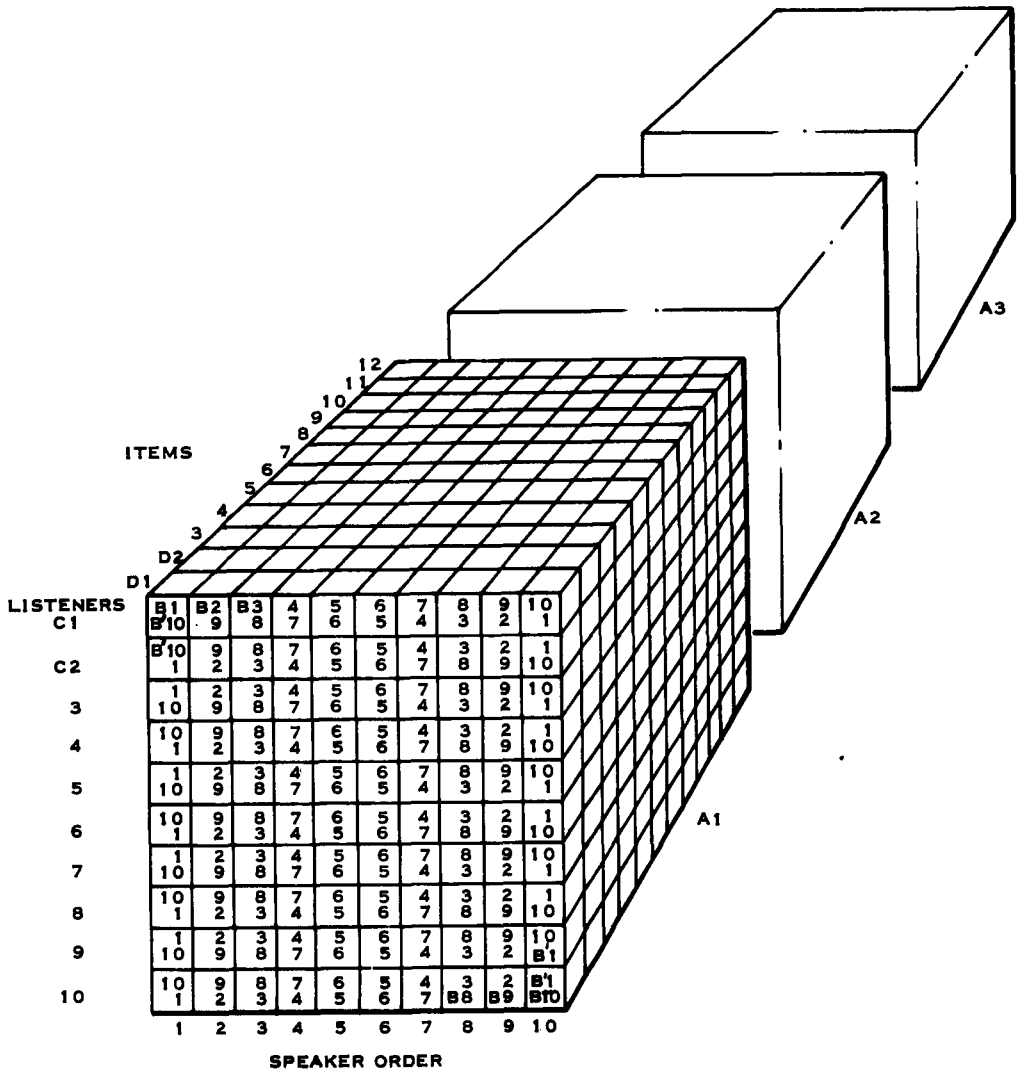


Figure 4. Voice Rating Experiment Design

SECTION V

RESULTS AND DISCUSSION

The results and discussion of the results will be presented relative to the two tasks in this experiment, the familiarization task and the voice characteristic rating task.

A. FAMILIARIZATION TASK

The subjects' ratings on all ten speakers were analyzed separately, first for each of the four separate rating days; then the pooled ratings across the four rating days (the average rating) were analyzed. The data was analyzed by Attneave's[26] method of graded dichotomies for the scaling of judgements. This method avoids the assumption of equal scale intervals, and yields results similar to those obtained by Thurstone's[27] paired comparison techniques. The scale values obtained for each of the speakers relates the degree to which his voice was rated as familiar by the subjects. A small scale value indicates a judgement of low familiarity, while a large scale value indicates a judgment of high familiarity. The results of the graded dichotomies analysis are presented in Table VI.

Examination of Table VI indicates that on the first day the ten speakers were rated from −0.1253 to 1.2198 in terms of familiarity. Note that on the first day the listeners had never heard these ten speakers' voices before.

It is interesting to note that on the first day Speaker 10 was rated to be about as familiar as Speaker 3 was on the fourth day. Thus, there appear to be definite differences among speakers according to the judgements of listeners in terms of familiarity, and that these differences are maintained fairly consistently in time; e.g., Speaker 3 was judged to be the least familiar, while Speaker 10 was judged the most familiar on both the first and fourth days.

The results indicate that listeners rated the speakers higher on the familiarity scale after repeated hearings. Only three speakers (Speakers 3, 4, and 8) showed a slight reversal of this tendency. Speaker 3 was reversed on Day 3 and Speakers 4 and 8 were reversed slightly on Day 4.

The fact that the listeners judged the speakers as varying considerably in familiarity (even when they had never heard the voices before, as was true on Day 1) suggests that ratings such as those obtained by Meeker and Nelson[20] should be interpreted with caution. If different sets of speakers are used to evaluate different types of speech processing systems, and judgements of varying degrees of judged recognizability are obtained relative to the various systems, the judged differences may well reflect basic differences in degree of perceived familiarity of the speakers' voices rather than differences in the effect of the various speech processing systems on the voices. Simply stated, these results indicate that a task in which listeners indicate the extent

Table VI. Graded Dichotomies Scale Values of the Degree of Familiarity of Ten Speakers and Judged by Ten Listeners

Speaker	Scale Values				
	Day 1	Day 2	Day 3	Day 4	Average D1-D4
1	0.3023	2.3023	2.7240	3.3963	2.1812
2	0.1473	1.1398	1.3815	2.5328	1.3003
3	-0.1253	1.1348	0.9515	1.5763	0.8843
4	0.5423	0.9598	2.2265	2.1088	1.4593
5	0.8623	1.4348	2.0065	3.3963	1.9249
6	0.7323	1.1023	1.4343	2.2388	1.3769
7	0.3798	1.8848	2.8315	3.1063	2.0506
8	0.6048	0.8023	2.1165	1.8963	1.3549
9	0.8848	1.0223	1.5115	2.9213	1.5849
10	1.2198	2.0273	2.4365	3.7163	2.3499

to which they believe various speakers' voices are recognizable based on how familiar the voices sound may have quite different results from a task in which the listeners are required to actually name the various speakers as an indication of recognizability.

B. VOICE CHARACTERISTIC RATING TASK

In three general analyses of variance, each of the four variables—speakers, listeners, days, and items—had significant variance. Thus such gross analyses did not lead to meaningful interpretation. The speaker x listener x day analysis (Table VII) was then investigated for each of the 12 items separately to indicate how the listeners agreed in their ratings for the speakers on each item and how stable were their ratings over time (three days). The summary tables for each of these 12 analyses are in Appendix A. The results of these analyses as to speaker variance and listener variance are summarized in Table VIII. Table VIII shows that all speakers were differentiated on each of the 12 items. On five items (No. 3, 4, 7, 8, and 11) there was no significant listener difference, and listener judgements for these items were stable over the three days (see the individual analyses in the appendix).

Thus, the results of this analysis indicate that Form IV can be used by listeners to reliably differentiate among speakers. Some of the items indicate differences among listeners' perceptions of the speakers' voice characteristics, and five items indicate listener agreement and consistency over time in reliably differentiating among the speakers.

Table VII. Summary Table of Analysis of Variance; Ten Speakers, Ten Listeners, and Three Days Over All 12 Items of Form IV

EMS	Source	Sum of Squares (S. S.)	df	Mean Square (M. S.)	F	df	P
A x B	(A) Speakers	741.5278	9	82.3920	24.6517	(9,81)	<0.001
A x B	(B) Listeners	114.1667	9	12.6852	3.7954	(9,81)	<0.001
B x C	(C) Days	45.9539	2	22.9770	16.0814	(2,18)	<0.001
Within	A x B	270.7222	81	3.3422	0.8055	(81,3300)	
A x B x C	A x C	6.7239	18	0.3736	0.3871	(18,162)	
A x B x C	B x C	25.7183	18	1.4288	1.4807	(18,162)	
Within	A x B x C	156.3261	162	0.9650	0.0684	(162,3300)	
	Within	13692.1667	3300	4.1491			
	TOTAL	15053.3056	3599				

Table VIII. Summary Table of the Analyses of Variance Computed for Each of the 12 Items from Form Iv
(As these are correlated, the probability statement associated with F is to be interpreted as inferred statistical significance as expected from noncorrelated observations.)

Item	M. S. Speaker	M. S. Listener	M. S. Interaction	F Speaker	F Listener
1	21.7262	16.0226	5.8357	3.7230^{xxx}	2.7456^{xx}
2	88.8967	18.7559	2.7864	31.9038^{xxx}	6.7312^{xxx}
3	55.6033	6.4922	3.7910	14.6672^{xxx}	1.7125
4	121.1856	7.4967	2.5189	48.1105^{xxx}	2.9762^{xx}
5	55.5392	2.2356	4.1969	13.2334^{xxx}	0.5327
6	35.8700	19.3589	6.0947	5.8854^{xxx}	3.1763^{xx}
7	45.6078	19.9484	11.8407	3.8518^{xxx}	1.6847
8	52.4967	7.7263	4.7905	10.9585^{xxx}	1.6128
9	55.9111	6.8741	2.8592	19.5548^{xxx}	2.4042^{x}
10	58.2756	15.4681	2.2056	26.4217^{xxx}	7.0131^{xxx}
11	71.5792	5.0681	3.0788	23.2491^{xxx}	1.6461
12	70.1408	13.7630	4.2205	16.6191^{xxx}	3.2610^{xx}

x = P < 0.05 for F 1.96
xx = P < 0.01 for F 2.56
xxx = P < 0.001 for F 3.38
df = (9,81)

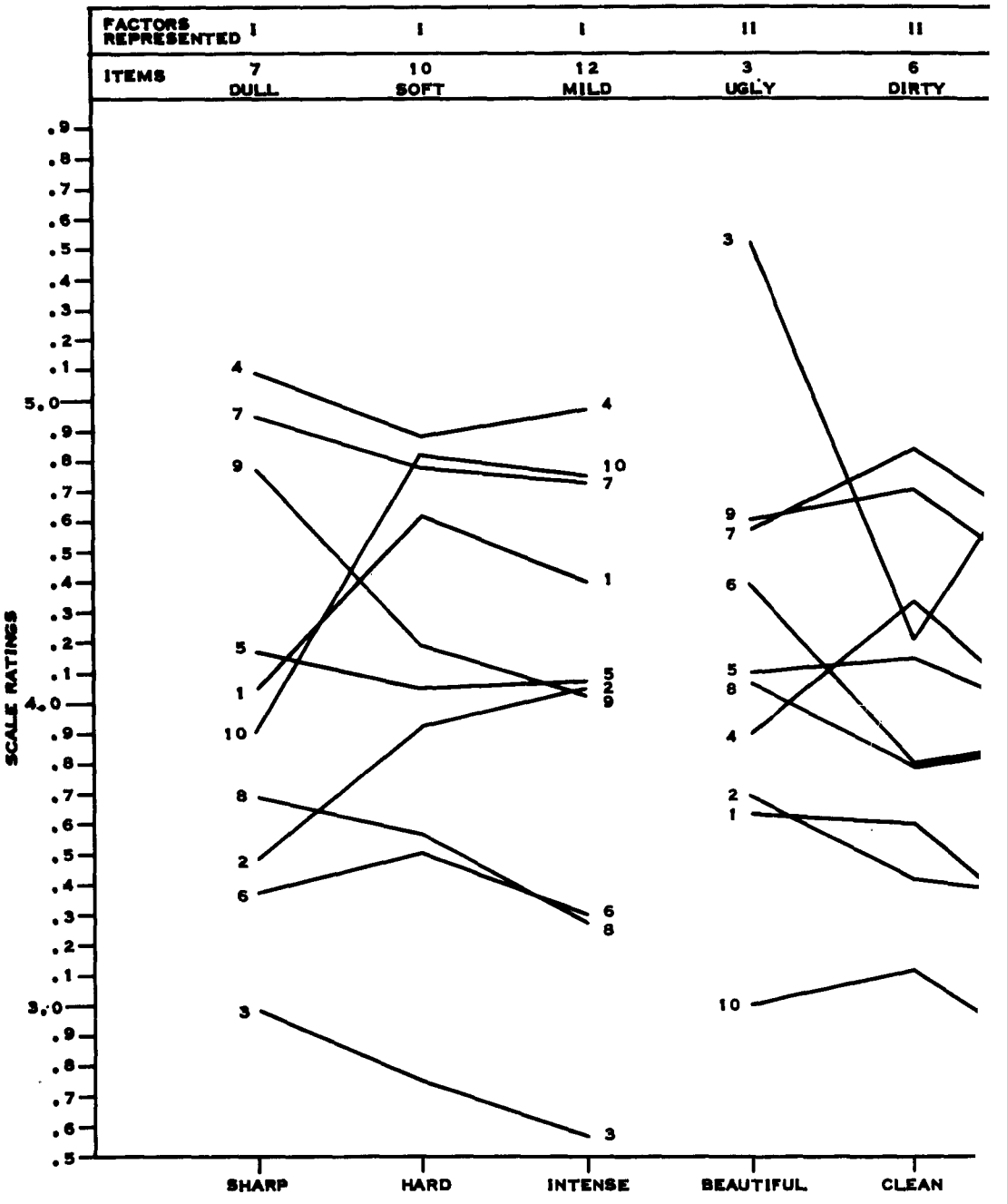
Table IX. Product Moment Correlations Between Two Groups' Ratings on 12 Items of Form IV for Each of Ten Speakers

Speaker Number	Correlation (r) First half (listeners 1—5) vs. Second half (listeners 6—10)
1	0.8960
2	0.8393
3	0.9669
4	0.8871
5	0.6725
6	0.8958
7	0.7006
8	0.7358
9	0.7935
10	0.9003

In a further effort to determine the extent to which this technique is reliable, the subjects were divided into two groups of five and the ratings for both groups on the 12 items were correlated for each of the ten speakers. These correlations are found in Table IX.

These correlations indicate the homogeneity of the listeners' ratings on each of the 12 items of Form IV for each speaker. From the correlations in Table IX we can determine the extent to which the variance can be predicted for one group on the basis of performance in the other group. The amount of variance accounted for in these correlations ranges from 93 percent for Speaker 3 to 45 percent for Speaker 5. For the majority of the speakers, 70 percent or more of the variance in the two groups' ratings on the 12 items is accounted for by these correlations.

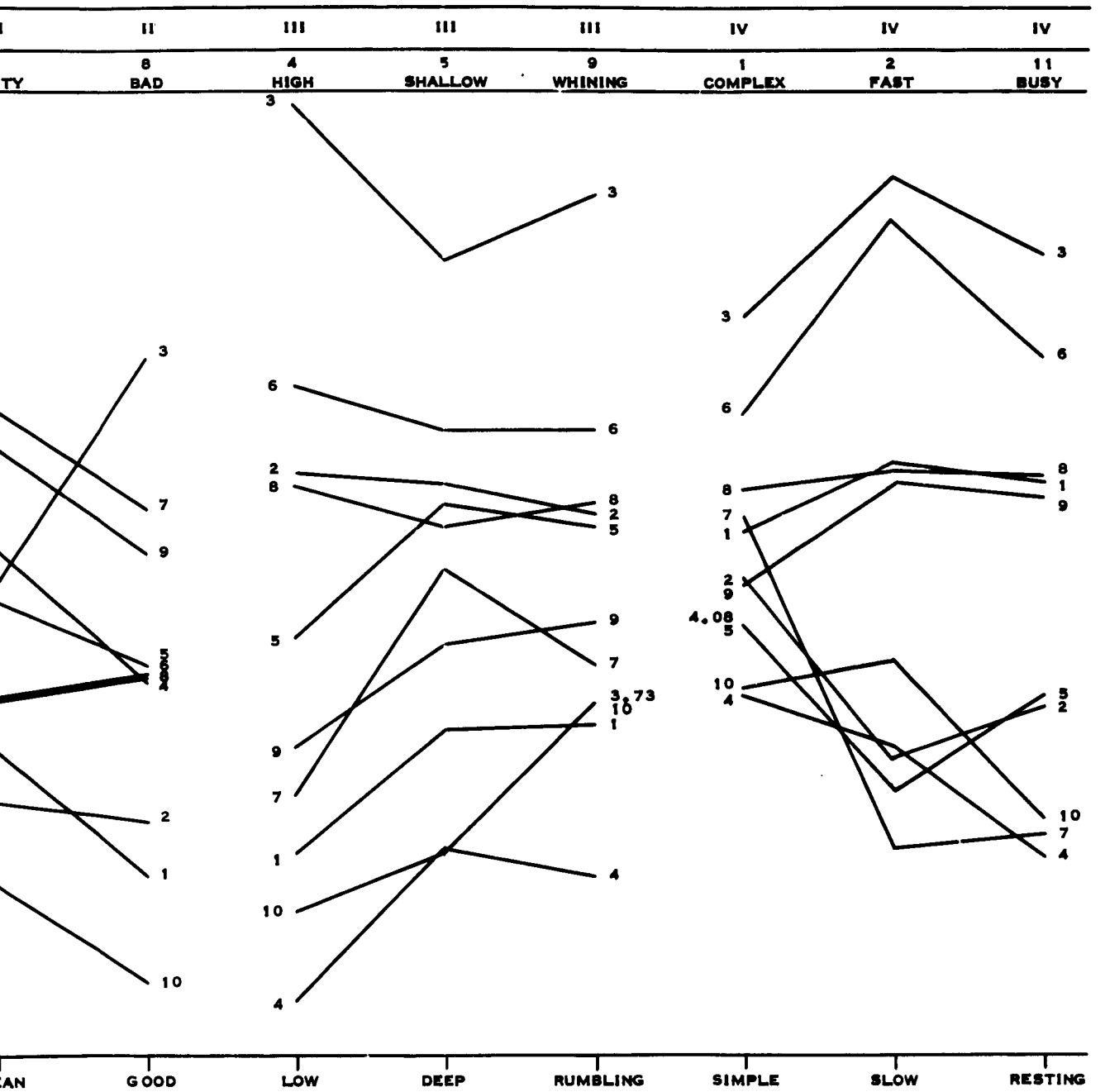
The manner in which the various speakers were perceived to have different voice characteristics, indicated by the listeners' judgements on the 12 items of Form IV, is graphically displayed in Figure 5. On the basis of the analysis conducted, the profiles on Items 3, 5, 7, 8, and 11 for each of the ten speakers, can be regarded as reliable indications as to speaker differences based on perceived voice characteristics. Re-examination of Table VIII leads us to believe that Items 4 and 9 should also be included in this group of reliable indicators due to the relatively small MS listener values as compared with Items 1, 2, 6, 10, and 12. The inclusion of Item 7 (in the group of reliable discriptions) should be regarded somewhat skeptically



RATINGS ON EACH SPEAKER FOR EACH ITEM OF FORM IV, AVERAGED OVER 10 LISTE
EACH ITEM REPRESENTING A FACTOR IS POLARIZED TO THE OTHER 2 ITEMS FOR TH

6678A

1



LISTENERS 2 TRIALS ON 3 DAYS,
OR THAT FACTOR.

Figure 5. Listeners' Ratings of Speakers by Items Over Three Days

because of its high MS listener value and the fact that the large MS interaction value was responsible for the nonsignificant F ratio for listener variance.

Figure 5 can be interpreted as follows.

The ordinate is the portion of the seven-point scale separating the adjective pairs used by the listeners in rating the speakers' voices. Listeners' ratings on the items were scored by assigning values one through seven, going from left to right, to the seven points separating the adjective pairs. Thus an item rated by a listener.

Simple ____ __X__ ____ ____ ____ ____ ____ Complex would be scored as 2. As each rating in Figure 5 reflects the average of ten listeners' responses on each of three days, the range of the original one to seven scale was reduced to 2.6 to 5.95.

The abscissa indicates the items (adjective pairs). Each of the four factors is represented by three items. The lines under the four factors represent the ten speakers and are so labeled.

This profile information indicates the extent to which the speakers' voices were judged to be different. Speaker 3 can be described as having a voice with some of the following characteristics: intense, ugly, high, and fast; whereas Speaker 4 was judged as dull, somewhat dirty, low, and resting. Each of the speakers can be differentiated on the basis of their ratings on the various items.

The following profile characteristics indicate the usefulness of the rating form method to indicate perceived differences among speakers' voices.

The speakers' voices were all rated differently on each item rather than clustering around the same point on the scale.

No two speakers were judged identically.

There was considerable agreement in listener judgements indicated by the wide scale range and the scale differences between speakers. (This is substantiated by the analysis of variance and correlation data.)

The three ratings representing each of the four factors for each of the speakers clustered in the same scale area. For example on the three items representing Factor III, Speaker 3 was judged high, shallow, and whining, and not high, deep, and whining.

As was stated earlier, the items employed in Form IV were selected to represent the four factors observed in the two previous analyses; thus, the item correlations over all speakers were factor-analyzed to determine whether the items maintained their factor representation as hypothesized.

The sums of the ten listeners' ratings (for two trials on each of the three days) on each of the ten speakers on the 12 items of Form IV were computed

Table X. Sum of Ten Subjects' Ratings (For Two Trials on Each of Three Days) on Ten Speakers by Use of 12-Item Schematic Differential Form IV

	Item	Speaker										Mean	$(S.D.)^2$
		1	2	3	4	5	6	7	8	9	10		
1	Simple–complex	264	254	310	228	245	290	267	273	252	230	261.3	24.2200
2	Slow–fast	279	215	341	217	207	332	195	277	274	236	257.3	48.9919
3	Beautiful–ugly	218	222	331	234	246	264	274	244	276	180	248.9	38.7465
4	Low–high	194	278	357	162	241	296	207	274	217	181	240.7	57.2015
5	Shallow–deep	279	225	176	304	229	213	243	234	260	305	246.8	38.7242
6	Dirty–clean	282	294	246	239	250	271	209	272	217	312	259.2	31.2883
7	Dull–sharp	256	291	320	194	249	297	202	278	212	264	256.3	40.3659
8	Good–bad	189	201	302	231	235	233	269	232	259	166	231.7	37.6485
9	Rumbling–whining	222	268	337	189	265	286	235	270	244	224	254.0	38.8536
10	Hard–soft	277	235	165	293	243	210	287	214	251	289	246.4	39.6666
11	Resting–busy	274	226	324	193	228	302	198	276	272	201	249.4	43.9618
12	Intense–mild	264	243	154	298	245	198	284	197	242	285	241.0	43.5178
	Mean	249.8333	246.0000	280.2500	231.8333	240.2500	266.0000	239.1667	253.4167	248.0000	239.4167		
	$(S.D.)^2$	33.0450	29.4873	71.5846	44.0357	13.7848	41.0041	34.3750	27.0169	21.5948	49.0382		

(Table X). From this data the correlations of speakers over the 12 items and the correlations of items over speakers were computed.

The speaker correlations are presented in Table XI. The highest negative correlation is between Speakers 3 and 4. Comparing this correlation with the profile information in Figure 5 reveals that the profiles for Speakers 3 and 4 are quite dissimilar for the 12 items. Conversely, the highest positive correlation in Table 11 is between Speakers 6 and 8. The profile information in Figure 5 indicated that Speakers 6 and 8 are quite similar on all 12 items.

Such an observation suggests the hypothesis that speakers with high positive correlations and similar profiles would confuse speaker differentiations (e. g., speaker recognition or speaker identification) more often than speakers with high negative correlations and dissimilar profiles.

Table XIII. Square-Root Factor Analysis of 12 Items
Over Each of Ten Speakers Using Form IV
(Pivot variables chosen on basis of largest sum of the three items representing Factor I, II, III, and IV—e. g., variables 12, 3, 5, and 1, respectively)

Item		I Intensity	II Quality	III Pitch	IV Rate	h^2
1	Simple—complex	0.8570	0.2513	0.1020	0.4437	1.0048
2	Slow—fast	0.7994	-0.0378	-0.4169	0.3161	0.9142
3	Beautiful—ugly	0.6599	0.7513	0.0000	0.0000	0.9999
4	Low—high	0.9459	0.0374	0.2814	-0.0005	0.9753
5	Deep—shallow	0.8588	0.2407	0.4522	0.0000	0.9999
6	Clean—dirty	0.0289	-0.9391	0.0687	0.1452	0.9085
7	Dull—sharp	0.7859	-0.4630	0.2102	0.1580	0.9012
8	Good—bad	0.5197	0.8335	0.0003	-0.1047	0.9758
9	Rumbling—whining	0.9355	0.0701	0.2758	0.0263	0.9568
10	Soft—hard	0.9860	0.0163	0.0988	-0.0940	0.9911
11	Resting—busy	0.8897	0.0169	-0.2351	0.2217	0.8963
12	Mild—intense	1.000	0.0000	0.0000	0.0000	1.0000

The item correlations are found in Table XII. The correlations were factor analyzed to indicate how effectively the items selected for Form IV were used by the listeners to discriminate among the various speakers and to determine whether the items represented the factors as hypothesized (Table XIII). By checking Table IV, we can see which factors the various items were hypothesized to represent. Examination of the factor-analysis summary in Table XIII leads to the conclusion that the items selected to represent the four factors in Form IV are the items that best represent the four factors in this analysis.

Further research suggested by the results of this experiment includes determining further the reliability of Form IV, determining the validity of the form for predicting judgements of similarity-dissimilarity, relating the perceived characteristics of speakers' voices to the physical characteristics of the speakers' voices, and using Form IV to evaluate the changes in perceived characteristics of speakers' voices which accompany degradations in the physical characteristics in the various speech processing systems (i. e., vocoders) in the development of a fidelity criterion. These lines of research are to be pursued, and hopefully the results will lead to a technique by which certain elements of the speech processing system might be specified to meet various functional requirements of the system.

Table XI. Correlations of Ten Speakers Over 12 Items on Form IV
(Correlations based on total of ten listeners' judgements
for two trials on each of three days.)

Speaker	1	2	3	4	5	6	7	8	9	10
1	1.0000	0.0592	-0.5062	0.4781	-0.2948	-0.0433	-0.1507	-0.0172	0.0939	0.7857
2		1.0000	0.1172	-0.3678	0.6220	0.2707	-0.3902	0.4389	-0.8999	0.3372
3			1.0000	-0.9276	-0.0426	0.8671	-0.5405	0.8019	0.0320	-0.7585
4				1.0000	-0.1010	-0.8438	0.6596	-0.8261	0.2495	0.6415
5					1.0000	-0.2068	0.2996	-0.0549	-0.5406	-0.0858
6						1.0000	-0.7724	0.9359	-0.0395	-0.3971
7							1.0000	-0.8155	0.2791	0.0355
8								1.0000	-0.2063	-0.2708
9									1.0000	-0.3242
10										1.0000

Table XII. Correlations of 12 Items Over Ten Speakers Using Form IV
(Correlations based on total of ten listeners' judgements
for two trials on each of three days)

Item	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	0.7699	0.7543	0.8485	-0.8426	-0.1413	0.6470	0.6095	0.8361	-0.8185	0.8387	-0.8570
2		1.00	0.4991	0.6445	-0.4889	0.1061	0.6413	0.3119	0.6351	-0.7251	0.9370	-0.7994
3			1.00	0.6523	-0.7476	-0.6865	0.1707	0.9692	0.6700	-0.6629	0.5998	-0.6599
4				1.00	-0.9486	0.0505	0.8233	0.5192	0.9790	-0.9741	0.7543	-0.9459
5					1.00	0.1702	-0.6585	-0.6471	-0.4450	0.8954	-0.6618	0.8588
6						1.00	0.5895	-0.8031	0.0204	-0.0292	0.0095	-0.0289
7							1.00	-0.0155	0.8063	-0.7952	0.6614	-0.7859
8								1.00	0.5393	-0.5316	0.4131	-0.5197
9									1.00	-0.9564	0.7505	-0.9355
10										1.00	-0.8275	0.9860
11											1.00	-0.8897
12												1.00

The results of the present experiment are summarized as follows.

The four factors found in earlier experiments were identified again as adequate to account for listeners' perceptions of speakers' voices.

Speakers were discriminable on the basis of their ratings on four factors.

Each item representing a factor was used by the listeners to discriminate between speakers.

The majority of the items reflected differences only between speakers; some also reflected differences between the listeners.

Ratings were stable over time on some items, but not all.

Some of the potential uses of the techniques developed in this research are:

To determine the extent to which voices previously unheard seem familiar

To differentiate among various speakers in terms of perceived voice characteristics

To specify the relationship between perceived voice characteristics and physical characteristics

To specify elements in the speech processing systems necessary to preserve desired preceived voice characteristics

To evaluate various speech processing devices in terms of the extent to which processing disturbs perceived characteristics

To specify certain fidelity criteria for speech processing systems

To select speakers whose voices withstand minimum system requirements for intelligibility and recognizability, thus permitting increased bandwidth compression

To select listeners for communication systems who are most consistent and efficient in differentiating among speakers on the basis of perceived characteristics.

SECTION VI

REFERENCES

1. F. Goldman-Eisler, "On the Variability of the Speed of Talking and on its Relation to the Length of Utterances in a Conversation," British Journal of Psychology (General Section), Vol. 45 (1954), p. 94.

2. Peter Ladefoged and D. E. Broadbent, "Information Conveyed by Vowels," Journal of the Acoustical Society of America, Vol. 29 (1957), p. 98.

3. Dayton C. Miller, The Science of Musical Sounds (New York: The Macmillan Co., 1922), p. 58.

4. Virgil A. Anderson, Training the Speaking Voice (New York: Oxford University Press, 1942), p. 107.

5. Robert Curry, The Mechanism of the Human Voice (London: J. and A. Churchill, Ltd., 1940), p. 102.

6. G. W. Gray and C. M. Wise, The Bases of Speech (New York: Harper and Brothers, 1946), p. 12.

7. L. S. Judson and A. T. Weaver, Voice Science (New York: F. S. Crofts and Co., 1942), p. 289.

8. Frances McGehee, "The Reliability of the Identification of the Human Voice," Journal of General Psychology, Vol. 17 (1937), pp. 249-271.

9. Frances McGehee, "An Experimental Study in Voice Recognition," Journal of General Psychology, Vol. 31 (1944), pp. 53-65.

10. I. Pollack, J. M. Pickett, and W. H. Sumby, "On the Identification of Speakers by Voice," Journal of the Acoustical Society of America, Vol. 26 (1954), pp. 403-406.

11. R. W. Peters, "Studies in Extra Messages: Listener Identification of Speakers' Voices Under Conditions of Certain Restrictions Imposed Upon the Voice Signal," USNSAM Joint Project Report No. 30 (1954).

12. R. W. Peters, "Studies in Extra Messages: The Effect of Various Modifications of the Voice Signal Upon the Ability of Listeners to Identify Speakers' Voices," USNSAM Joint Project Report No. 61 (1956).

13. Gretchen A. Skalbeck, "An Experimental Study of Several Factors in Speaker Recognition," Unpublished Masters Thesis, University of Washington (1955).

14. John W. Black and John J. Dreher, "Nonverbal Messages in Voice Communication," USNSAM Joint Project Report No. 45 (1955).

15. A. S. Howell, "Feasibility Study of Transistorized Coder-Multiplexer," Report 1820-TN-1, Project 5570 of Contract AF30(602)-2262, General Dynamics Electronics (December 1960).

16. L. V. Surgent and L. H. Yost, "Measurement of Speaker Recognition," Paper presented at 61st meeting of the Acoustical Society of America, May 10-13, 1961.

17. J. N. Shearme and J. N. Holmes, "An Experiment Concerning the Recognition of Voices," Language and Speech, Vol. 2 (1959), pp. 123-131.

18. Victor E. McGee, "The Determination of a Perceptual Space for the Quality of Filtered Speech," E. T. S. Research Bulletin (RB-61-21), Princeton University (1961).

19. J. A. Williamson, "An Investigation of Several Factors Which Affect the Ability to Identify Voices as Same or Different," Unpublished Dissertation for Diploma in Phonetics, University of Edinburgh (1961).

20. W. F. Meeker and A. L. Nelson, "Vocoder Evaluation Studies," Report No. 3, Contract AF19(604)-6151, RLA (June 1962).

21. C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, The Measurement of Meaning (Urbana: University of Illinois Press, 1957).

22. Lois L. Elliott and P. H. Tannenbaum, "Factor Structure of Semantic Differential Responses to Visual Form and Prediction of Factor Scores from Structural Characteristics of the Stimuli," Unpublished Research, USAF Aerospace Medical Center, Brooks AFB, Texas.

23. W. H. Lichte, "Attributes of Complex Tones," Journal of Experimental Psychology, Vol. 28 (1941), pp. 445-480.

24. R. W. Peters, "Research on Psychological Parameters of Sound," WADD Technical Report, Vol. 60 (1960), p. 249.

25. E. Uldall, "Attitudinal Meanings Conveyed by Intonation Contours," Language and Speech, Vol. 3 (1960), pp. 223-234.

26. Fred Attneave, "A Method of Graded Dichotomies for the Scaling of Judgements," Psychological Review, Vol. 56 (1949), p. 334-340.

27. L. L. Thurstone, "Psychophysical Analysis," American Journal of Psychology, Vol. 56 (1927), pp. 368-389.

SECTION VII

SCIENTISTS CONTRIBUTING TO REPORT

Persons contributing to this report were Gary L. Holmgren, Dr. E. J. Morrison, Dr. W. D. Voiers, and Kenneth Berry. The latter three contributed as consultants. Mr. Holmgren was the principal investigator. A resume of his education and experience is as follows.

M. A. in Psychology, Texas Christian University.
B. A. in Psychology, Texas Christian University.

Member of Psi Chi, American Psychology Association, Sigma Xi, Acoustical Society of America, IEEE.

Presented a paper before the Southwestern Psychological Association in 1962 on Intermodal Transfer in a Paired-Associates Learning Task, and before the Southern Society for Philosophy and Psychology on Stimulus Components and Stimulus Patterning in Auditory Perception.

Mr. Holmgren is assigned to a study and development program directed toward designing and producing a digital voice communication system. His responsibility in this work includes investigating the determinants of speaker recognition. He also works closely with the group developing statistical models of the communication system. Before joining Texas Instruments in 1962, Mr. Holmgren was a Research Assistant and Teaching Research Fellow at Texas Christian University for four years. His research concerned perception of auditory patterns and the relationships between visual and auditory perception. Work on his Ph. D. Will be completed at Texas Christian University in 1963.

APPENDIX A

ANALYSIS OF VARIANCE SUMMARY TABLES FOR EACH OF THE 12 ITEMS (ADJECTIVE PAIRS) EMPLOYED IN FORM IV

Speakers x Listeners x Days—Summary Table

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	195.536	9	21.7262	3.7230	(9, 81)	<0.001
A x B	(B) Subjects	144.203	9	16.0226	2.7456	(9, 81)	<0.01
B x C	(C) Days	27.980	2	13.9900	6.1589	(2, 18)	<0.01
Error	A x B	472.695	81	5.8357	4.6179	(81, 162)	<0.001
Error	A x C	23.754	18	1.3197	1.0443	(18, 162)	
Error	B x C	40.887	18	2.2715	1.7975	(18, 162)	<0.05
	Error	204.715	162	1.2637			
	Total	1109.770	299				

Item 1

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	800.070	9	88.8967	31.9038	(9, 81)	<0.001
A x B	(B) Subjects	168.803	9	18.7559	6.7312	(9, 81)	<0.001
B x C	(C) Days	19.887	2	9.9435	2.7348	(2, 18)	
Error	A x B	225.696	81	2.7864	3.3599	(81, 162)	<0.001
Error	A x C	10.98	18	0.6100	0.7356	(18, 162)	
Error	B x C	65.447	18	3.6359	4.3843	(18, 162)	<0.001
	Error	134.354	162	0.8293			
	Total	1425.2370	299				

Item 2

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	500.430	9	55.6033	14.6672	(9, 81)	<0.001
A x B	(B) Subjects	58.430	9	6.4922	1.7125	(9, 81)	
B x C	(C) Days	11.387	2	5.6935	2.8723	(2, 18)	
Error	A x B	307.067	81	3.7910	4.7710	(81, 162)	<0.001
Error	A x C	22.880	18	1.2711	1.5997	(18, 162)	
Error	B x C	35.680	18	1.9822	2.4946	(18, 162)	<0.001
	Error	128.723	162	0.7946			
	Total	1064.597	299				

Item 3

Speakers x Listeners x Days—Summary Table (Continued)

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	1090.670	9	121.1856	48.1105	(9, 81)	<0.001
A x B	(B) Subjects	67.470	9	7.4967	2.9762	(9, 81)	<0.01
B x C	(C) Days	19.227	2	9.6135	10.6556	(2, 18)	<0.001
Error	A x B	204.029	81	2.5189	2.8827	(81, 162)	<0.001
Error	A x C	11.640	18	0.6467	0.7401	(18, 162)	
Error	B x C	16.240	18	0.9022	1.0325	(18, 162)	
	Error	141.561	162	0.8738			
	Total	1550.837	299				

Item 4

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	499.853	9	55.5392	13.2334	(9, 81)	<0.001
A x B	(B) Subjects	20.120	9	2.2356	0.5327	(9, 81)	
B x C	(C) Days	34.667	2	1.7334	0.9231	(2, 18)	
Error	A x B	339.945	81	4.1969	2.8384	(81, 162)	<0.001
Error	A x C	32.667	18	1.8148	1.2274	(18, 162)	
Error	B x C	33.800	18	1.8778	1.2700	(18, 162)	
	Error	239.535	162	1.4786			
	Total	1200.587	299				

Item 5

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	322.830	9	35.8700	5.8854	(9, 81)	<0.001
A x B	(B) Subjects	174.230	9	19.3589	3.1763	(9, 81)	<0.01
B x C	(C) Days	1.647	2	0.8235	3.8988	(2, 18)	<0.05
Error	A x B	493.668	81	6.0947	4.8548	(81, 162)	<0.001
Error	A x C	23.620	18	1.3122	1.0452	(18, 162)	
Error	B x C	38.020	18	2.1122	1.6825	(18, 162)	<0.05
	Error	203.382	162	1.2554			
	Total	1257.397	299				

Item 6

Speakers x Listeners x Days—Summary Table (Continued)

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	410.470	9	45.6078	3.8518	(9, 81)	<0.001
A x B	(B) Subjects	179.536	9	19.9484	1.6847	(9, 81)	
B x C	(C) Days	35.780	2	17.8900	2.9556	(2, 18)	
Error	A x B	959.096	81	11.8407	1.2701	(81, 162)	
Error	A x C	145.620	18	8.0900	0.8678	(18, 162)	
Error	B x C	108.954	18	6.0530	0.6493	(18, 162)	
	Error	1510.314	162	9.3229			
	Total	3349.770	299				

Item 7

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	472.470	9	52.4967	10.9585	(9, 81)	<0.001
A x B	(B) Subjects	69.537	9	7.7263	1.6128	(9, 81)	
B x C	(C) Days	9.007	2	4.5035	1.4477	(2, 18)	
Error	A x B	388.029	81	4.7905	4.2654	(81, 162)	<0.001
Error	A x C	41.060	18	2.2811	2.0311	(18, 162)	<0.01
Error	B x C	55.993	18	3.1107	2.7697	(18, 162)	<0.001
	Error	181.941	162	1.1231			
	Total	1218.037	299				

Item 8

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	503.200	9	55.9111	19.5548	(9, 18)	<0.001
A x B	(B) Subjects	61.867	9	6.8741	2.4042	(9, 81)	<0.05
B x C	(C) Days	13.627	2	6.8135	5.9037	(2, 181)	<0.05
Error	A x B	231.597	81	2.8592	2.8741	(81, 162)	<0.001
Error	A x C	8.440	18	0.4689	0.4714	(18, 162)	
Error	B x C	20.773	18	1.1541	1.1601	(18, 162)	
	Error	161.163	162	0.9948			
	Total	1000.667	299				

Item 9

Speakers x Listeners x Days—Summary Table (Continued)

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	524.480	9	58.2756	26.4217	(9, 18)	<0.001
A x B	(B) Subjects	139.213	9	15.4681	7.0131	(9, 18)	<0.001
B x C	(C) Days	1.847	2	0.9235	0.2823	(2, 18)	
Error	A x B	178.651	81	2.2056	1.5010	(81, 162)	<0.01
Error	A x C	23.220	18	1.2900	0.8780	(18, 162)	
Error	B x C	58.887	18	3.2715	2.2264	(18, 162)	<0.01
	Error	238.049	162	1.4694			
	Total	1164.347	299				

Item 10

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	644.213	9	71.5792	23.2491	(9, 81)	<0.001
A x B	(B) Subjects	45.613	9	5.0681	1.6461	(9, 81)	
B x C	(C) Days	12.527	2	6.2635	2.8465	(2, 18)	
Error	A x B	249.386	81	3.0788	2.7050	(81, 162)	<0.001
Error	A x C	16.807	18	0.9337	0.8203	(18, 162)	
Error	B x C	39.607	18	2.2004	1.9332	(18, 162)	<0.05
	Error	184.394	162	1.1382			
	Total	1192.547	299				

Item 11

EMS	Source	Sum of Squares	df	Mean Square	F	df	P
A x B	(A) Speakers	631.267	9	70.1408	16.6191	(9, 81)	<0.001
A x B	(B) Subjects	123.867	9	13.7630	3.2610	(9, 81)	<0.01
B x C	(C) Days	7.647	2	3.8235	1.3668	(2, 18)	
Error	A x B	341.864	81	4.2205	3.4305	(81, 162)	<0.001
Error	A x C	17.353	18	0.9641	0.7836	(18, 162)	
Error	B x C	50.353	18	2.7974	2.2738	(18, 162)	<0.001
	Error	199.316	162	1.2303			
	Total	1371.667	299				

Item 12

AD–

Texas Instruments Incorporated, Dallas, Texas. SPEAKER RECOGNITION, by G. L. Holmgren. May 1963. 52 p. illus. tables (Proj. 4610: Task 461002) (Report 14-73801-14; AFCRL-63-119) (Contract AF19(628)-345) Unclassified report

Reported research is concerned with defining a perceptual space within which listeners locate voices, to the end that the effects of manipulating speaker, hardware, and listener characteristics can be measured, and eventually, that specifications for elements of the communication system can be prepared to produce the desired system characteristics. In the experiments, taped speech samples were rated by listeners using Osgood's semantic differential method. Previous study indicated that only four basic

UNCLASSIFIED

1. Speech transmission
2. Voice communication systems

UNCLASSIFIED

AD–

Texas Instruments Incorporated, Dallas, Texas. SPEAKER RECOGNITION, by G. L. Holmgren. May 1963. 52 p. illus. tables (Proj. 4610: Task 461002) (Report 14-73801-14; AFCRL-63-119) (Contract AF19(628)-345) Unclassified report

Reported research is concerned with defining a perceptual space within which listeners locate voices, to the end that the effects of manipulating speaker, hardware, and listener characteristics can be measured, and eventually, that specifications for elements of the communication system can be prepared to produce the desired system characteristics. In the experiments, taped speech samples were rated by listeners using Osgood's semantic differential method. Previous study indicated that only four basic

UNCLASSIFIED

1. Speech transmission
2. Voice communication systems

UNCLASSIFIED

dimensions were required to account for ratings given speakers on a large number of characteristics. In a second experiment, a reduced number of characteristics, selected from the original list as best representing the four necessary factors, was used by listeners to rate speakers from AFCRL's speaker library. The experimental design allowed examination of the effects on ratings due to differences between listeners, due to repetition of the rating task, and to order of speaker presentation. Results of these examinations and the following are presented. (1) The adequacy of original factors to account for listeners' ratings, (2) the differentiation between speakers, (3) the reliability of ratings, and (4) the familiarity of previously unheard voices.

UNCLASSIFIED

dimensions were required to account for ratings given speakers on a large number of characteristics. In a second experiment, a reduced number of characteristics, selected from the original list as best representing the four necessary factors, was used by listeners to rate speakers from AFCRL's speaker library. The experimental design allowed examination of the effects on ratings due to differences between listeners, due to repetition of the rating task, and to order of speaker presentation. Results of these examinations and the following are presented. (1) The adequacy of original factors to account for listeners' ratings, (2) the differentiation between speakers, (3) the reliability of ratings, and (4) the familiarity of previously unheard voices.

UNCLASSIFIED