1960

Very low frequency characteristics of speech

Ralph Floyd Schauer
Iowa State University

Recommended Citation

Schauer, Ralph Floyd, "Very low frequency characteristics of speech " (1960). *Retrospective Theses and Dissertations*. 2749.
https://lib.dr.iastate.edu/rtd/2749

VERY LOW FREQUENCY CHARACTERISTICS OF SPEECH

by

Ralph Floyd Schauer

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1960

TABLE OF CONTENTS

INTRODUCTION

It has long been realized that speech contains certain salient elements which enable the listening ear and the associated auditory processes to extract semantic and other data which leads to the subjective hearing experience. In addition, speech contains a great deal of irrevalent structure which the hearing sense is able to delineate and to discard. This implies that in the handling of the speech signal in communication systems, circuit and system components should be designed to preserve the information bearing elements while discarding the superfluous features. (9).

The superfluous nature of a signal, that is the redundant and descriptive information above and beyond that necessary for comprehension, is frequently a function of the anatomy of the source. Information can be represented in a number of different ways. The particular method chosen will introduce additional form or structure which characterizes the representation. In this light speech can be considered as a modulated carrier in which the sequence of vocal cord pulses is modulated by the slowly varying vocal cavities. However, the structure concerning the source of signal generally provides little protection against noise during transmission. For this reason, in communication systems predetermined structure is often inserted in a signal so that it can be transmitted intact through a noisy channel. Familiar examples are frequency modulation and the various pulse modulation systems. The elimination of the structure peculiar to the source does not greatly increase the noise susceptibility of the signal. (9).

Efficient transmission of the signal requires the structure peculiar to the source be removed and a noise-resistant structure inserted. This process results in an over-all economy of channel capacity required to transmit the message. Studies of noise-resistant structures have resulted in noise-resistant codes which insert these structures into the signals. Such codes usually assume that each elementary signal or pulse is of equal importance, and no account is taken of the source structure. This approach is justified if all of the superfluous source structure has been removed prior to the application of the noise-resistant code.

The function of removing the source structure is concerned with both the message signals and their source. In the case of the speech signal the origin of the source structure can be considered as the acoustical cavity formed by the mouth and the throat. This cavity is a multiply resonant system whose poles or resonances are functions of its dimensions. In the production of speech, the shape of this cavity is varied relatively slowly by gestures of the tongue and jaw, thereby changing the resonances. During the production of vowel sounds, the cavity is excited by a series of nearly-periodic pulses generated by the vocal cords. These pulses evoke a repeated acoustic transient which comprises the sound wave. Sounds excited in this way are called voiced. The positions of the poles of the cavity, often referred to as formant frequencies, are different for the various vowel sounds and for the same vowel sound, are different for various speakers. In the case of consonants, excitation is provided by air passing turbulently through constrictions usually located toward the front of the vocal tract. Cavities both in front and behind the source tend to affect the quality or phonetic value of this "unvoiced" sound. The nature of the source results in a broad energy distribution with significant contributions as high as 5 - 8 kilocycles per second. The nature of the attack and decay of such sounds is also important. Some few sounds fall between these two classifications, voiced and unvoiced. Thus speech production can be thought of as a low-frequency modulation of either a pulse or noise carrier wave. (9).

Historically, most attempts of characterization of speech have been predicated upon a fixed shape for the vocal tract although in connected speech the changes in the shape of the tract occur relatively slowly compared to the frequency content of the excitation. Speech production can then be assumed to be a quasi-steady state process. This assumption permits the time variations of the tract to be described rather accurately as a succession of steady states. This realization has led to a number of methods for characterizing speech by means of a few slowly varying parameters. The three most investigated and reported mathematical models are:

1. Positions of vocal tract poles as a function of time along with pitch frequency as function of the same variable.
2. Time varying or "short-time" spectrum or sound spectrogram.

3. Short-time autocorrelation function. (9).

There are very extensive applications for equipment which can automatically extract the intelligence from speech. The application to narrow band transmission is the most fundamental and general. A large reduction in channel capacity is possible when only the information bearing structure is to be transmitted. (33).

It was reported early that the formant frequencies carry an important part of the information required for the identification of speech sounds. (33). Flanagan and others have reported extensively on formant frequencies and their variation among speakers. (15, 16, 17, 18, 19, 33, 41). It has been recognized that formant transitions also contain cues to recognition of the speech elements. (30, 35, 40). Harris has found that speech formed from building blocks, with one block for each vowel and each consonant, is not only unnatural but almost unintelligible because the influences between adjacent speech sounds is missing. (29). Flanagan and House have reported a formant-coding speech compression system which utilizes the following information: frequencies of the three formants, amplitudes of voicing and of friction, fundamental vocal frequency, and frequency of the spectral maximum of the fricative excitation. (20). An automatic recognizer of spoken digits which essentially breaks the sound into two formant frequencies and compares these against stored patterns has been reported by Davis and others. (10).

The sound spectrogram, developed at the Bell Telephone Laboratories, has also been a useful method of characterizing speech.

The spectrogram essentially displays the distribution of average power in various speech sounds as a function of frequency. The signal is divided into bands and the square of the amplitude in each band over a measured time interval is recorded. (1, 37, 46). It is possible for a trained person to read the sound spectrogram almost as easily as a phonetic transcription of the utterance. (9). In studies conducted to determine in what factors the information is contained, it was observed that a large number of phonemes are intimately associated with rapid changes in the spectral content of the sound. The observed fact that steady sounds tend to lose their meaning is indicative that changes in the spectrum are important. Therefore Kock and Miller have proposed dynamic spectrograms which involve the differentation of the time-amplitude pattern for different points in the spectrum. (34).

The information structure of a sound spectrogram has been used in several different ways. In the "Vocoder" spectral analysis is used to compress the voice signals. In this device each of a number of filters is used to produce a signal corresponding to the energy within a narrow band of frequencies. The output of the filters then indicates the energy distribution of the sound as a function of time. This signal is transmitted and reconstructed although with an inherent loss of naturalness at the receiver. (8, 9, 11, 25, 31, 37, 49). The energy distribution as a function of time has also been used to attempt automatic recognition of sounds. In this case the derived distribution function is compared with stored patterns and matched to the nearest of these. The stored patterns are energy spectra of either phonemes or complete words. Successes have been achieved using this technique, notably with "Audrey" of the Bell Telephone Laboratories, for the speech of a single speaker. However, less success has been obtained for several speakers unless adjustment of the equipment for each new source is made. (8, 12, 13, 14, 25, 26, 44). Some work has been done in utilizing a digital computer to analyze the real-time spectral data and to solve the recognition problem by using correlation techniques. (22, 23, 24).

Various other methods have been used to try to characterize speech. These included amplitude-dichotomization, time-quantization as reported by Licklider (36) and short-term autocorrelation analysis as reported by Biddulph (2) and Davenport (7).

It has been noted that investigators recognized that there is a slow frequency modulation that occurs in speech and that this modulation occurs at syllabic rates which are limited to a maximum value of about 15 cps. The importance of this variation to the intelligence contained in the sound is evident in the work of Peterson (41), Kork and Miller (34), and Harris (29).

The purpose of the investigation reported here was to determine what information structure was present in the very low frequency characteristics of speech. The band of frequencies from 0-15 cycles per second must contain this structure because muscle control is limited to this syllabic rate. Therefore, it was proposed to study the amplitude and frequency characteristic of a speech signal within this range of frequencies. A very simple vocabulary of the ten spoken digits, zero to nine, was used throughout the investigation.

METHOD OF APPROACH

The problem as outlined in the introduction was essentially to determine what information structure exists in the band of frequencies from 0-15 cycles per second for speech. Because the body is not capable of physical change or motion at rates greater than about 15 cycles per second, the upper limit of the frequency band was fixed by this consideration. A pair of signals was derived for each spoken word, in this case one of the ten digits. One of the signals is a combination of components of the amplitude envelope. The other is composed of components of the low frequency modulation. In both signals the component frequencies are limited to the band of 0-15 cycles per second. These two signals are an analog representation of the spoken digit.

A digital representation of the analog signals was obtained by sampling the signals at uniform intervals of time. Actually this task was performed by hand although commercial equipment is available which will perform this function. The digital approximations to the signals were then used to determine and define the information structure in the digit characterizations. The Cyclone digital computer was used to aid in the evaluation of the information content of the derived signals.

Correlation studies with the digital computer appeared to lead to a promising model for the comprehension process for the simple vocabulary. Initial efforts were directed to grouping or classifying the signals from a single speaker. Later the invariance of these signals among speakers was checked. The structure of the time rate of change of the frequency modulation signal was examined and found to contain a great deal of information or intelligence. This finding suggested a technique which was used to duplicate the results of an experiment reported by the Bell Telephone Laboratories staff (42) but conducted in an entirely different way. Attempts were made to determine how much of the complete sound spectrum contained the information represented by these characterizations. The selected frequency band was also investigated to determine if the full band from 0-15 cycles per second was necessary for the information structure found.

EQUIPMENT

A block diagram of the measuring equipment used to obtain the amplitude and frequency signals for a given spoken digit is shown in Figure 1. The Cyclone digital computer and its peripheral equipment make up the rest of the complete system, with a human link performing the necessary matching functions between the analog and digital systems. An Electro-voice, model 647, dynamic microphone was used at the input of the system of Figure 1. The audio amplifier shown in the diagram was a Knight 30 watt, high fidelity amplifier, model KN-530, modified to operate into the next blocks of the system. The amplifier was operated with the rolloff response eliminated and the extended high frequency boosted with the treble control. The bass control has essentially no effect on the system response.

The complete circuits contained in the next blocks of the system, those which actually form the amplitude and frequency characterization signals, are shown in Figure 2. An output from the audio amplifier is taken from the plates of the output driver stages to form the amplitude signal. This output first is rectified in the bridge diode rectifier and then filtered. An output representing the very low frequency voltage E_A components in the amplitude envelope, is developed across the filter capacitor. The cutoff frequency of the RC filter is essentially 15 cycles per second.

The output transformer of the audio amplifier is used to drive 10 General Ceramic ferrite cores. These cores are type S3 material

Figure 1. Block diagram of the system to record amplitude and frequency characterizations.

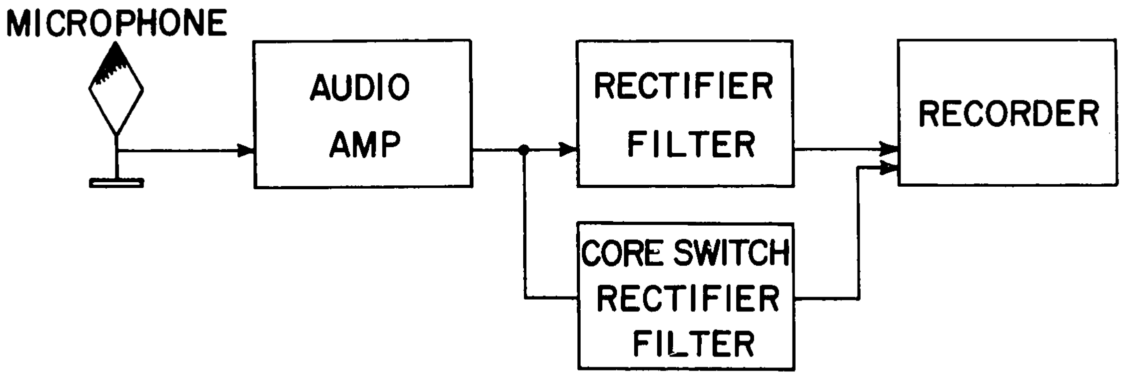
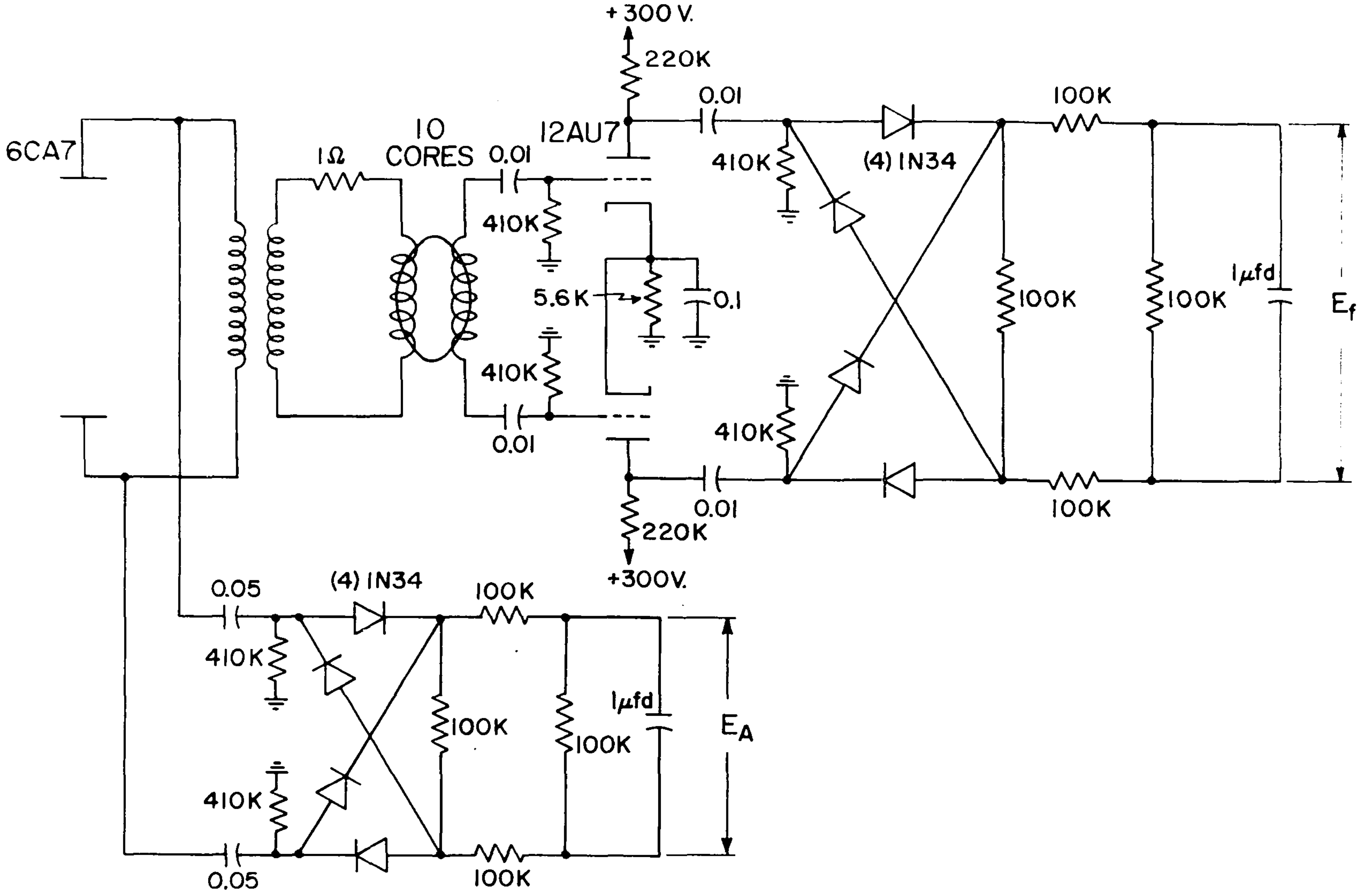

MICROPHONE
AUDIO AMP
RECTIFIER FILTER
CORE SWITCH RECTIFIER FILTER
RECORDER

Figure 2. Circuitry for deriving the amplitude and frequency characterization signals.


6CA7
1Ω
10 CORES
0.01
12AU7
+300 V.
220K
0.01
410K
(4) 1N34
100K
410K
5.6K
0.1
410K
0.01
410K
100K
100K
1μfd
E_f
0.01
220K
+300V.
100K
0.05
(4) 1N34
100K
410K
100K
100K
1μfd
E_A
410K
0.05
100K

and need a drive of about 300 milliampere-turns to be switched. The ten cores are simply stacked with ten turns on the primary winding and ten turns on the secondary winding. The output of the cores is amplified, rectified, and then filtered. The voltage E_f then is a linear function of the frequency of the drive, and changes in E_f indicate the frequency modulation present in the input to the core switch. The gain of the audio amplifier was adjusted so that it was just below the threshold at which noise would switch the cores. The RC filter in this circuit also has a nominal cutoff of 15 cycles per second.

The relationship between the amplitude output voltage, E_A, and the input to the phase inverter stage of the audio amplifier is shown in Figure 3. The gain falls off at low frequencies in the audio amplifier but at high frequencies is essentially constant. This variation of gain with frequency has little effect on the operation of the system as long as there is sufficient input to switch the cores in the frequency detection circuitry. It will be shown that the derived amplitude signal contains virtually no information. The variation of the voltage, E_f, versus frequency of a constant-amplitude input to the phase inverter stage of the audio amplifier is shown in Figure 4. It can be seen that for input signals with magnitudes over one volt there is sufficient drive to prevent non-linearity below about 10 kilocycles per second, a frequency normally considered above the range of most speech. It can also be seen that E_f is a linear function of the input frequency

Figure 3. Output voltage, E_A , versus input voltage as a function of frequency.

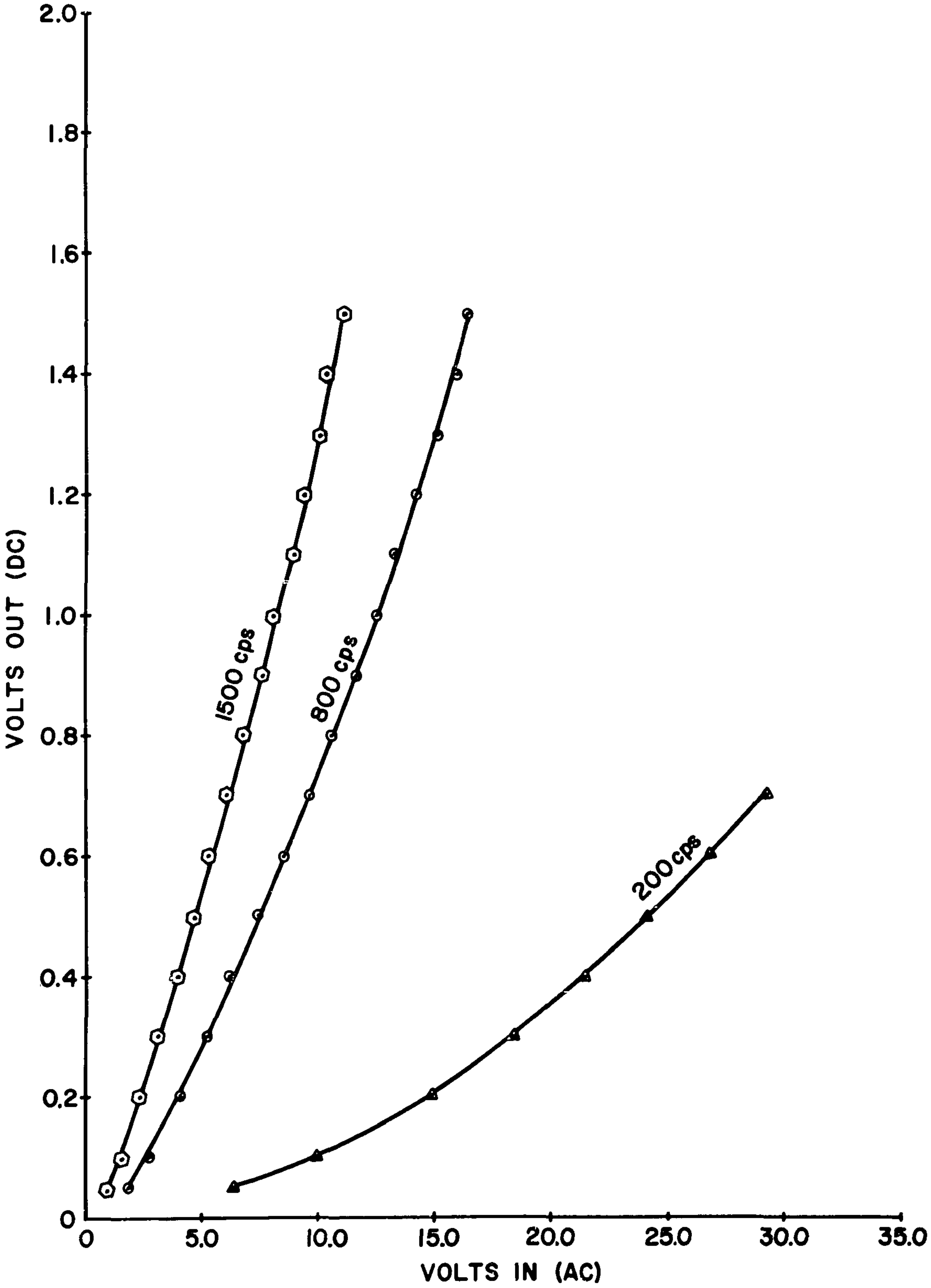
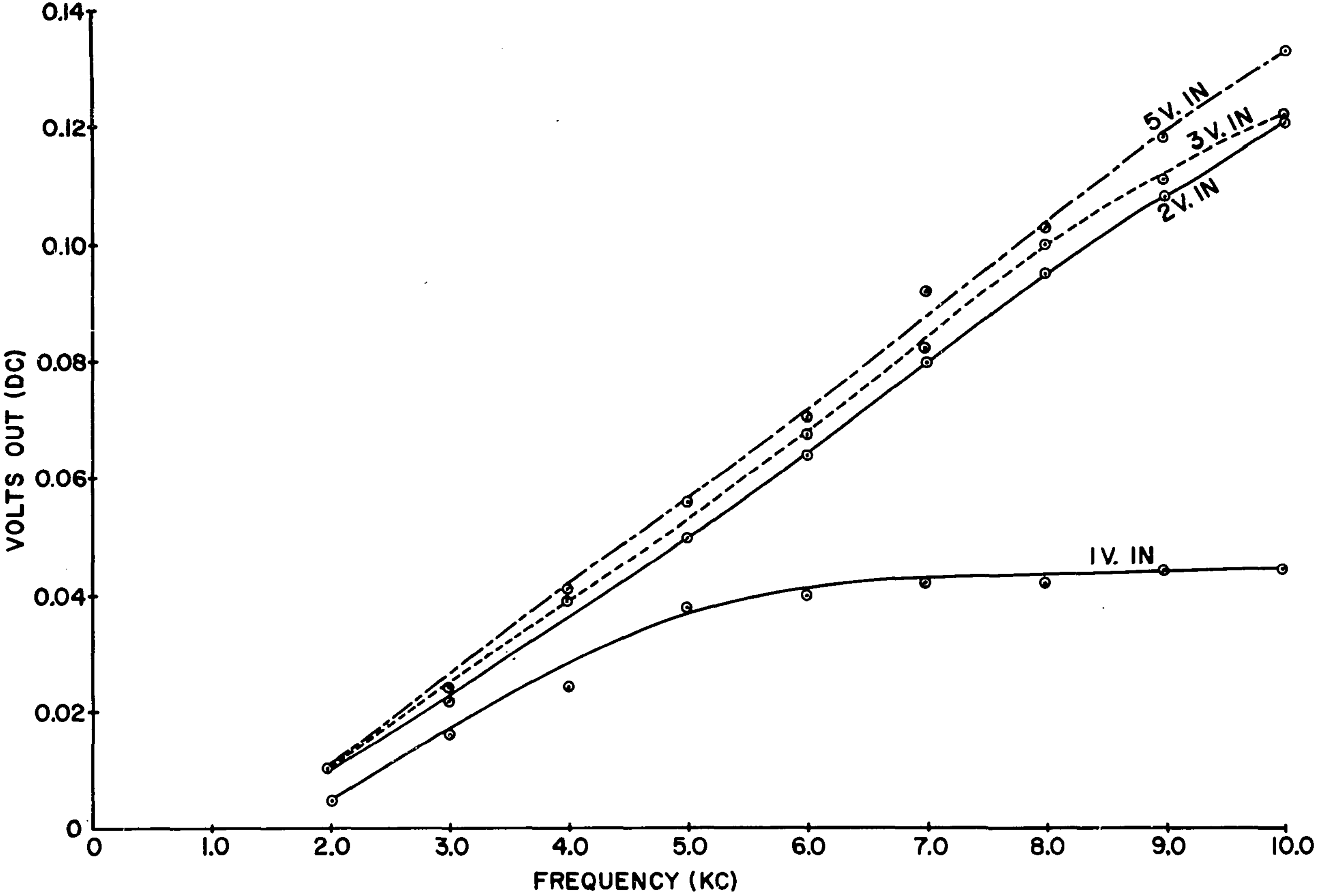

2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
VOLTS OUT (DC)
0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
VOLTS IN (AC)
1500 cps
800 cps
200 cps

Figure 4. Output voltage, E_f , versus frequency for a constant input.



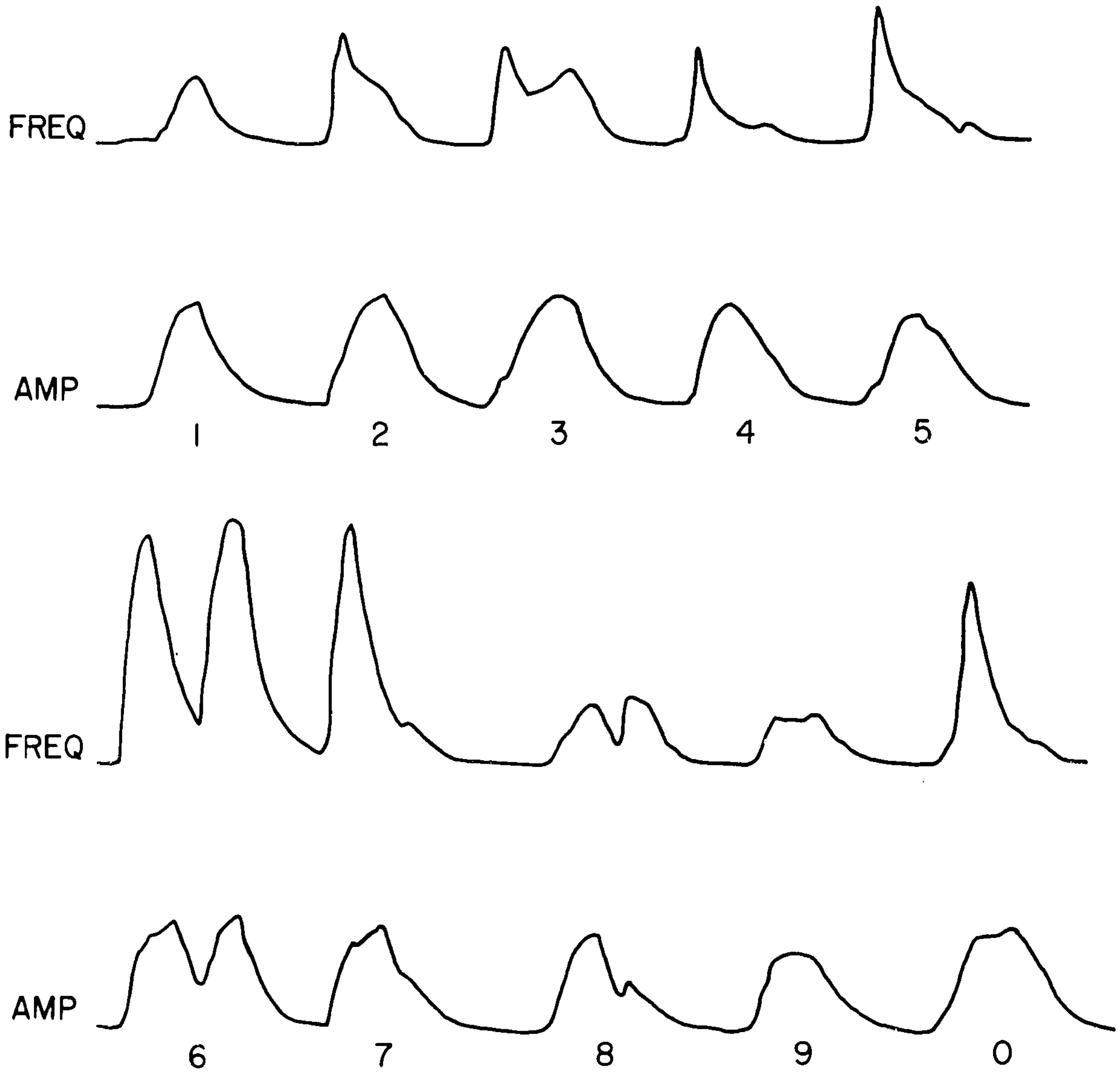
5V. IN
3V. IN
2V. IN
1V. IN
VOLTS OUT (DC)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
FREQUENCY (KC)

over the range of speech frequencies.

The outputs of the two filters, E_A and E_f , are then plotted using a two channel Sanborn recorder, model 60-1300. This recorder has a cutoff frequency of 40 cycles per second, and therefore the amplitude and frequency signals are well within the range of flat response. For several of the conducted tests the preamplifier in the recorder was used to increase the level of the frequency signal. Since the signals were all normalized in the correlation studies, the absolute amplitude values of the signals were not important except that the gain was kept constant throughout any given recording run. Figure 5 shows typical amplitude and frequency characteristics for the spoken digits "zero" to "nine" obtained from the Sanborn recorder running at a speed of 50 millimeters per second.

The correlations studies of the speech characterization signals were done on the Cyclone digital computer, a part of the computational facilities at Iowa State University of Science and Technology. This machine is a medium speed, general purpose, Illiac-type computer. The analog information was first converted to a digital representation by essentially sampling the signals at a 50 cycles per second rate. The conversion was actually done by hand although it could have been implemented more conveniently with available commercial equipment. The digital information was used to perform the correlation studies discussed in the next section.

Figure 5. Amplitude and frequency characteristics of spoken digits obtained from the output of the Sanborn recorder.


FREQ
AMP
1
2
3
4
5
FREQ
AMP
6
7
8
9
0

CORRELATION STUDIES OF THE SPEECH CHARACTERIZATION SIGNALS

In order to determine what information structure was contained in the derived signals, it was decided that a classification scheme was necessary. Since each spoken digit can be uniquely recognized by a listener, then the degree of information contained in a particular derived signal could be measured by the recognizability of this signal. It was therefore proposed to run correlation studies on these signals in order to determine their degree of similarity or dissimilarity.

As mentioned in the previous section, the Cyclone digital computer was used to perform these studies. Library routine K-2, entitled, "Product Moment Correlations, Means, Standard Deviations, Variances and Covariances", was used to perform the calculations. The library routine was read into the machine, followed by a parameter tape and lastly the data tape. The parameter tape specified the sample size, the number of variables, the number of significant digits to be retained in the correlation matrix for output and the number of significant digits to be retained for the means, standard deviations, and variance-covariance matrix for output. Either of the latter outputs could be suppressed by indicating zero significant digits.

The product moment correlation coefficient is a measure of the degree of relation of two variables. It can be shown to range between +1 and -1. This program computed the matrix of product moment correlations between each pair of a set of variables. Because this matrix is symmetrical, it was necessary to print out only half of the off-diagonal coefficients.

The product moment correlation coefficient may be written in terms of the observed data as

$$r_{x\,y} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\left[\Sigma(x - \bar{x})^2\,\Sigma(y - \bar{y})^2\right]^{\frac{1}{2}}}$$

For computational purposes this can be rewritten in terms of x, y, xy, and s, the sample size as

$$r_{x\,y} = \frac{s\ \Sigma\ xy - \Sigma\ x\,\Sigma\ y}{\left(\left[s\,\Sigma\ x^2 - (\Sigma x)^2\right]\ \left[s\Sigma y^2 - (\Sigma y)^2\right]\right)^{\frac{1}{2}}}$$

By using this form the observation points can be stored in the memory one at a time, the sums and product-sums being formed point by point. When the observations have all been read into the machine, the correlations are calculated and the matrix of coefficients is printed out.

The frequency and amplitude correlation matrices for the ten digit vocabulary as spoken by the author are given in Tables 1a and 1b. These matrices were calculated from recorded data similar to that shown in Figure 5. When analyzing the speech of a single person, the sampling interval was determined by the length of the longest signal. In every case this was the representation of "six". The amplitude signal, because of its initial, very positive rise, was used to initiate the sampling process and its trailing edge to stop the process. An amplitude level was selected such that the only change occurring was due to the decay time of the RC filter, and when the amplitude signal had dropped to this level, the sampling of both signals was stopped.

It is evident from the frequency matrix, Table 1a, that the frequency characteristics can be classified in the following groups:

Group a - 1-3-4-5-9
Group b - 2-7-0
Group c - 6
Group d - 8

A minimum value of r_{xy} of +0.8000 was used to determine the above groups. That is, those coefficients with values greater than +0.8000 indicated the digits belonged to the same basic group. It was determined by examination of a large number of correlation matrices that this was the highest value that could be used reliably in making this decision. For some circumstances even lower values may be dictated.

The amplitude matrix, Table 1b, is interesting because of its uniformity among the calculated coefficients. From the matrix it would appear that the amplitude characteristic of "six" is the only one that materially differs from the others. This is verified by Figure 5. Consequently, it would be impossible to distinguish the other nine digits on the basis of the amplitude characteristic alone. This implies that there is very little information structure

Table 1a. Frequency correlation matrix for the ten digits spoken by a single speaker.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.5089	+0.9587	+0.8594	+0.9859	-0.1003	+0.5226	+0.7338	+0.9752	+0.7920
	+1.0000	+0.4914	+0.4069	+0.4620	+0.1761	+0.9130	+0.4719	+0.3681	+0.6914
		+1.0000	+0.9534	+0.9263	-0.2061	+0.4457	+0.6552	+0.9646	+0.6778
			+1.0000	+0.8061	-0.2919	+0.3552	+0.5562	+0.8874	+0.5989
				+1.0000	-0.1591	+0.4581	+0.7313	+0.9674	+0.7514
					+1.0000	+0.3014	+0.2043	-0.1736	+0.1744
						+1.0000	+0.5195	+0.3690	+0.8389
							+1.0000	+0.7159	+0.7552
								+1.0000	+0.6775
									+1.0000

Table 1b. Amplitude correlation matrix for the ten digits spoken by a single speaker.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.9647	+0.9609	+0.9665	+0.9495	+0.4462	+0.9591	+0.9063	+0.9540	+0.9162
	+1.0000	+0.9306	+0.9102	+0.9127	+0.3805	+0.9803	+0.9000	+0.8855	+0.8295
		+1.0000	+0.9874	+0.9940	+0.3640	+0.9446	+0.8795	+0.9813	+0.9625
			+1.0000	+0.9888	+0.3822	+0.9308	+0.8840	+0.9912	+0.9781
				+1.0000	+0.3762	+0.9330	+0.8996	+0.9896	+0.9753
					+1.0000	+0.4043	+0.5109	+0.4220	+0.4043
						+1.0000	+0.9099	+0.9069	+0.8537
							+1.0000	+0.8996	+0.8655
								+1.0000	+0.9905
									+1.0000

in the amplitude envelope components of frequencies less than about 15 cycles per second. Therefore, most of the remainder of the investigation was concerned with only the frequency signal.

It has been shown that the frequency characteristics of the ten digits for a single speaker can be classified into four or five groups. It is evident that it would be possible to only make a group designation of an unknown signal rather than a unique digit designation using only the derived frequency characteristic. The next problem investigated was the invariance of these characteristics among several speakers. The recorded signals would be expected to vary because of the differences of pitch among speakers. However, the signals are normalized by the computer program as the problem is run, and it is the information structure of the normalized signal that is of interest.

The correlation matrices of the frequency signal for each of the ten digits as spoken by ten men and women are given in Tables 2a - 2j. Six of the ten persons are men. No attempt was made to classify the language characteristics of the speakers themselves. A midwest speech characteristic was common in all the subjects.

The time base for any one digit frequency signal was made the same for all ten speakers. This meant that some patterns were condensed while others were elongated. Essentially the average length for a given digit was used as the reference to which all time bases were changed. Normally only very small changes were necessary except

Table 2a. Correlation matrix of the frequency signal for "one" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola *	Reuter*	Magee*	Anderson*
+1.0000	+0.9546	+0.9811	+0.7899	+0.9447	+0.9560	+0.9917	+0.9505	+0.8621	+0.9596
	+1.0000	+0.9345	+0.8907	+0.9582	+0.9746	+0.9670	+0.8428	+0.9526	+0.9503
		+1.0000	+0.8090	+0.8944	+0.9148	+0.9643	+0.9219	+0.8428	+0.9654
			+1.0000	+0.8091	+0.8734	+0.8047	+0.6272	+0.9531	+0.8972
				+1.0000	+0.9739	+0.9567	+0.8839	+0.9075	+0.8940
					+1.0000	+0.9681	+0.8632	+0.9442	+0.9442
						+1.0000	+0.9376	+0.8756	+0.9599
							+1.0000	+0.7133	+0.8713
								+1.0000	+0.9106
									+1.0000

*Female

Table 2b. Correlation matrix of the frequency signal for "two" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Reuter	Magee	Anderson
+1.0000	+0.9688	+0.9856	+0.8921	+0.9038	+0.6407	+0.9459	+0.9299	+0.9769	+0.8475
	+1.0000	+0.9768	+0.9076	+0.7891	+0.5042	+0.8794	+0.9276	+0.9657	+0.8695
		+1.0000	+0.8947	+0.8524	+0.5213	+0.9139	+0.9672	+0.9796	+0.8885
			+1.0000	+0.7577	+0.5299	+0.9089	+0.8128	+0.9449	+0.6618
				+1.0000	+0.8142	+0.9491	+0.7923	+0.8561	+0.6336
					+1.0000	+0.7606	+0.3518	+0.5436	+0.2300
						+1.0000	+0.8260	+0.9383	+0.6701
							+1.0000	+0.9418	+0.9279
								+1.0000	+0.8394
									+1.0000

Table 2c. Correlation matrix of the frequency signal for "three" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.8848	+0.7633	+0.6292	+0.5350	+0.6780	+0.8843	+0.6768	+0.8234	+0.4329
	+1.0000	+0.5565	+0.2942	+0.7586	+0.8890	+0.9089	+0.4477	+0.9539	+0.5618
		+1.0000	+0.8926	+0.3605	+0.3345	+0.7071	+0.9618	+0.5251	+0.2860
			+1.0000	+0.1114	+0.0484	+0.5206	+0.8886	+0.2136	+0.1590
				+1.0000	+0.9206	+0.7744	+0.2286	+0.7503	+0.9068
					+1.0000	+0.8189	+0.1885	+0.8997	+0.7735
						+1.0000	+0.6265	+0.8777	+0.6242
							+1.0000	+0.4393	+0.1881
								+1.0000	+0.5615
									+1.0000

Table 2d. Correlation matrix of the frequency signal for "four" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9298	+0.6653	+0.2847	+0.2906	+0.5361	+0.5051	+0.4578	+0.1009	-0.4747
	+1.0000	+0.7391	+0.2951	+0.3229	+0.6372	+0.6732	+0.5028	+0.2483	-0.2869
		+1.0000	+0.7022	+0.4796	+0.8562	+0.8581	+0.8993	+0.1506	-0.3910
			+1.0000	+0.7462	+0.7446	+0.5619	+0.8639	+0.2272	-0.3225
				+1.0000	+0.8029	+0.5807	+0.6025	+0.6124	+0.0503
					+1.0000	+0.8252	+0.7942	+0.3629	-0.2252
						+1.0000	+0.8114	+0.4878	+0.0964
							+1.0000	+0.2394	-0.2885
								+1.0000	+0.6604
									+1.0000

Table 2e. Correlation matrix of the frequency signal for "five" as spoken by ten men and women.

Schauer	Rickey	Casson	MCConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.8868	+0.3134	+0.9424	+0.7030	+0.4005	-0.0530	+0.6703	-0.0293	+0.1904
	+1.0000	+0.4783	+0.8020	+0.7205	+0.5924	+0.2365	+0.6730	+0.2493	+0.4098
		+1.0000	+0.2182	+0.7339	+0.9540	+0.6906	+0.4985	+0.7557	+0.7557
			+1.0000	+0.5894	+0.3109	-0.0569	+0.7699	-0.0468	+0.1628
				+1.0000	+0.8169	+0.3597	+0.6952	+0.4685	+0.6689
					+1.0000	+0.7787	+0.6361	+0.8324	+0.8751
						+1.0000	+0.4464	+0.9336	+0.8467
							+1.0000	+0.5133	+0.6673
								+1.0000	+0.9490
									+1.0000

Table 2f. Correlation matrix of the frequency signal for "six" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	-0.1927	+0.1760	+0.4160	+0.0179	+0.6600	+0.0810	+0.6461	+0.0639	+0.6863
	+1.0000	+0.6616	-0.2583	+0.4235	-0.1020	+0.2132	-0.0714	+0.5402	-0.2810
		+1.0000	+0.1396	+0.7727	+0.2407	+0.6970	+0.4343	+0.6785	-0.0103
			+1.0000	+0.5648	+0.8594	+0.6122	+0.7922	+0.5391	+0.8418
				+1.0000	+0.5290	+0.8898	+0.5030	+0.8812	+0.2288
					+1.0000	+0.5228	+0.7452	+0.5618	+0.8658
						+1.0000	+0.5433	+0.6474	+0.2445
							+1.0000	+0.4910	+0.8162
								+1.0000	+0.3326
									+1.0000

Table 2g. Correlation matrix of the frequency signal for "seven" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9242	+0.9746	+0.7994	+0.9477	+0.7503	+0.8645	+0.8666	+0.6251	+0.9881
	+1.0000	+0.9522	+0.9004	+0.9759	+0.9152	+0.9665	+0.9829	+0.8342	+0.9283
		+1.0000	+0.8642	+0.9854	+0.8392	+0.9200	+0.9160	+0.7335	+0.9737
			+1.0000	+0.9122	+0.9394	+0.8841	+0.9168	+0.9386	+0.8311
				+1.0000	+0.8967	+0.9624	+0.9588	+0.8111	+0.9626
					+1.0000	+0.9213	+0.9563	+0.9634	+0.7720
						+1.0000	+0.9638	+0.8487	+0.8942
							+1.0000	+0.8895	+0.8741
								+1.0000	+0.6634
									+1.0000

Table 2h. Correlation matrix of the frequency signal for "eight" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.6065	+0.6072	+0.1778	+0.6776	+0.7033	+0.4516	+0.4422	+0.6296	+0.3081
	+1.0000	+0.9164	+0.1511	+0.9828	+0.6896	+0.7779	+0.7706	+0.9877	+0.5468
		+1.0000	+0.2773	+0.9417	+0.7889	+0.9056	+0.9391	+0.9472	+0.3002
			+1.0000	+0.2387	+0.4166	+0.4983	+0.2970	+0.2242	-0.1712
				+1.0000	+0.7768	+0.8324	+0.7889	+0.9878	+0.4714
					+1.0000	+0.8646	+0.7031	+0.7727	-0.0434
						+1.0000	+0.8675	+0.8419	-0.0243
							+1.0000	+0.8244	+0.1535
								+1.0000	+0.4631
									+1.0000

Table 2i. Correlation matrix of the frequency signal for "nine" as spoken by ten men and women

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.5762	+0.7634	+0.5176	+0.7320	+0.5712	+0.6223	+0.7591	+0.7944	+0.9414
	+1.0000	+0.8637	+0.9552	+0.9336	+0.9816	+0.9807	+0.8212	+0.9269	+0.7864
		+1.0000	+0.7564	+0.9558	+0.8910	+0.8702	+0.9304	+0.8716	+0.9148
			+1.0000	+0.8863	+0.4467	+0.9521	+0.6753	+0.9119	+0.7207
				+1.0000	+0.9613	+0.9442	+0.8555	+0.9472	+0.8955
					+1.0000	+0.9700	+0.7951	+0.9144	+0.7857
						+1.0000	+0.8454	+0.9555	+0.8246
							+1.0000	+0.8381	+0.8965
								+1.0000	+0.9208
									+1.0000

Table 2j. Correlation matrix of the frequency signal for "zero" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9792	+0.9079	+0.7277	+0.8761	+0.9180	+0.4739	+0.8831	+0.6741	+0.8946
	+1.0000	+0.8503	+0.8082	+0.9084	+0.9252	+0.4015	+0.9412	+0.7824	+0.9492
		+1.0000	+0.6797	+0.8708	+0.8943	+0.7686	+0.7236	+0.5622	+0.7546
			+1.0000	+0. 9164	+0.8261	+0.3871	+0.8787	+0.9625	+0.8957
				+1.0000	+0.9574	+0.6156	+0.8993	+0.8735	+0.9191
					+1.0000	+0.6193	+0.9132	+0.8091	+0.9287
						+1.0000	+0.2846	+0.2586	+0.3232
							+1.0000	+0.9089	+0.9912
								+1.0000	+0.9139
									+1.0000

in those cases where the person spoke very rapidly. Changing the time base of some signals did cause a shift in the characteristic structure of the signals. This accounts for some of the apparently poor correlation for some digits.

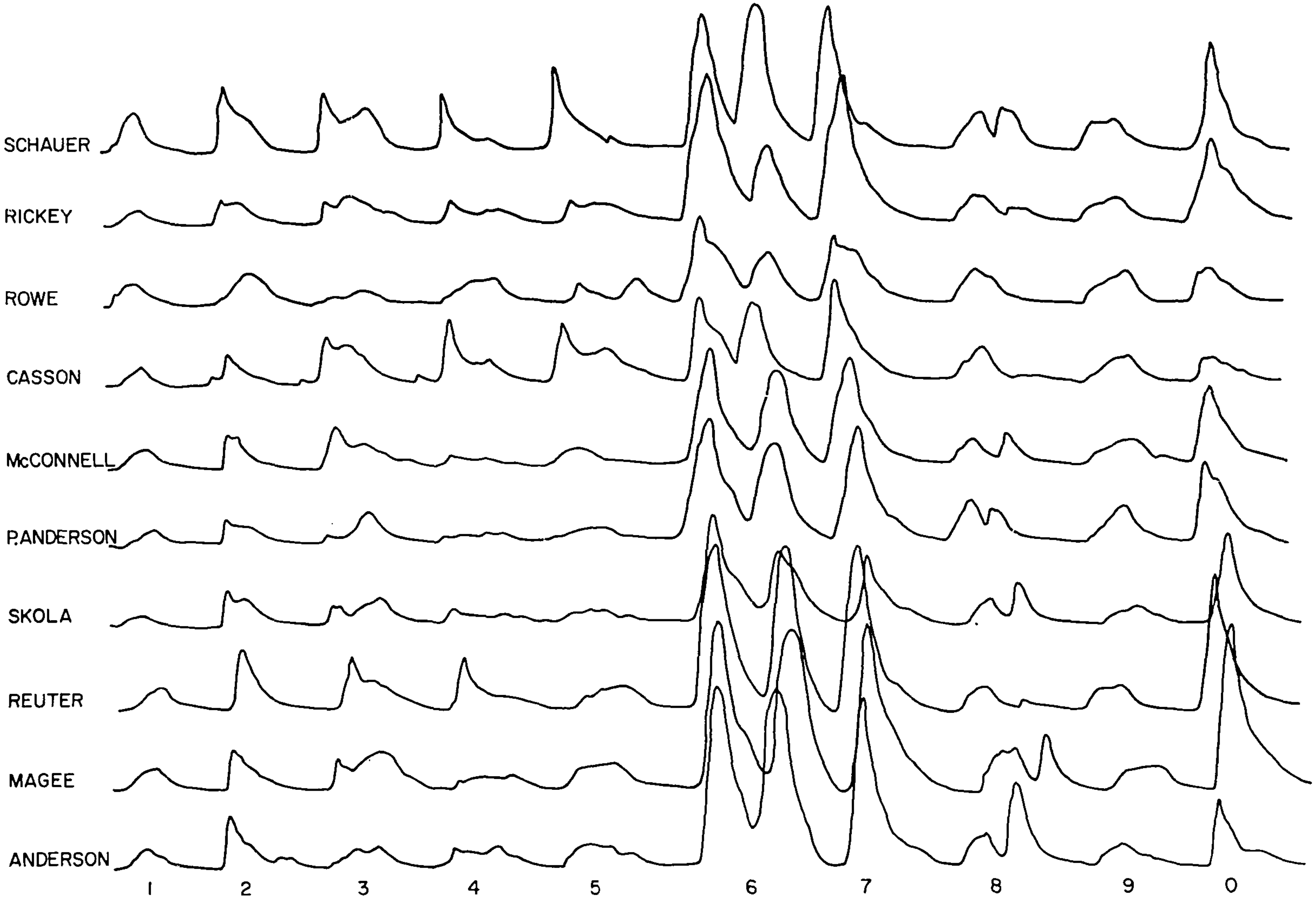
The most interesting finding of this particular study is that there is no way of distinguishing between male and female speakers. Examination of the ten matrices revealed that certain groupings were possible. From Table 2a it is apparent that all signals are highly correlated, meaning that essentially the same structure exists in the frequency signal of "one" for all ten speakers. For "two" all are highly correlated except that of P. Anderson. Two groupings were possible for the digit "three". The first included Schauer, Rickey, Rowe, P. Anderson, Skola, Magee, and Anderson, while the second was made up of Casson, McConnell and Reuter. The evaluation of the correlation matrix for "four" is not as clear cut as some of the others. Two small groups are possible. Schauer and Rickey can be classified together and Casson, P. Anderson, Skola, and Reuter. McConnell, Rowe, Magee and Anderson are not highly correlated with any other signal.

Table 2e showed that for "five" Schauer, Rickey, and McConnell are highly correlated while all the others, excluding Reuter, are grouped together. The correlation coefficients for McConnell, P. Anderson, Reuter, and Anderson indicated one group for the digit "six" while Rowe, Skola, and Magee made up another. The rest are not highly correlated with any other signal. Table 2g indicated all signals are highly correlated for "seven". For "eight" only Schauer, McConnell, and Anderson are not highly correlated with any other signal. All but Schauer were grouped for "nine", and for "zero" all but Skola.

It is apparent from the above that, except for several instances, the majority of the signals from different speakers for a given digit were highly correlated. There also were no well defined divisions among the groups of speakers. This would indicate that the effects of variables such as pitch have been minimized in this speech characterization technique. The enunciation of the digit by any speaker appeared to have the most effect on the frequency characteristic.

The above correlative evaluation of individual speech characteristic was repeated with the additional condition that an attempt was made to get all speakers to talk at essentially the same speed. The characteristics were analyzed and are shown in Figure 6. In the recording phase the speakers repeated the digits after the author. Only the repeated digits were recorded however. It was likely that this technique also introduced some "learning" in the speakers which would affect enunciation particularly. However, part of the purpose of this second run was to determine the effects of enunciation on the characteristics. The analysis of the data was made in exactly the same way as the previously reported run. The results of the correlation study are given in Tables 3a - 3j.

Figure 6. Frequency characterization for the ten digit vocabulary for ten different speakers.


SCHAUER
RICKEY
ROWE
CASSON
McCONNELL
P. ANDERSON
SKOLA
REUTER
MAGEE
ANDERSON
1
2
3
4
5
6
7
8
9
0

The "one" signals were all highly correlated as can be seen from Table 3a. Table 3b indicated a high degree of correlation for "two" among the speakers except for McConnell and Anderson. From Figure 6 it can be seen that even for these two speakers the general shape is the same but the trailing edges fall off more rapidly. The signals representing "three" are much less highly correlated. In fact, three groupings can be made from the correlation matrix, Table 3c, or Figure 6. Schauer, Rickey, Casson, Rowe, and Reuter can be grouped on the basis of the correlation matrix. In general, these signals can be described as having an initial peak, falling to some nearly steady value, and then decaying toward the residual value. P. Anderson, Skola, and Magee form a second group in which the signals can be thought of as having an initial peak, falling to a minimal value, rising to the maximum signal level, and then decaying. The signals of McConnell and Anderson, the third group, are best described as being rather broad with only a slight dip in the amplitude.

Two groups also can be obtained for "four" from Table 3d. The first group, characterized by a signal with an initial peak, consists of Schauer, Rickey, Casson, Skola, and Reuter. A second group, McConnell, P. Anderson, Magee and Anderson, have a signal of nearly constant value throughout the region of interest. The enunciation of the fricative consonants appears to determine this grouping. Only the characteristics for Schauer and Casson are not highly correlated for "five". Again the emphasis of the

Table 3a. Correlation matrix of the frequency signal for "one" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9879	+0.9330	+0.9590	+0.9284	+0.9019	+0.9730	+0.9455	+0.8773	+0.9658
	+1.0000	+0.9227	+0.9646	+0.9257	+0.8877	+0.9656	+0.9418	+0.8780	+0.9653
		+1.0000	+0.9758	+0.9886	+0.9876	+0.9606	+0.9877	+0.9692	+0.8797
			+1.0000	+0.9860	+0.9624	+0.9783	+0.9852	+0.9636	+0.9449
				+1.0000	+0.9857	+0.9678	+0.9923	+0.9839	+0.9026
					+1.0000	+0.9476	+0.9761	+0.9701	+0.8550
						+1.0000	+0.9777	+0.9473	+0.9646
							+1.0000	+0.9777	+0.9159
								+1.0000	+0.8755
									+1.0000

Table 3b. Correlation matrix of the frequency signal for "two" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9586	+0.9449	+0.7786	+0.8530	+0.9532	+0.9689	+0.9658	+0.9732	+0.8062
	+1.0000	+0.8976	+0.8708	+0.9066	+0.9517	+0.9673	+0.9445	+0.9206	+0.7620
		+1.0000	+0.7228	+0.8969	+0.8362	+0.8718	+0.9541	+0.9737	+0.7338
			+1.0000	+0.9166	+0.7599	+0.8223	+0.6936	+0.7277	+0.3899
				+1.0000	+0.7583	+0.8273	+0.8285	+0.8635	+0.4924
					+1.0000	+0.9788	+0.9321	+0.8756	+0.8644
						+1.0000	+0.9246	+0.9070	+0.7907
							+1.0000	+0.9578	+0.8852
								+1.0000	+0.7731
									+1.0000

Table 3c. Correlation matrix of the frequency signal for "three" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.6531	+0.8418	+0.3788	+0.7832	+0.4171	+0.7031	+0.7334	+0.7391	+0.5707
	+1.0000	+0.8429	+0.5390	+0.5964	+0.1702	+0.2339	+0.6654	+0.5110	+0.6069
		+1.0000	+0.3157	+0.8414	+0.1058	+0.3822	+0.8504	+0.4725	+0.4561
			+1.0000	+0.2454	+0.7997	+0.5569	+0.2754	+0.7628	+0.8692
				+1.0000	+0.1022	+0.5615	+0.9825	+0.3989	+0.4312
					+1.0000	+0.7218	+0.0489	+0.8508	+0.7261
						+1.0000	+0.4877	+0.8369	+0.6424
							+1.0000	+0.3528	+0.4573
								+1.0000	+0.7702
									+1.0000

Table 3d. Correlation matrix of the frequency signal for "four" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Reuter	Magee	Anderson
+1.0000	+0.3194	+0.9192	-0.3861	+0.2504	+0.0027	+0.6554	+0.6905	+0.1180	-0.2244
	+1.0000	+0.6017	+0.4239	+0.9349	+0.5958	+0.8775	+0.8197	+0.6626	+0.7002
		+1.0000	-0.2216	+0.5134	+0.1510	+0.8449	+0.8858	+0.2445	-0.0062
			+1.0000	+0.3889	+0.8658	+0.1393	-0.1161	+0.7430	+0.8239
				+1.0000	+0.5591	+0.8135	+0.7860	+0.7338	+0.7364
					+1.0000	+0.4232	+0.1872	+0.8297	+0.7976
						+1.0000	+0.9293	+0.5042	+0.4254
							+1.0000	+0.3357	+0.2409
								+1.0000	+0.8345
									+1.0000

Table 3e. Correlation matrix of the frequency signal for "five" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.3755	+0.7661	-0.2240	+0.2257	+0.2029	-0.0429	+0.0348	+0.2264	+0.3277
	+1.0000	+0.7970	+0.7273	+0.8742	+0.8750	+0.8508	+0.8723	+0.9452	+0.8653
		+1.0000	+0.2934	+0.5500	+0.5786	+0.4761	+0.4796	+0.6578	+0.6000
			+1.0000	+0.7647	+0.8424	+0.9575	+0.9365	+0.8474	+0.6835
				+1.0000	+0.9206	+0.8731	+0.9074	+0.9471	+0.9561
					+1.0000	+0.9346	+0.9587	+0.9714	+0.8309
						+1.0000	+0.9752	+0.9462	+0.7901
							+1.0000	+0.9631	+0.8388
								+1.0000	+0.8911
									+1.0000

Table 3f. Correlation matrix of the frequency signal for "six" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.6044	+0.9481	+0.6849	+0.9016	+0.7334	+0.8292	+0.8395	+0.8204	+0.8164
	+1.0000	+0.6828	+0.9431	+0.7632	+0.8312	+0.7901	+0.8035	+0.5885	+0.5556
		+1.0000	+0.7862	+0.8640	+0.7350	+0.8626	+0.8199	+0.7699	+0.8385
			+1.0000	+0.8368	+0.8454	+0.8898	+0.8544	+0.7015	+0.6017
				+1.0000	+0.8912	+0.9326	+0.9723	+0.9455	+0.6278
					+1.0000	+0.7845	+0.9458	+0.8105	+0.4271
						+1.0000	+0.8905	+0.8907	+0.6879
							+1.0000	+0.9198	+0.5276
								+1.0000	+0.4763
									+1.0000

Table 3g. Correlation matrix of the frequency signal for "seven" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.8268	+0.9777	+0.6506	+0.8439	+0.8127	+0.9799	+0.9622	+0.9873	+0.9940
	+1.0000	+0.8285	+0.8776	+0.9977	+0.9963	+0.7893	+0.9384	+0.8666	+0.7903
		+1.0000	+0.6934	+0.8461	+0.8137	+0.9909	+0.9436	+0.9839	+0.9840
			+1.0000	+0.8777	+0.8837	+0.6584	+0.7574	+0.7358	+0.6128
				+1.0000	+0.9957	+0.8083	+0.9453	+0.8821	+0.8092
					+1.0000	+0.7738	+0.9260	+0.8512	+0.7723
						+1.0000	+0.9217	+0.9743	+0.9854
							+1.0000	+0.9732	+0.9430
								+1.0000	+0.9829
									+1.0000

Table 3h. Correlation matrix of the frequency signal for "eight" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.5817	+0.5667	+0.6523	+0.7944	+0.8788	+0.7568	+0.3628	+0.7020	+0.7583
	+1.0000	+0.3997	+0.7211	+0.6606	+0.8381	+0.5348	+0.9226	+0.6285	+0.0535
		+1.0000	+0.9052	+0.2227	+0.6221	+0.1651	+0.3253	+0.2882	+0.2949
			+1.0000	+0.4002	+0.8153	+0.2840	+0.6646	+0.4231	+0.1829
				+1.0000	+0.7821	+0.9303	+0.3761	+0.8752	+0.6603
					+1.0000	+0.7237	+0.6819	+0.7682	+0.4226
						+1.0000	+0.2464	+0.8914	+0.6581
							+1.0000	+0.3516	-0.2308
								+1.0000	-0.4758
									+1.0000

Table 3i. Correlation matrix of the frequency signal for "nine" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.7991	+0.7077	+0.7572	+0.5189	+0.6480	+0.8496	+0.8846	+0.8530	+0.9479
	+1.0000	+0.9752	+0.9849	+0.9107	+0.9462	+0.9639	+0.9662	+0.9802	+0.7454
		+1.0000	+0.9742	+0.9212	+0.9822	+0.9566	+0.9155	+0.9330	+0.6683
			+1.0000	+0.9081	+0.9477	+0.9429	+0.9598	+0.9557	+0.7033
				+1.0000	+0.9107	+0.8446	+0.8052	+0.8569	+0.4880
					+1.0000	+0.9327	+0.8606	+0.8838	+0.5709
						+1.0000	+0.9388	+0.9444	+0.8005
							+1.0000	+0.9705	+0.8343
								+1.0000	+0.8186
									+1.0000

Table 3j. Correlation matrix of the frequency signal for "zero" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.8452	+0.8657	+0.8583	+0.8608	+0.9089	+0.9138	+0.9732	+0.9304	+0.4274
	+1.0000	+0.8426	+0.8477	+0.6577	+0.7659	+0.6432	+0.8245	+0.7179	+0.1922
		+1.0000	+0.9766	+0.9128	+0.9461	+0.8748	+0.9350	+0.9223	+0.6320
			+1.0000	+0.9227	+0.9433	+0.8709	+0.9378	+0.9042	+0.6620
				+1.0000	+0.9707	+0.9739	+0.9465	+0.9635	+0.8064
					+1.0000	+0.9584	+0.9685	+0.9779	+0.6740
						+1.0000	+0.9604	+0.9854	+0.7242
							+1.0000	+0.9747	+0.5999
								+1.0000	+0.6775
									+1.0000

fricative consonant appears to account for this difference in the signals. For the digit "six" by Anderson seems to be the only badly correlated signal. This, from Figure 6, appears to be true because the second peak occurs sooner than for most of the other signals. The rounded leading edge of the McConnel "seven" signal explains the lack of high correlation for that signal with the others which are highly correlated. There are no clear cut groups from Table 3h for "eight". It appears from Figure 6 that the enunciation of the final "t" plays an important role in determining the characteristic of the digit. For "nine" the correlation is good except for Schauer and Anderson. In the case of Schauer, the signal rises more rapidly, and for Anderson, it falls off more slowly than most of the others. From Table 3j the signals of Rickey and Anderson are the poorest correlated. The "zero" of Rickey has a broader peak while that of Anderson falls back to a steady level. Most of the signals decay without a plateau. Again there was no line of distinction between male and female speakers although in some cases all the signals were not highly correlated.

It has previously been shown that essentially no information structure exists in the derived amplitude signals but that some information is carried in the frequency signal. At least enough information was present to permit groupings as reported in the discussion of Table 1a. Since it has been known that a steady tone loses its intelligibility if sustained for a long period of time, it was decided to further investigate the rate of change of the modulating frequencies. This investigation was also performed on the basis of correlation studies on the Cyclone digital computer. An approximation to the derivative of the frequency signal was used for the analysis. Since a digital representation was available, it was decided to use the first difference in the observations in lieu of the first derivative of the frequency signal.

For the frequency characteristics shown in Figure 5, the matrices of Tables 4a and 4b were calculated. It is obvious from Table 4a that only a broad classification is possible for each digit other than "six" and "eight". These two could be identified or recognized on the basis of the frequency signal alone. The correlation matrix of the first differences of the frequency signal is presented in Table 4b. Study of this matrix indicates that only the digits "two", "three", "four", and "five" have highly correlated first difference signals. This can be explained by observation of the emphasis on the enunciation of the initial consonants sounds of these digits. It appears that the characteristics, as shown in Figure 5, are actually exaggerated in this respect.

The pronunciation of the digits "five" and "nine" has been exaggerated in some voice communication systems in order to try to increase the probability of correct recognition of a sequence of digits. Essentially this has been accomplished by pronouncing the two digits as though they each had two syllables. The frequency and amplitude characteristics of the ten digits with the pronunciation of "five" and "nine" exaggerated are shown in Figure 7. A correlation study was run on the frequency signal and its first difference.

to determine what effect these changes might make in the information structure of these signals. The results of this study are listed in Tables 5a and 5b. From Table 5a, one can group the frequency signals as shown below:

Group a - 1-5-9

Group b - 2-3-4-7-0

Group c - 6

Group d - 8

Again it is evident that there is insufficient information in the frequency signal alone to uniquely identify the spoken digit. However, the correlation matrix for the first difference frequency has as its largest cross correlation coefficient a value of +0.7108. Therefore, on the basis of the first difference signals it would be possible to recognize any signal from any other using a correlation technique.

The first difference frequency correlation matrices were calculated for all signals shown in Figure 6. The matrix for the author is given in Table 4b and the other speakers in Tables 6a-6i. It will be noticed that generally the first differences are much less correlated than the frequency characteristics themselves. Depending upon the speaker, however, there may be several first difference signals which are too highly correlated to be distinguished one from another by using this technique. However, no attempt had been made to train these speakers to enunciate properly. In Tables 7a-7j the correlation matrices of the first difference frequency signals for each of the

Table 4a. Correlation matrix of the frequency signal for the ten digits.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.6321	+0.5454	+0.4423	+0.5765	+0.2173	+0.7657	+0.5801	+0.9168	+0.9021
	+1.0000	+0.7178	+0.9440	+0.9453	+0.2184	+0.8760	+0.3111	+0.8033	+0.6844
		+1.0000	+0.8009	+0.8346	-0.1566	+0.7161	+0.7248	+0.7702	+0.5754
			+1.0000	+0.9622	+0.1141	+0.8045	+0.3508	+0.6906	+0.5388
				+1.0000	+0.1834	+0.8952	+0.4372	+0.8036	+0.6734
					+1.0000	+0.3006	-0.1533	+0.1490	+0.3104
						+1.0000	+0.5286	+0.8751	+0.9153
							+1.0000	+0.5851	+0.6074
								+1.0000	+0.8716
									+1.0000

Table 4b. Correlation matrix of the first difference frequency signal for the ten digits.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.1397	-0.0942	-0.0244	-0.0096	+0.3747	+0.3283	+0.3251	+0.4528	+0.6149
	+1.0000	+0.7897	+0.9021	+0.7415	+0.3436	+0.4903	+0.0814	+0.5687	+0.0447
		+1.0000	+0.8321	+0.8874	+0.0860	+0.4352	+0.2047	+0.5077	-0.1216
			+1.0000	+0.8306	+0.1216	+0.4231	+0.0737	+0.4763	-0.0967
				+1.0000	+0.3220	+0.5661	+0.1301	+0.5983	-0.0425
					+1.0000	+0.5381	+0.0256	+0.4337	+0.4505
						+1.0000	+0.2549	+0.6420	+0.7264
							+1.0000	+0.1932	+0.2788
								+1.0000	+0.3549
									+1.0000

Figure 7. Frequency and amplitude characteristics of the ten digits with the pronunciation of "five" and "nine" exaggerated.

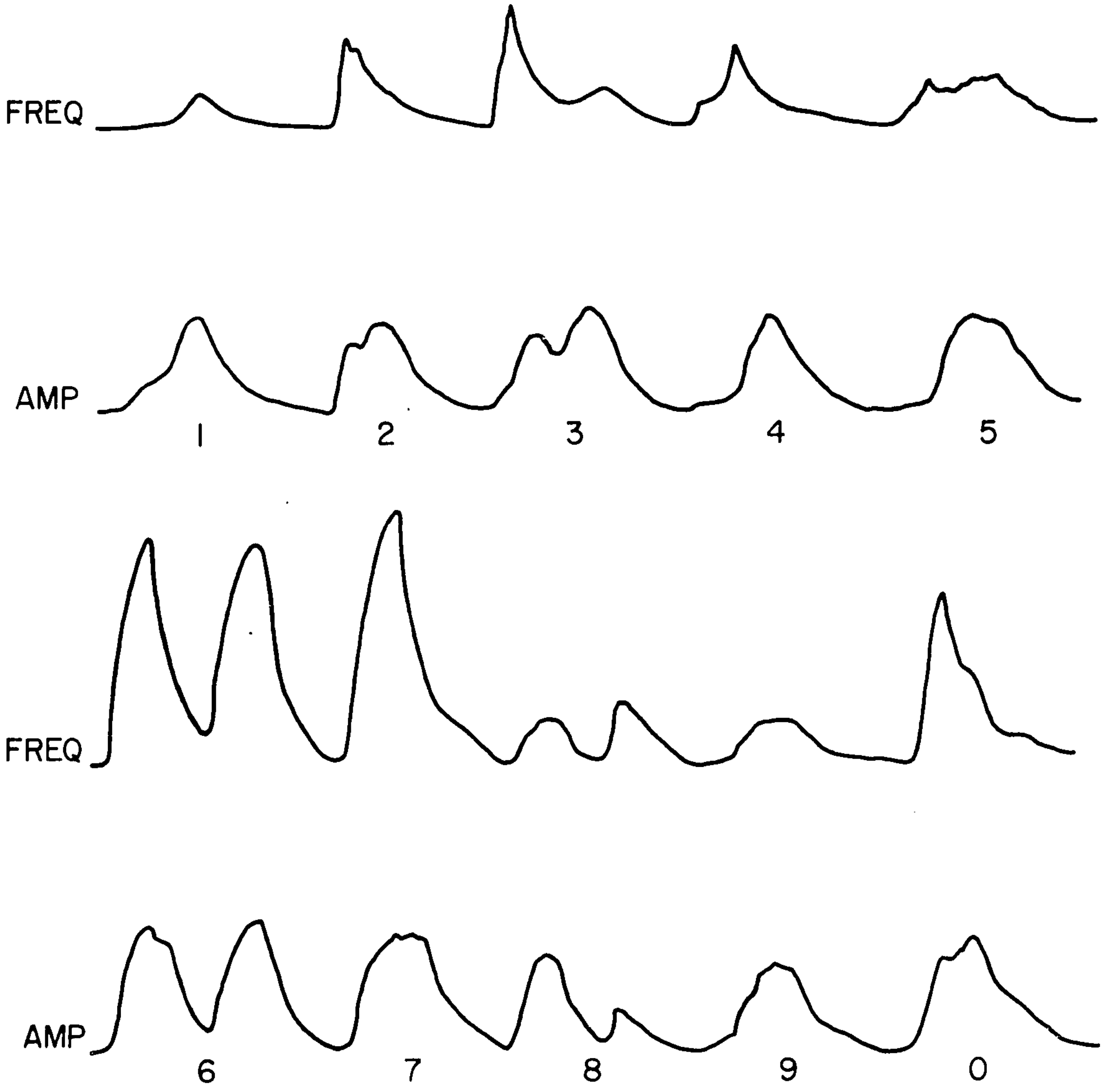

FREQ
AMP
1
2
3
4
5
FREQ
AMP
6
7
8
9
0

Table 5a. Correlation matrix of the frequency signal of ten digits with pronunciation of "five" and "nine" exaggerated.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.0314	+0.0585	+0.4755	+0.7181	-0.2461	+0.2526	+0.0290	+0.9054	+0.2500
	+1.0000	+0.8460	+0.5265	+0.2000	+0.2801	+0.8094	+0.2744	+0.0924	+0.7003
		+1.0000	+0.5264	+0.4458	+0.4907	+0.7694	+0.5884	+0.1607	+0.7734
			+1.0000	+0.6540	+0.3350	+0.8892	+0.4118	+0.6982	+0.8195
				+1.0000	+0.1667	+0.4944	+0.5805	+0.8622	+0.5466
					+1.0000	+0.3960	+0.4906	-0.1219	+0.4700
						+1.0000	+0.4127	+0.4320	+0.9501
							+1.0000	+0.2171	+0.4616
								+1.0000	+0.4147
									+1.0000

Table 5b. Correlation matrix of the first difference frequency signal of ten digits with pronunciation of "five" and "nine" exaggerated.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	-0.0327	-0.0976	-0.0975	+0.1170	-0.2416	-0.1894	-0.2372	+0.5589	-0.1077
	+1.0000	+0.5406	+0.2169	+0.1092	+0.3943	+0.5742	+0.1484	-0.0873	+0.0950
		+1.0000	+0.1683	+0.2508	+0.4551	+0.5604	+0.3341	-0.2056	+0.4973
			+1.0000	+0.1087	+0.1826	+0.7108	+0.2089	+0.3052	+0.0898
				+1.0000	+0.1622	+0.3588	+0.2789	+0.3603	+0.5029
					+1.0000	+0.5647	+0.2711	-0.2959	+0.5310
						+1.0000	+0.3711	+0.0016	+0.5854
							+1.0000	-0.2871	+0.3324
								+1.0000	-0.1090
									+1.0000

Table 6a. Correlation matrix of the first difference frequency signal for Rickey.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.2736	+0.2839	+0.2551	+0.2024	+0.6950	+0.8295	+0.6490	+0.3354	+0.8044
	+1.0000	+0.7163	+0.1698	+0.1765	+0.5840	+0.4507	+0.5415	+0.3917	+0.3103
		+1.0000	-0.1472	-0.0225	+0.4313	+0.3082	+0.3248	+0.4413	+0.2947
			+1.0000	+0.6825	+0.1839	+0.3818	+0.4633	+0.1416	+0.2016
				+1.0000	+0.2241	+0.3844	+0.2330	+0.4371	+0.2973
					+1.0000	+0.7739	+0.5804	+0.5567	+0.6740
						+1.0000	+0.7304	+0.4271	+0.9455
							+1.0000	+0.2048	+0.5771
								+1.0000	+0.4475
									+1.0000

Table 6b. Correlation matrix of the first difference frequency for Casson.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	-0.0847	+0.1294	-0.1417	-0.0935	+0.2212	+0.1841	+0.5275	+0.5568	+0.2208
	+1.0000	+0.5243	+0.8889	+0.6571	+0.4424	+0.5751	+0.0267	+0.1836	+0.6344
		+1.0000	+0.6394	+0.4921	+0.4107	+0.4531	+0.2907	+0.3071	+0.6586
			+1.0000	+0.8575	+0.5106	+0.7793	+0.1185	+0.1047	+0.7727
				+1.0000	+0.4277	+0.8732	+0.2059	+0.1615	+0.7790
					+1.0000	+0.6529	+0.1780	+0.2451	+0.4977
						+1.0000	+0.2761	+0.2302	+0.7486
							+1.0000	+0.7627	+0.2924
								+1.0000	+0.3716
									+1.0000

Table 6c. Correlation matrix of the first difference frequency signal for McConnell.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.2781	+0.2140	+0.3856	+0.6487	+0.3919	+0.6255	+0.1584	+0.3772	+0.4041
	+1.0000	+0.7558	+0.4461	+0.2724	+0.5927	+0.5725	+0.3636	+0.3643	+0.8627
		+1.0000	+0.7669	+0.4683	+0.4733	+0.4200	+0.2640	+0.4568	+0.8348
			+1.0000	+0.5297	+0.4416	+0.5758	+0.3779	+0.1961	+0.5951
				+1.0000	+0.4164	+0.6397	+0.3202	+0.6516	+0.5227
					+1.0000	+0.6274	+0.2108	+0.2562	+0.6308
						+1.0000	+0.4073	+0.3604	+0.5889
							+1.0000	+0.0371	+0.3530
								+1.0000	+0.4082
									+1.0000

Table 6d. Correlation matrix of the first difference frequency signal for Rowe.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.5506	+0.4655	+0.1325	+0.3408	+0.4871	+0.6075	+0.7606	+0.5076	+0.4607
	+1.0000	+0.1076	+0.2604	+0.2622	+0.6238	+0.5444	+0.4757	+0.5175	+0.6549
		+1.0000	+0.3753	+0.7444	+0.0870	+0.3983	+0.5754	+0.6803	+0.1939
			+1.0000	+0.6313	-0.0127	+0.2972	+0.2817	+0.1792	+0.3096
				+1.0000	+0.0685	+0.3957	+0.5398	+0.3791	+0.3541
					+1.0000	+0.7317	+0.5573	+0.5084	+0.5255
						+1.0000	+0.8127	+0.6062	+0.5488
							+1.0000	+0.6417	+0.6347
								+1.0000	+0.5321
									+1.0000

Table 6e. Correlation matrix of the first difference frequency signal for P. Anderson.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.1469	-0.2167	-0.0019	+0.2182	+0.5129	+0.5585	+0.4250	+0.7206	+0.3643
	+1.0000	+0.2511	+0.4177	+0.1201	+0.2226	+0.2127	+0.1900	-0.0143	+0.1046
		+1.0000	+0.4119	+0.5375	-0.1772	-0.1706	-0.1250	+0.2703	-0.1185
			+1.0000	+0.5014	+0.1462	+0.3670	+0.3729	-0.0886	+0.2906
				+1.0000	+0.1867	+0.3829	+0.3327	+0.3904	+0.3159
					+1.0000	+0.6799	+0.3564	+0.3263	+0.5290
						+1.0000	+0.6321	+0.3410	+0.6698
							+1.0000	+0.1105	+0.4114
								+1.0000	+0.2910
									+1.0000

Table 6f. Correlation matrix of the first difference frequency signal for Skola.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.3401	-0.1474	+0.0975	+0.3455	+0.3186	+0.2505	+0.1662	+0.6950	+0.4076
	+1.0000	+0.1678	+0.2953	+0.0176	+0.2122	+0.0913	+0.0961	+0.3667	+0.2682
		+1.0000	+0.5592	+0.5040	+0.3095	+0.4815	+0.4019	+0.2021	+0.4324
			+1.0000	+0.3475	+0.5813	+0.8888	+0.2606	+0.2932	+0.6952
				+1.0000	+0.1640	+0.3910	+0.3952	+0.4337	+0.3906
					+1.0000	+0.7073	+0.0093	+0.2960	+0.7989
						+1.0000	+0.2263	+0.2646	+0.8294
							+1.0000	+0.0648	+0.2517
								+1.0000	+0.3446
									+1.0000

Table 6g. Correlation matrix of the first difference frequency signal for Reuter.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	-0.0227	+0.3228	+0.2068	+0.3617	+0.4393	+0.5428	+0.6052	+0.4450	+0.5791
	+1.0000	+0.4530	+0.3018	+0.1939	+0.2925	+0.5690	+0.2010	+0.5491	+0.3742
		+1.0000	+0.8590	+0.3795	+0.4713	+0.7484	+0.3979	+0.6370	+0.8029
			+1.0000	+0.2701	+0.3999	+0.5610	+0.3748	+0.4146	+0.5965
				+1.0000	+0.1848	+0.4075	+0.3331	+0.5673	+0.3131
					+1.0000	+0.6465	+0.4357	+0.4753	+0.5912
						+1.0000	+0.6712	+0.6470	+0.9280
							+1.0000	+0.3168	+0.6360
								+1.0000	+0.5713
									+1.0000

Table 6h. Correlation matrix of the first difference frequency signal for Magee .

				Digits					
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.4243	+0.3124	+0.0382	+0.3161	+0.6473	+0.6893	+0.9066	+0.2221	+0.7154
	+1.0000	-0.0220	+0.0157	+0.2133	+0.4777	+0.2991	+0.3827	+0.1072	+0.7208
		+1.0000	+0.7247	+0.7060	+0.2811	+0.3359	+0.5860	+0.0962	+0.1983
			+1.0000	+0.7181	+0.4393	+0.0073	+0.3399	+0.2726	+0.2079
				+1.0000	+0.5642	+0.1477	+0.4462	+0.4561	+0.2697
					+1.0000	+0.3517	+0.5981	+0.2969	+0.8143
						+1.0000	+0.6970	+0.0719	+0.5444
							+1.0000	+0.2505	+0.6892
								+1.0000	+0.0897
									+1.0000

Table 6i. Correlation matrix of the first difference frequency signal for Anderson.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	-0.0190	+0.2355	+0.4402	+0.6111	+0.4106	+0.4639	-0.0946	+0.7257	+0.0078
	+1.0000	+0.0813	-0.0353	+0.4420	+0.2675	+0.1567	+0.0462	+0.4808	+0.6365
		+1.0000	+0.2191	+0.5157	-0.0181	+0.1853	+0.5791	+0.5160	+0.2089
			+1.0000	+0.3658	+0.2753	+0.4542	+0.2901	+0.3514	+0.0223
				+1.0000	+0.4593	+0.4887	+0.1047	+0.8045	+0.3227
					+1.0000	+0.6573	-0.1803	+0.4091	+0.4728
						+1.0000	+0.1779	+0.5672	+0.5400
							+1.0000	+0.2362	+0.2048
								+1.0000	+0.4610
									+1.0000

Table 7a. Correlation matrix of first difference frequency signal for "one" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.9004	+0.8288	+0.8599	+0.8291	+0.7267	+0.8199	+0.7081	+0.6545	+0.8115
	+1.0000	+0.7638	+0.8355	+0.7857	+0.6461	+0.7616	+0.6277	+0.6196	+0.7475
		+1.0000	+0.8497	+0.8876	+0.9092	+0.8183	+0.8071	+0.8080	+0.6418
			+1.0000	+0.8742	+0.8421	+0.8053	+0.7893	+0.7990	+0.8736
				+1.0000	+0.8962	+0.7997	+0.8851	+0.8232	+0.7311
					+1.0000	+0.7743	+0.8305	+0.8176	+0.5968
						+1.0000	+0.8148	+0.8835	+0.8032
							+1.0000	+0.9369	+0.7165
								+1.0000	+0.7374
									+1.0000

Table 7b. Correlation matrix of first difference frequency signals for "two" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.7882	+0.8049	+0.5340	+0.5340	+0.7502	+0.8557	+0.8868	+0.9400	+0.6671
	+1.0000	+0.7101	+0.7897	+0.6969	+0.8552	+0.9002	+0.9362	+0.6493	+0.7840
		+1.0000	+0.4161	+0.8285	+0.4352	+0.6115	+0.8266	+0.8581	+0.3947
			+1.0000	+0.6289	+0.6402	+0.6469	+0.6142	+0.3623	+0.4937
				+1.0000	+0.2681	+0.4349	+0.6424	+0.5712	+0.1703
					+1.0000	+0.9440	+0.8449	+0.5174	+0.9508
						+1.0000	+0.9023	+0.6694	+0.8800
							+1.0000	+0.7851	+0.8243
								+1.0000	+0.4567
									+1.0000

Table 7c. Correlation matrix of first difference frequency signal for "three" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.4859	+0.5097	+0.1078	+0.5467	+0.4649	+0.7210	+0.5003	+0.7984	+0.4109
	+1.0000	+0.6505	+0.2974	+0.0937	+0.3252	+0.0673	+0.1225	+0.7110	+0.3017
		+1.0000	+0.2748	+0.5334	+0.3725	+0.4611	+0.4751	+0.5528	+0.2307
			+1.0000	+0.1293	+0.4990	+0.1913	+0.1267	+0.3825	+0.4195
				+1.0000	+0.2136	+0.6863	+0.9307	+0.3262	+0.1996
					+1.0000	+0.3227	+0.1401	+0.6155	+0.3957
						+1.0000	+0.6951	+0.5153	+0.3285
							+1.0000	+0.3120	+0.2175
								+1.0000	+0.4107
									+1.0000

Table 7d. Correlation matrix of first difference frequency signal for "four" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.0504	+0.6563	+0.2036	-0.0400	+0.4973	+0.5184	+0.1723	+0.7131	+0.0363
	+1.0000	+0.4513	+0.5559	+0.9193	+0.2101	+0.7998	+0.8760	+0.2540	+0.5298
		+1.0000	+0.2717	+0.3744	+0.4584	+0.7596	+0.7038	+0.3671	+0.1251
			+1.0000	+0.4813	+0.5946	+0.4842	+0.3565	+0.5128	+0.4480
				+1.0000	+0.1621	+0.7117	+0.8404	+0.2485	+0.5704
					+1.0000	+0.3937	+0.1838	+0.4695	+0.3288
						+1.0000	+0.8664	+0.4447	+0.3186
							+1.0000	+0.1555	+0.3840
								+1.0000	+0.5137
									+1.0000

Table 7e. Correlation matrix of first difference frequency signal for "five" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.4349	+0.6960	+0.0464	+0.1604	+0.3931	+0.0539	+0.3317	+0.3671	+0.2581
	+1.0000	+0.8724	+0.2567	+0.2529	+0.4121	+0.5050	+0.6143	+0.7813	+0.3662
		+1.0000	+0.1692	+0.2061	+0.4232	+0.3931	+0.4883	+0.7145	+0.3040
			+1.0000	+0.5942	+0.6737	+0.7737	+0.7455	+0.6488	+0.4468
				+1.0000	+0.6356	+0.6863	+0.6176	+0.6348	+0.6543
					+1.0000	+0.6851	+0.7316	+0.6870	+0.4217
						+1.0000	+0.7615	+0.7661	+0.5371
							+1.0000	+0.8185	+0.5602
								+1.0000	+0.5586
									+1.0000

Table 7f. Correlation matrix of first difference frequency signal for "six" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.7174	+0.8629	+0.7195	+0.8296	+0.6731	+0.7031	+0.7158	+0.7216	+0.7496
	+1.0000	+0.6621	+0.7378	+0.8228	+0.8466	+0.6365	+0.8544	+0.6990	+0.5640
		+1.0000	+0.6916	+0.7205	+0.5832	+0.7120	+0.6593	+0.6707	+0.6912
			+1.0000	+0.8680	+0.6980	+0.8840	+0.8139	+0.8327	+0.5850
				+1.0000	+0.8126	+0.8059	+0.9153	+0.8952	+0.5921
					+1.0000	+0.6130	+0.8337	+0.7450	+0.3611
						+1.0000	+0.7617	+0.8879	+0.6215
							+1.0000	+0.8703	+0.4881
								+1.0000	+0.5384
									+1.0000

Table 7g. Correlation matrix of first difference frequency signal for "seven" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.5550	+0.8507	+0.6862	+0.6050	+0.6004	+0.8554	+0.8347	+0.9478	+0.9630
	+1.0000	+0.5211	+0.6208	+0.9648	+0.9451	+0.4491	+0.8338	+0.5943	+0.5367
		+1.0000	+0.5463	+0.5611	+0.5080	+0.9536	+0.6966	+0.8424	+0.9364
			+1.0000	+0.6658	+0.6623	+0.5733	+0.6605	+0.6922	+0.6213
				+1.0000	+0.9704	+0.4835	+0.8053	+0.6162	+0.5764
					+1.0000	+0.4515	+0.7946	+0.5840	+0.5402
						+1.0000	+0.6352	+0.7943	+0.9192
							+1.0000	+0.8882	+0.8033
								+1.0000	+0.9494
									+1.0000

Table 7h. Correlation matrix of first difference frequency signal for "eight" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.3697	+0.1153	+0.3911	+0.4198	+0.7255	+0.4356	+0.2748	+0.3233	+0.7250
	+1.0000	+0.1417	+0.5928	+0.6162	+0.6408	+0.6390	+0.6652	+0.5808	+0.1232
		+1.0000	+0.7322	-0.0262	+0.1578	-0.1172	+0.1284	+0.0468	+0.0386
			+1.0000	+0.3142	+0.5490	+0.2360	+0.5392	+0.1761	+0.2120
				+1.0000	+0.6201	+0.9186	+0.2177	+0.6905	+0.5580
					+1.0000	+0.6220	+0.5422	+0.4788	+0.4821
						+1.0000	+0.1897	+0.8142	+0.4499
							+1.0000	+0.0682	-0.0087
								+1.0000	+0.2021
									+1.0000

Table 7i. Correlation matrix of first difference frequency signal for "nine" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.7222	+0.5473	+0.5900	+0.5413	+0.5911	+0.7627	+0.7086	+0.7577	+0.6627
	+1.0000	+0.8677	+0.9281	+0.8173	+0.8097	+0.7278	+0.9044	+0.9054	+0.5454
		+1.0000	+0.8725	+0.7234	+0.8657	+0.7906	+0.8122	+0.7438	+0.4907
			+1.0000	+0.7689	+0.7921	+0.6804	+0.9120	+0.8373	+0.5382
				+1.0000	+0.6364	+0.6724	+0.6786	+0.7189	+0.4478
					+1.0000	+0.7509	+0.6779	+0.7383	+0.3047
						+1.0000	+0.6935	+0.6635	+0.5760
							+1.0000	+0.8044	+0.6459
								+1.0000	+0.5768
									+1.0000

Table 7j. Correlation matrix of first difference frequency signal for "zero" as spoken by ten men and women.

Schauer	Rickey	Casson	McConnell	Rowe	P. Anderson	Skola	Rueter	Magee	Anderson
+1.0000	+0.5315	+0.3782	+0.3606	+0.5277	+0.5364	+0.8034	+0.8612	+0.7504	-0.0848
	+1.0000	+0.5182	+0.5224	+0.3316	+0.4556	+0.3118	+0.5940	+0.4623	+0.0468
		+1.0000	+0.7472	+0.8142	+0.7289	+0.5905	+0.6417	+0.7373	+0.6157
			+1.0000	+0.8369	+0.6863	+0.6305	+0.7177	+0.6360	+0.7748
				+1.0000	+0.8077	+0.8735	+0.8149	+0.8273	+0.7043
					+1.0000	+0.7907	+0.7339	+0.8991	+0.3662
						+1.0000	+0.9000	+0.8961	+0.3656
							+1.0000	+0.8788	+0.3388
								+1.0000	+0.3436
									+1.0000

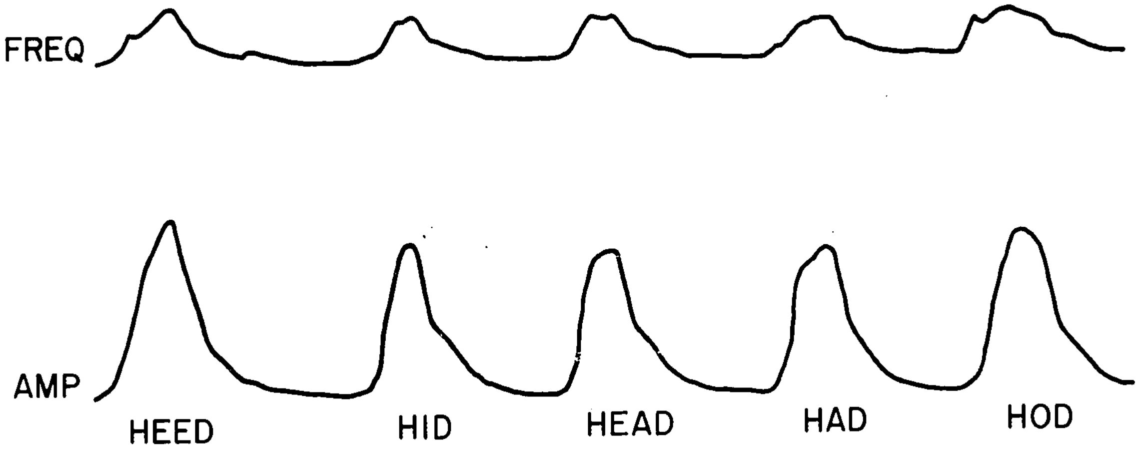
ten digits spoken by ten men and women are given. As was expected, taking the first differences accentuated the irregularities of the signals, and consequently those digits which were not well correlated for the ten speakers on the basis of the frequency signal are even more poorly correlated. On the basis of the first differences, it would appear that the individual differences of speakers are too accentuated to permit recognition based on the first difference frequency characteristics unless the frequency characteristics themselves are highly correlated.

In a vowel study program conducted at the Bell Telephone Laboratories by Peterson and Barney, (42), it was found that when observers disagreed with speakers on the classification of a vowel sound, the two classifications were nearly always in adjacent position of the vowel loop as plotted by the investigators. The test involved studying the ability of listeners to classify ten vowel sounds when these sounds appeared in words with the same initial and final consonants. In this case the consonants selected were "h" and "d". For example, results indicated that the probability of mistaking "hid" for "head", and vice versa, was higher than mistaking "heed" for "hide". The results of this experiment were duplicated using correlation studies of the derived frequency and first difference frequency signals for the same ten words. The frequency and amplitude characteristics of the ten words are given in Figure 8.

If the words were plotted around the vowel path, they would be arranged in the following order: "heed, hid, head, had, hod, hawed, who'd", with "heard, hud and hood" inside the enclosed path. The correlation matrices of the frequency, amplitude, and first difference frequency signals for the ten selected words are given in Tables 8a-8c. It is apparent that nearly all of the signals appear to be highly correlated as evidenced by the frequency and amplitude matrices. Certainly little information concerning the word is evident in these two signals on the basis of this analysis. However, the first difference frequency matrix, as shown in Table 8c, indicates that these sounds are not well correlated using the first difference signals. In fact, the highest cross correlation coefficients occurred for words such as "head" and "hid" having vowel sounds adjacent on the vowel loop or between those words inside the loop and those on the loop. No general statement was made to cover those sounds inside the loop in the original report. Essentially the same results as reported by Peterson and Barney (42) were obtained using the correlation studies on the derived speech signals.

In the investigation reported to this point, the derived frequency and amplitude signals were obtained by using the full speech spectrum extending from about 100 to 10,000 cycles per second. Figures 9, 10, and 11 show the results of using RC filters to determine just how much of the spectrum was necessary to sustain the same information content. To obtain the signals of Figure 9, a high pass RC filter was inserted at the input to the phase inverter of the audio amplifier and the value of C adjusted to obtain the various cutoff frequencies. The loading on the previous stage was not changed appreciably by the addition of this

Figure 8. Frequency and amplitude characteristics of words used in vowel study program.


FREQ
AMP
HEED
HID
HEAD
HAD
HOD

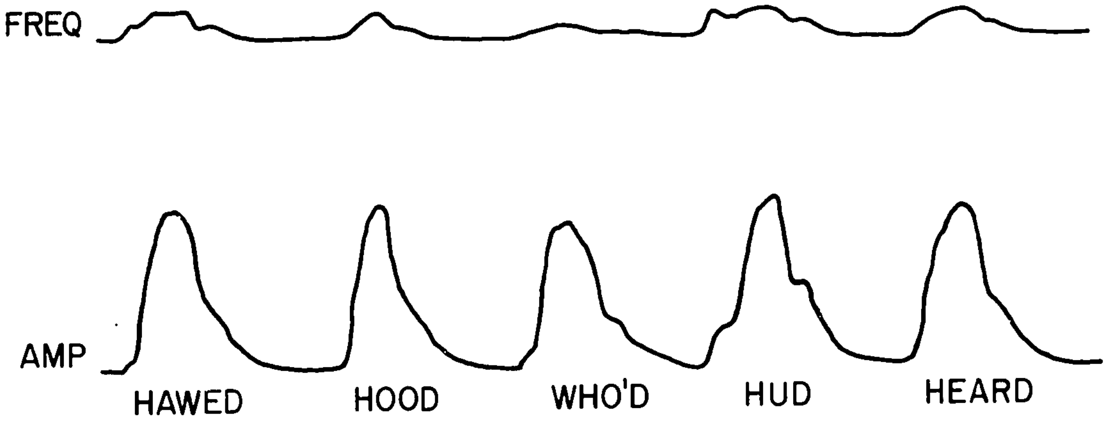

FREQ
AMP
HAWED
HOOD
WHO'D
HUD
HEARD

Table 8a. Correlation matrix of the frequency signal for the ten word vowel study.

heed	hid	head	had	hod	hawed	hood	who'd	hud	heard
+1.0000	+0.7170	+0.6055	+0.7828	+0.7604	+0.8592	+0.6067	+0.7583	+0.7600	+0.9532
	+1.0000	+0.9432	+0.9566	+0.8811	+0.9115	+0.9603	+0.8851	+0.7186	+0.8632
		+1.0000	+0.9404	+0.8974	+0.8652	+0.9700	+0.8580	+0.7325	+0.7716
			+1.0000	+0.9241	+0.9464	+0.9247	+0.8790	+0.8127	+0.9118
				+1.0000	+0.9492	+0.8682	+0.8906	+0.9265	+0.8685
					+1.0000	+0.8544	+0.9201	+0.8913	+0.9569
						+1.0000	+0.8353	+0.6588	+0.7629
							+1.0000	+0.8089	+0.8813
								+1.0000	+0.8344
									+1.0000

Table 8b. Correlation matrix of the amplitude signal for the ten word vowel study.

heed	hid	head	had	hod	hawed	hood	who'd	hud	heard
+1.0000	+0.7283	+0.6375	+0.7868	+0.9829	+0.9151	+0.5633	+0.8820	+0.9747	+0.9116
	+1.0000	+0.9676	+0.9681	+0.8114	+0.9278	+0.9493	+0.9514	+0.6290	+0.9242
		+1.0000	+0.9598	+0.7250	+0.8644	+0.9844	+0.8940	+0.5331	+0.8768
			+1.0000	+0.8494	+0.9567	+0.9281	+0.9654	+0.6926	+0.9640
				+1.0000	+0.9533	+0.6570	+0.9328	+0.9494	+0.9476
					+1.0000	+0.8189	+0.9921	+0.8481	+0.9936
						+1.0000	+0.8586	+0.4396	+0.8234
							+1.0000	+0.8051	+0.9870
								+1.0000	+0.8475
									+1.0000

Table 8c. Correlation matrix of the first difference frequency signal for the ten word vowel study.

heed	hid	head	had	hod	hawed	hood	who'd	hud	heard
+1.0000	+0.2067	+0.3004	+0.5227	+0.2399	+0.4868	+0.2401	+0.4587	+0.2950	+0.7965
	+1.0000	+0.7860	+0.5961	+0.4389	+0.5126	+0.8072	+0.6474	+0.1138	+0.4670
		+1.0000	+0.6621	+0.4480	+0.3849	+0.7471	+0.6730	+0.1716	+0.4682
			+1.0000	+0.4177	+0.4930	+0.5478	+0.6693	+0.2041	+0.7574
				+1.0000	+0.6154	+0.3909	+0.6345	+0.7108	+0.3893
					+1.0000	+0.4182	+0.7223	+0.6228	+0.6927
						+1.0000	+0.6058	-0.0091	+0.3976
							+1.0000	+0.4787	+0.6106
								+1.0000	+0.3136
									+1.0000

Figure 9. Effects of a high pass filter on frequency characteristic.

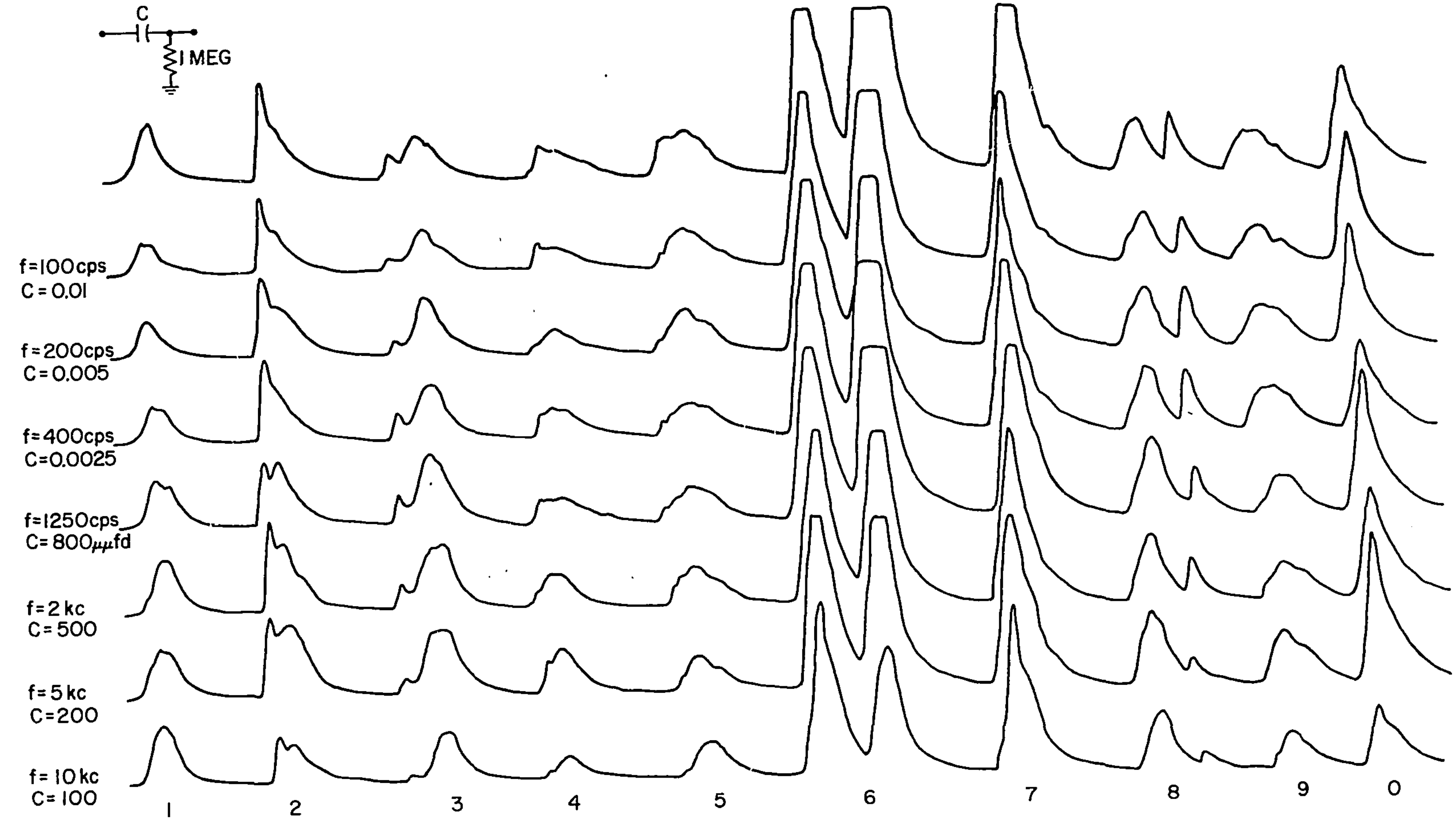

C
1 MEG
f=100cps
C=0.01
f=200cps
C=0.005
f=400cps
C=0.0025
f=1250cps
C=800μμfd
f=2 kc
C=500
f=5 kc
C=200
f=10 kc
C=100
1
2
3
4
5
6
7
8
9
0

Figure 10. Effects of single stage low pass filter on frequency characteristic.

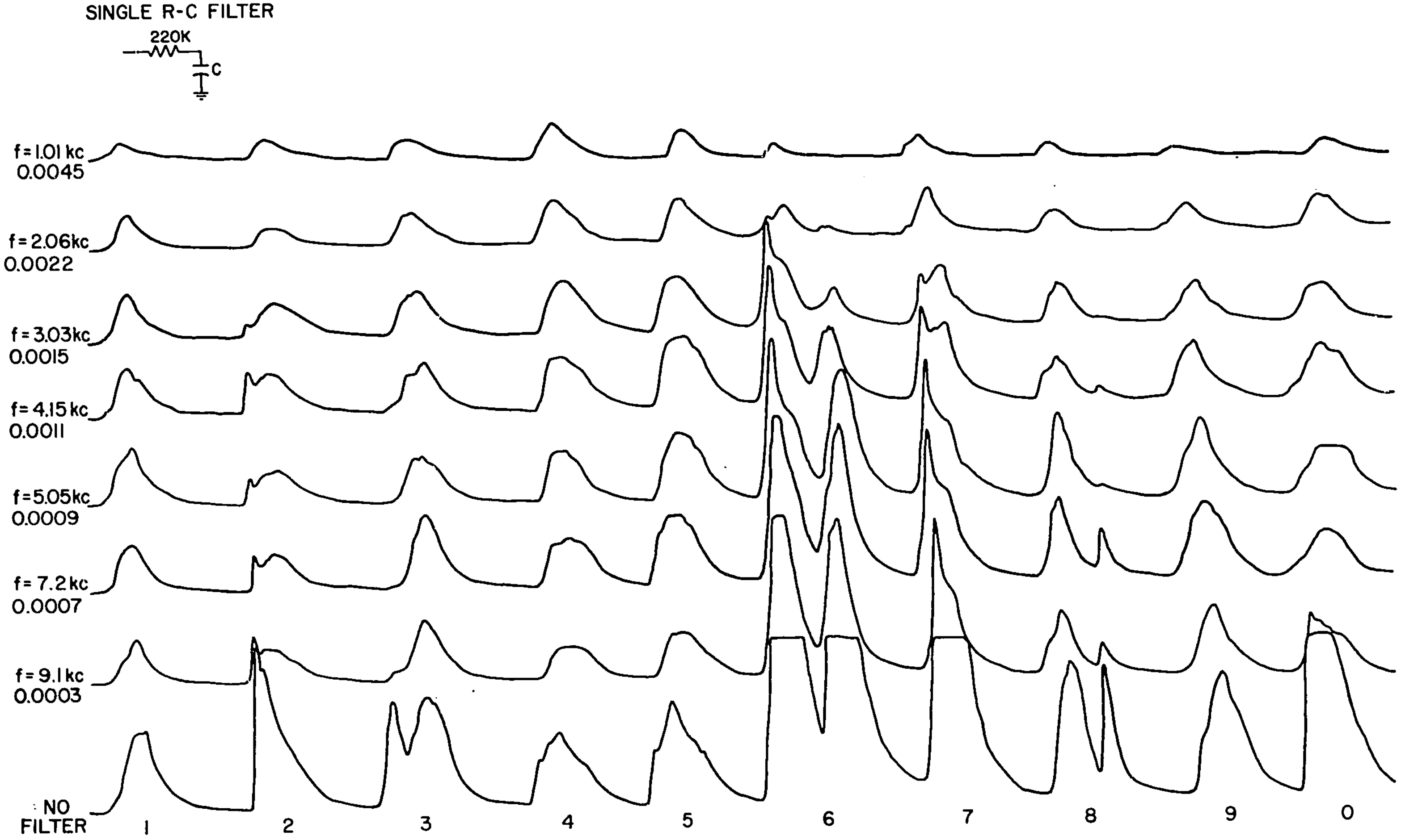
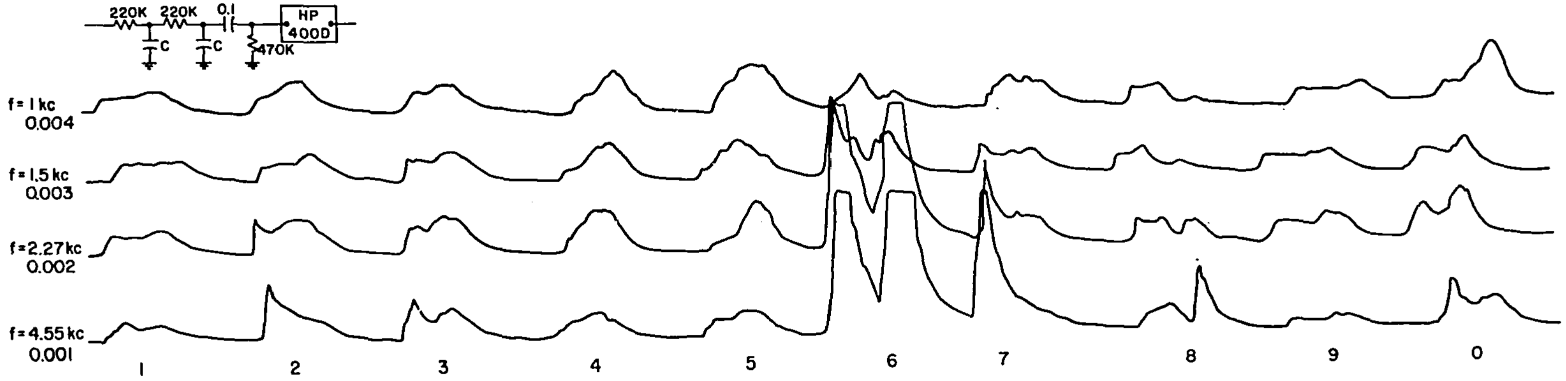

SINGLE R-C FILTER
220K
C
f = 1.01 kc
0.0045
f = 2.06 kc
0.0022
f = 3.03 kc
0.0015
f = 4.15 kc
0.0011
f = 5.05 kc
0.0009
f = 7.2 kc
0.0007
f = 9.1 kc
0.0003
NO FILTER
1
2
3
4
5
6
7
8
9
0

Figure 11. Effects of two stage low pass filter with additional amplification on frequency characteristic.


220K
220K
0.1
HP
400D
C
C
470K
f = 1 kc
0.004
f = 1.5 kc
0.003
f = 2.27 kc
0.002
f = 4.55 kc
0.001
1
2
3
4
5
6
7
8
9
0

filter network. It is clear that only when the nominal cutoff frequency is at the very upper end of the speech spectrum, are the dominant features of the frequency signal less distinct and the character of the signal lost.

The effects of a low pass RC filter, placed at the input of the phase inverter of the audio amplifier, are graphically shown in Figure 10. Values of R and C were selected to minimize the loading effects of the filter on the preceding amplifier stage. From these plots it is evident that the high frequency end of the spectrum is very important in forming the frequency characteristics of the digits. In Figure 11 additional frequency characteristics are shown. In this case a two stage low pass filter is used to give an attenuation of 40 db per decade. Because of the insertion loss, the amplifier stage of a Hewlett-Packard Vacuum tube voltmeter, model 400-D, was inserted ahead of the phase inverter of the audio amplifier. These plots also indicate that the high frequency end of the speech spectrum contains a large part of the information structure.

The simple RC filter stages are not the most effective filters that might be used to determine the spectral range necessary for reasonably distinct output signals. The attenuation per stage of the simple filter is only 20 db per decade. Consequently there is no assurance that the lower frequencies which contain the highest energy density will not contribute to the output signal even though they may be well below the cutoff frequency. The insertion loss of the filter is a second problem, particularly in the case of the low pass filter.

When the two-stage RC low pass filter was used, an additional amplifier section was necessary to assure that the input signals in the pass band were large enough to switch the cores.

A commercial filter, a Krohn-hite Instrument Company, ultra-low frequency, band pass filter, model 330-A, was used to overcome the attenuation-loss problem. The one drawback to this filter was the upper frequency limit of 2000 cycles per second. This filter has an attenuation of 70 db outside the selected band pass, and zero insertion loss. Characteristics obtained with this filter inserted in the input of the phase inverter of the audio amplifier are shown in Figure 12. For all runs the low end of the pass band was set at 0.02 cycles per second with the high end as indicated in the drawing. It is apparent that with the full pass band of the filter the distinctive characteristics of the frequency signals just begin to emerge from the broad base of the signals. It would appear from Figure 12 that insufficient information is present in the pass band from 0.02 - 2000 cycles per second to recognize individual signals. This is verified by the correlation studies of the frequency and amplitude signals shown in Figure 13. It is evident from the correlation matrix in Table 9a, that only the signals of "six" and "eight" can be separated from the entire group with any reasonable success. The rest of the signals are all essentially highly correlated. As shown in Table 9b, the first difference signals have lower cross correlation coefficients, although it would be difficult to distinguish between a "three", "five", or "nine".

The significance of frequency changes in the speech signal which

Figure 12. Effects of band pass filter on frequency characteristics.

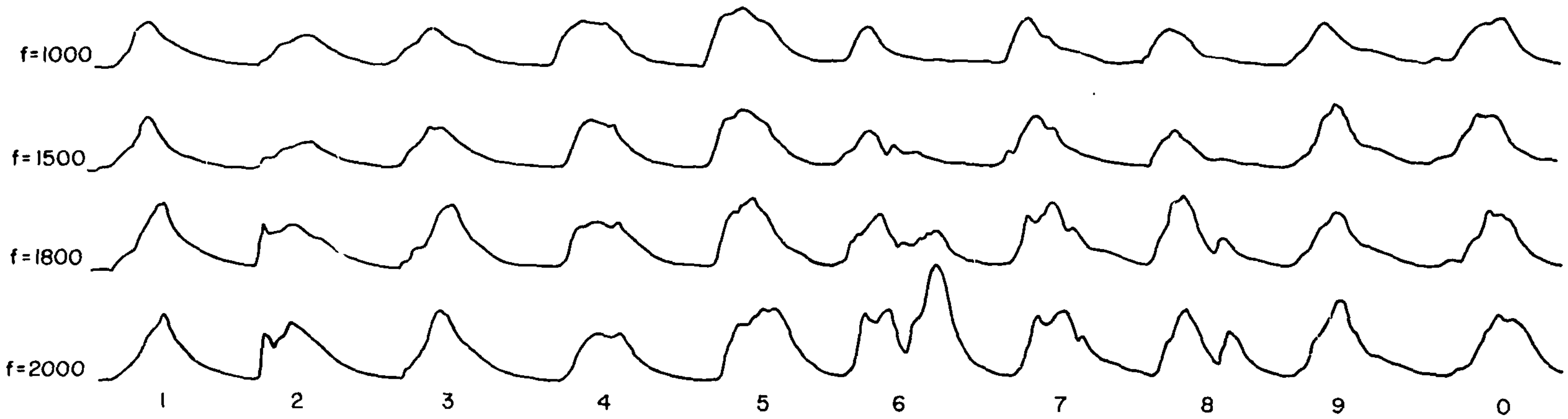

f=1000
f=1500
f=1800
f=2000
1
2
3
4
5
6
7
8
9
0

Figure 13. Frequency and amplitude characteristics of the ten digits using a pass band of 0.02 - 2000 cycles per second.

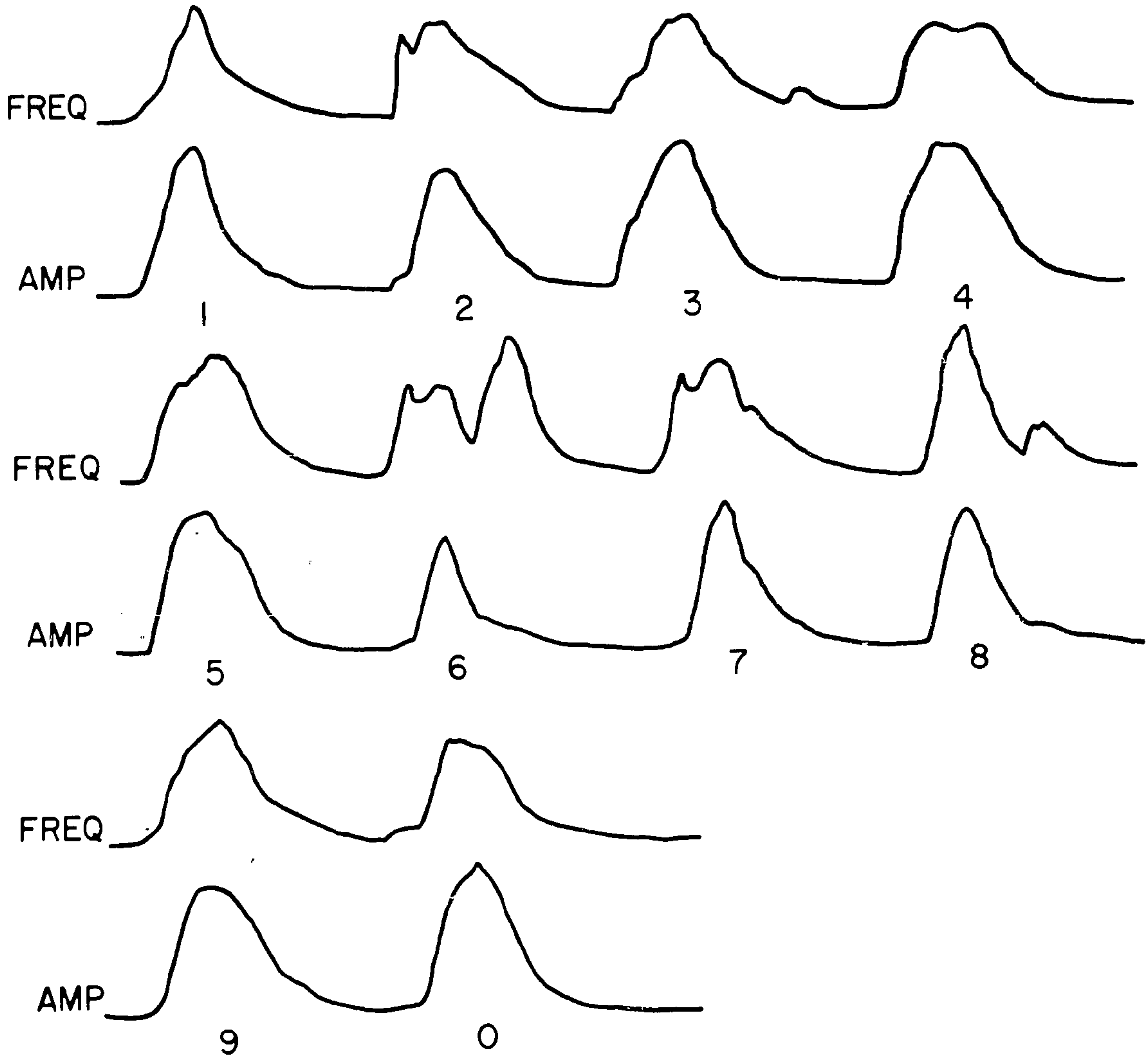

FREQ
AMP
1
2
3
4
FREQ
AMP
5
6
7
8
FREQ
AMP
9
0

Table 9a. Correlation matrix of frequency signals of the ten digits using a 0.02 - 2,000 cycles per second band pass filter.

				Digits					
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.7513	+0.9501	+0.7912	+0.8997	+0.2937	+0.8981	+0.5295	+0.9221	+0.9158
	+1.0000	+0.7519	+0.9298	+0.8614	+0.2836	+0.8839	+0.8588	+0.6964	+0.6397
		+1.0000	+0.8677	+0.9591	+0.2661	+0.8896	+0.4703	+0.9780	+0.9295
			+1.0000	+0.9627	+0.2517	+0.9237	+0.7354	+0.8401	+0.7432
				+1.0000	+0.2453	+0.9456	+0.6219	+0.9464	+0.8512
					+1.0000	+0.3180	+0.4067	+0.2332	+0.3212
						+1.0000	+0.7453	+0.8563	+0.7865
							+1.0000	+0.4404	+0.3711
								+1.0000	+0.9203
									+1.0000

Table 9b. Correlation matrix of first difference frequency signals of the ten digits using a 0.02 - 2,000 cycles per second band pass filter.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.2937	+0.7185	+0.2870	+0.5162	+0.2456	+0.5352	+0.2267	+0.5885	+0.7207
	+1.0000	+0.3233	+0.5965	+0.5496	+0.2884	+0.1291	+0.4069	+0.2876	+0.3567
		+1.0000	+0.5941	+0.7250	+0.1492	+0.5400	+0.2266	+0.7656	+0.7368
			+1.0000	+0.8238	+0.2298	+0.5382	+0.5182	+0.6012	+0.4288
				+1.0000	+0.1462	+0.6474	+0.3807	+0.7824	+0.5415
					+1.0000	+0.3615	+0.4474	-0.1248	+0.0092
						+1.0000	+0.4237	+0.5045	+0.2744
							+1.0000	+0.3154	+0.0736
								+1.0000	+0.6777
									+1.0000

Figure 14. Frequency signals for the ten digits with clipping on the peaks.

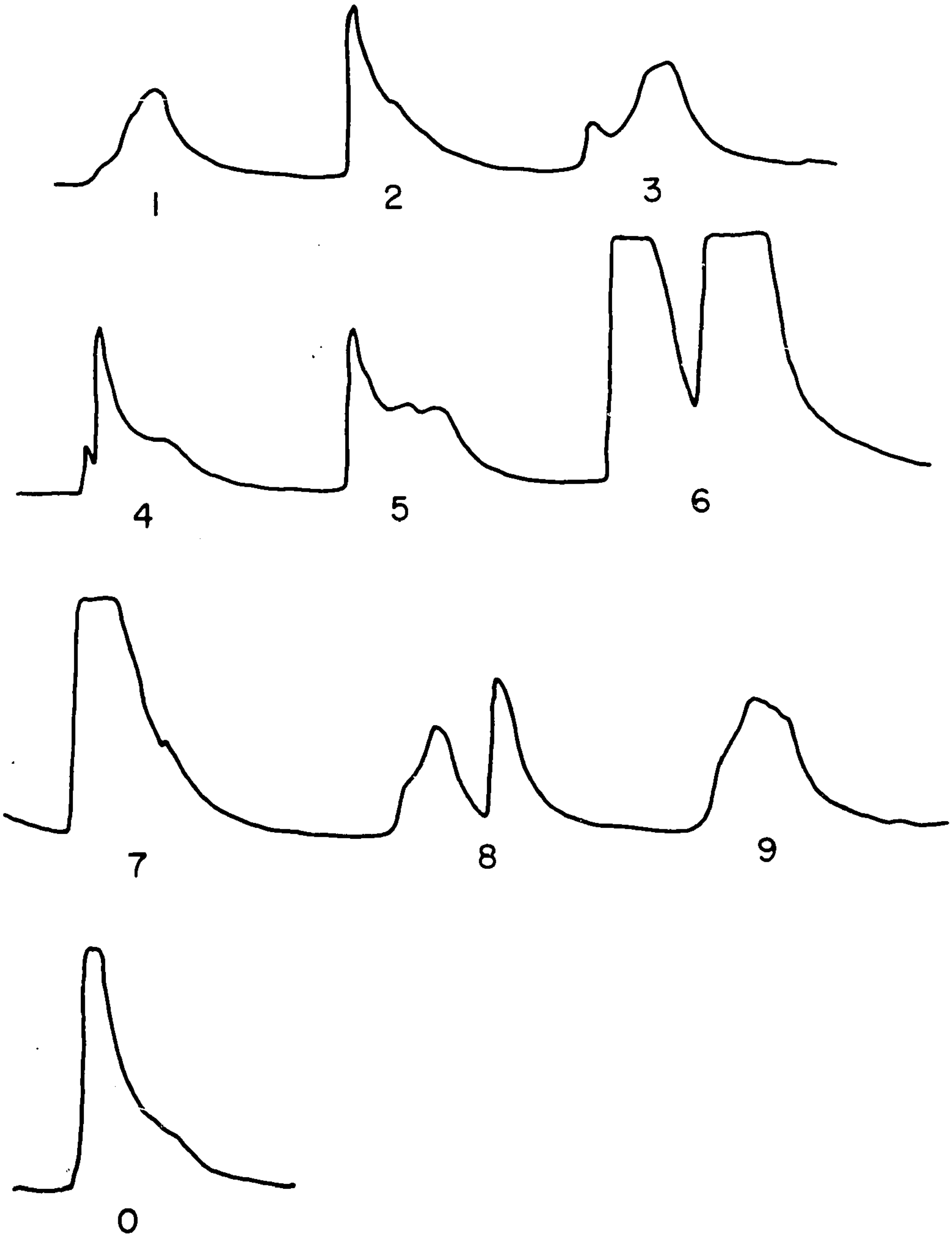

1
2
3
4
5
6
7
8
9
0

Table 10a. Correlation matrix of clipped frequency signals for the ten digits.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.3551	+0.5199	+0.6147	+0.5915	+0.2237	+0.7335	+0.5084	+0.9147	+0.6507
	+1.0000	+0.0176	+0.7011	+0.8974	+0.1116	+0.8011	+0.0360	+0.1953	+0.6347
		+1.0000	+0.3542	+0.4166	+0.2516	+0.3123	+0.4865	+0.7975	+0.3221
			+1.0000	+0.7486	+0.3566	+0.9023	+0.2814	+0.5346	+0.9655
				+1.0000	+0.1765	+0.8585	+0.2175	+0.5498	+0.6738
					+1.0000	+0.4073	+0.6634	+0.2111	+0.4554
						+1.0000	+0.3517	+0.5947	+0.9215
							+1.0000	+0.5041	+0.3459
								+1.0000	+0.5469
									+1.0000

Table 10b. Correlation matrix of first difference clipped frequency signals for the ten digits.

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.0580	-0.1189	-0.0729	+0.1180	+0.1640	+0.3331	+0.2566	+0.6736	+0.0382
	+1.0000	-0.0891	+0.1072	+0.8864	+0.1021	+0.4540	-0.0301	-0.0625	-0.0755
		+1.0000	+0.4648	-0.0005	+0.0416	+0.0187	-0.0960	+0.4204	+0.4404
			+1.0000	+0.1687	+0.1300	+0.1705	+0.0932	+0.1593	+0.7827
				+1.0000	+0.0893	+0.5499	-0.0792	+0.0555	-0.0821
					+1.0000	+0.6622	+0.1736	-0.0838	+0.4234
						+1.0000	+0.0729	+0.1460	+0.4172
							+1.0000	-0.0691	+0.1330
								+1.0000	+0.2208
									+1.0000

The significance of frequency changes in the speech signal which are large compared to the small changes associated with the medium frequency range of the speech spectrum was also studied. In this case the recorder was adjusted so that clipping occurred on the large peaks of the frequency signal. Figure 14 shows the frequency signals of the ten digits recorded in this manner. Correlation studies were conducted on the derived characteristics, and the results are given in Tables 10a and 10b. In Table 10a, it is apparent that four basic groups exist. Only the digits "six" and "eight" are clearly distinguishable. The matrix of Table 10b indicates that even the first differences of "two" and "five" are highly correlated. Clipping the signals apparently adds little to or detracts from the recognition problem.

The use of "oh" instead of "zero" was studied to determine its effect on the problem of classification of the ten frequency signals. Figure 15 contains the characteristics analyzed for the problem. The frequency signals were clipped, and "oh" was used in place of "zero" as indicated. Correlation studies were conducted on the resulting signals. The frequency signal for "oh" was highly correlated with all signals but "six". Even the first differences of "oh", "four", and "seven" are highly correlated. Therefore the use of "oh" instead of "zero" only compounded the recognition problem.

Figure 15. Frequency characteristics with clipping of the ten digits "one" through "oh".

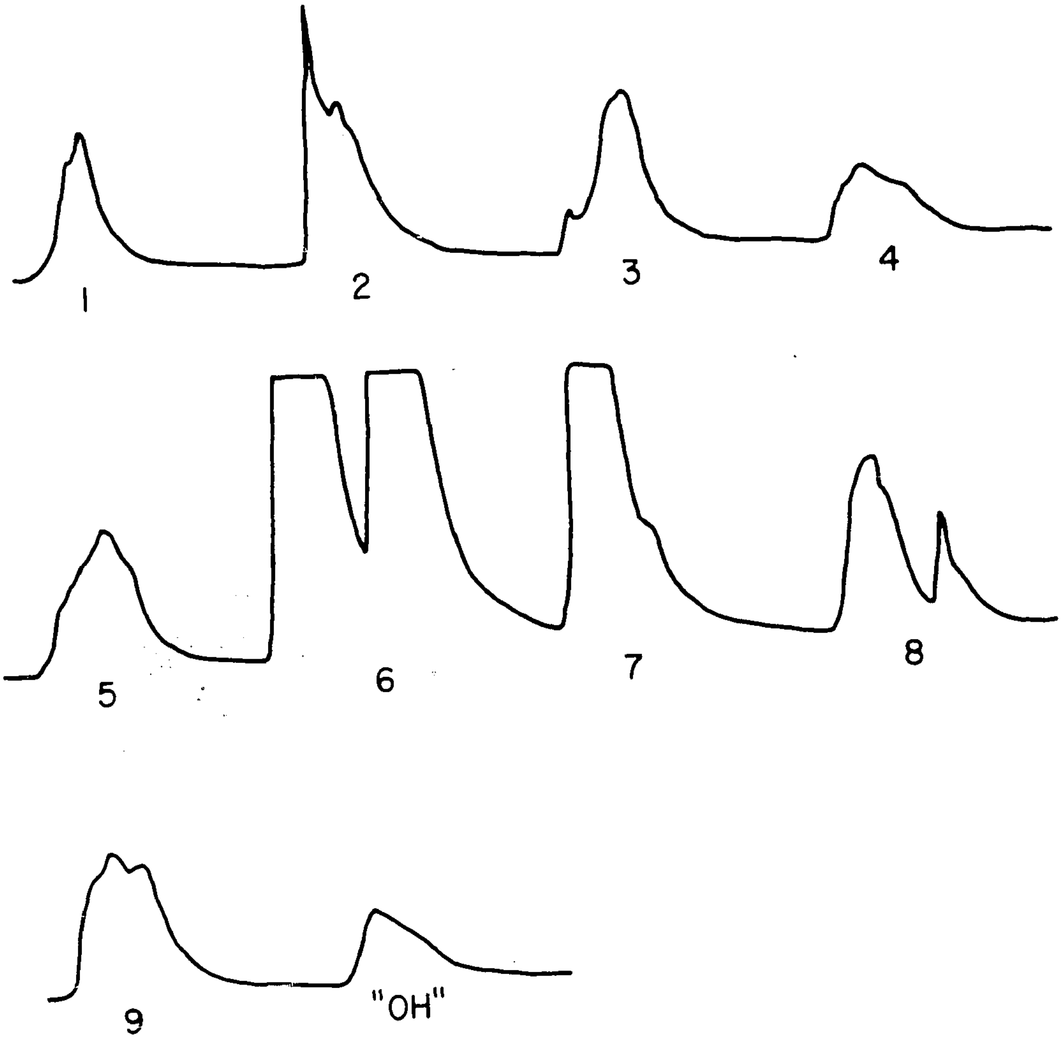

1
2
3
4
5
6
7
8
9
"OH"

Table 11a. Correlation matrix for clipped frequency signals for the ten digits with "zero" replaced by "oh".

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.6074	+0.8864	+0.9002	+0.7630	+0.3417	+0.9042	+0.8736	+0.8553	+0.9103
	+1.0000	+0.5411	+0.7024	+0.4185	+0.3864	+0.7668	+0.6525	+0.4776	+0.7426
		+1.0000	+0.8773	+0.9078	+0.2582	+0.8195	+0.7400	+0.9044	+0.8277
			+1.0000	+0.8861	+0.3191	+0.9628	+0.8776	+0.9393	+0.9820
				+1.0000	+0.1826	+0.7757	+0.6819	+0.9710	+0.8103
					+1.0000	+0.3950	+0.5659	+0.2225	+0.3600
						+1.0000	+0.9094	+0.8566	+0.9849
							+1.0000	+0.7808	+0.8987
								+1.0000	+0.8875
									+1.0000

Table 11b. Correlation matrix of first difference clipped frequency signals for the ten digits with "zero" replaced by "oh".

Digits									
1	2	3	4	5	6	7	8	9	0
+1.0000	+0.0317	+0.5305	+0.5175	+0.3630	+0.1126	+0.4345	+0.5331	+0.6300	+0.5544
	+1.0000	+0.2375	+0.1048	+0.0185	+0.6773	+0.0636	+0.0463	-0.1305	+0.1474
		+1.0000	+0.2973	+0.5601	+0.2773	+0.2070	+0.2290	+0.3352	+0.2031
			+1.0000	+0.6520	+0.1892	+0.6727	+0.5999	+0.6393	+0.8492
				+1.0000	-0.0077	+0.2150	+0.2150	+0.6900	+0.6091
					+1.0000	+0.4940	+0.4940	-0.0853	+0.2716
						+1.0000	+0.5727	+0.6762	+0.8397
							+1.0000	+0.3779	+0.6304
								+1.0000	+0.7451
									+1.0000

RESULTS

It has been shown that a pair of signals can be derived from speech which are indicative of the frequency and amplitude components varying at the rate of 15 cycles per second or less. These signals were obtained by using a rectifier and a filter for amplitude, and a core switch, rectifier and filter for frequency. The recorder was only necessary in this investigation because suitable analog to digital conversion equipment was not available in the laboratory.

The amplitude signal, analyzed by using correlation techniques, was found to contain essentially no information concerning identification of the sound from which it was derived. However, the signal would be very useful in an automated system to indicate when an input had been presented to the system. The sampling period could also be controlled in such a system by the duration of the amplitude signal.

It has been shown that the frequency signal does have some information structure associated with it. A correlation study of a very simple vocabulary consisting of the ten digits revealed normally four classifications among these ten signals, indicating insufficient information structure in the frequency signal to uniquely identify any arbitrary input.

However, it was found that the normalized frequency signal for a given digit was essentially independent of pitch and consequently insensitive to signal source with respect to the pitch variable. The results of correlation studies of ten speakers with the ten digit vocabulary have shown that it is impossible to distinguish between male and female speakers on the basis of the recorded signals alone. The variations in the individual frequency signals, when they are large enough to be important, are caused primarily by the enunciation differences of the speakers. It has been shown in the reported correlation studies that enunciation differences are most likely to occur for the fricative sounds. Because the energy is distributed over a wide range of frequencies for a fricative sound, the manner in which the sound is uttered may cause considerable differences in the frequencies present in the sound. It is apparent from the signals themselves that peaks in the characteristics, particularly those resulting from initial or final fricative consonants, may be smoothed or sometimes even missing depending upon the enunciation of the speaker. Therefore, training of individuals could practically eliminate this source of signal variation.

Correlation studies of the first differences of the frequency signals have shown that generally sufficient information is present in the difference signals to uniquely identify the individual digits of the vocabulary. It was also shown that the first difference signals for the same digit for the ten speakers were less highly correlated than the frequency signals themselves. This would be anticipated since the difference signal should accentuate any individual characteristics in the frequency signals. However, it was shown that the first difference frequency signals could be used to essentially duplicate the results of a vowel study test conducted and reported by members of the Bell Telephone Laboratories who used a completely different approach.

An investigation of the range of input frequencies necessary to retain a useable information structure in the signals revealed that the high end of the normal speech spectrum appeared to contribute most to the information structure. As the high frequency components of the speech were suppressed, the derived signals lost their distinctiveness and information content.

CONCLUSIONS

It has been shown that the amplitude and frequency signals derived to characterize speech do have sufficient information structure to be useful. It would appear that this technique of speech characterization would be valuable in areas of speech correction and therapy. A visual display is available immediately to show the effects of the enunciation of the speaker. Such a display could perhaps make evaluation of the speech problem and progress towards its solution much easier.

As indicated earlier the problems of speech transmission and recognition have been studied for many years. The method of characterization presented here could prove valuable in such areas. From an information theory viewpoint the transmission of normal speech is very wasteful of bandwidth. A bandwidth of about 7,000 cycles per second is a reasonable figure to completely identify unmodified speech. This bandwidth can be reduced only at the expense of the information concerning the individual source.

The transmission of the derived signal is one method which might be used to reduce this bandwidth requirement. The signal could be transmitted with a bandwidth of 20 cycles per second. This is a reduction in bandwidth by a factor of 350. The reduction in bandwidth alone would suggest further investigations into the problems of regenerating the sound from the transmitted derived signals. It would appear that it is possible to obtain a completely automated system, the output of which would be a continuous signal or a sampled digital representation of the derived frequency signal for any given input.

The problem of speech recognition is closely akin to that of speech transmission. Usually, however, it is desirable to automate the recognition or identification process as well as the transmission process. This implies that sufficient information structure must be contained in the analyzed signals to make recognition possible and that the form of the structure in the signals is known. Both of these conditions pose problems that must be investigated before a recognition system can be conceived. From the studies reported here a system using the frequency and first difference frequency signals would appear feasible. The frequency signal itself could be used to classify the unknown into one of several group patterns which have been indicated by these studies. The identification within the group would be performed on the basis of the first difference signal. Some correlation technique which compared the unknown input with the stored representations of the vocabulary would be used to perform both of these functions. Properly selected representations would make the problem of reliable recognition easier to solve. Further studies to determine what representations are most effective are necessary. Training of the speakers, themselves, might also be a partial answer to reliable recognition. The rate at which the speech signals need be sampled is so low that the correlation could be done during the sampling period so that an output would be available almost immediately. The speed of the system should be such that normal speech rates are possible. The ability to perform automatic speech recognition reliably has implications in the areas of digital data systems and automatic control systems as well as information transmission. Additional efforts on the recognition problem, possibly extending the ideas presented herein, appear to be warranted.

ACKNOWLEDGMENTS

The writer wishes to express his appreciation to Dr. V. W. Bolie for his interest and advice, and to Dr. R. M. Stewart, Jr. for many helpful suggestions.

BIBLIOGRAPHY

1. Benson, R. W. and Hirsh, Ira J. Some variables in audio spectrometry. J. Acoustical Soc. Am. 25:499-505, 1953.

2. Biddulph, R. Short-term autocorrelation analysis and correlatograms of spoken digits. J. Acoustical Soc. Am. 26:539-541, 1954.

3. Bogert, Bruce P. The VOBANC - a two-to-one speech bandwidth reduction system. J. Acoustical Soc. Am. 28:399-404, 1956.

4. Chang, Sze-Hou. Two schemes of speech compression system. J. Acoustical Soc. Am. 28:565-572, 1956.

5. Chao, Yuen Ren. Linguistic prerequisite for a speech writer. J. Acoustical Soc. Am. 28:1039-1045, 1956.

6. Cooper, Franklin S., Delattre, Pierre C., Liberman, Alvin M., Borst, John M., and Gerstman, Louis V. Some experiments on the perception of synthetic speech sounds. J. Acoustical Soc. Am. 24:597-606, 1952.

7. Davenport, Wilbur B., Jr. An experimental study of speechwave probability distribution. J. Acoustical Soc. Am. 24:390-399, 1952.

8. David, Edward E., Jr. Ears for computers. Scientific American 192:92-97, 1955.

9. __________. Signal theory in speech transmission. Inst. Radio Eng. Trans. Professional Group on Circuit Theory. 3:232-244, 1956.

10. Davis, K. H., Biddulph, R. and Balashek, S. Automatic recognition of spoken digits. J. Acoustical Soc. Am. 24:637-642, 1952.

11. Dudley, Homer. Fundamentals of speech synthesis. J. Audio Eng. Soc. 3:177-185, 1955.

12. __________. Phonetic pattern recognition vocoder for narrow band speech transmission. J. Acoustical Soc. Am. 30:733-739, 1958.

13. __________ and Balashek, S. Automatic recognition of phonetic patterns in speech. J. Acoustical Soc. Am. 30:721-732, 1958.

14. Fairbanks, Grant. Test of phonemic differentiation. The rhyme test. J. Acoustical Soc. Am. 30:716-720, 1958.

15. Flanagan, James L. Automatic extraction of formant frequencies from continuous speech. J. Acoustical Soc. Am. 28:110-118, 1956.

16. __________. Band width and channel capacity necessary to transmit the formant information of speech. J. Acoustical Soc. Am. 28:592-596, 1956.

17. __________. Difference limen for formant amplitude. J. Speech and Hearing Disorders. 22:205-212, 1957.

18. __________. Difference limen for the intensity of a vowel sound. J. Acoustical Soc. Am. 27:1223-1225, 1955.

19. __________. A difference limen for vowel formant frequency. J. Acoustical Soc. Am. 27:613-617, 1955.

20. __________ and House, Arthur S. Development and testing of a formant-coding speech compression system. J. Acoustical Soc. Am. 28:1040-1050, 1956.

21. Fletcher, Harvey and Galt, Rogers H. The perception of speech and its relation to telephony. J. Acoustical Soc. Am. 22:89-151, 1950.

22. Forgie, Carma, Graves, M. L. and Frick, F. C. Automatic recognition of spoken digits. (Abstract) J. Acoustical Soc. Am. 30:669-, 1958.

23. Forgie, James W. and Forgie, Carma D. Results obtained from a vowel recognition computer program. J. Acoustical Soc. Am.

24. __________, __________, and Hughes, George W. A real-time speech input system for a digital computer. (Abstract). J. Acoustical Soc. Am. 30:698-699, 1958.

25. Fry, D. B. and Denes, P. Mechanical speech recognition. In Jackson, Willis, ed. Communication theory, pp. 426-436. New York, N. Y. Academic Press, Inc. 1953.

26. __________ and __________. On presenting the output of a mechanical speech recognizer. J. Acoustical Soc. Am. 29:364-367, 1957.

27. Green, David M., Birdsall, Theodore G., and Tanner, Wilson P., Jr. Signal detection as a function of signal intensity and duration. J. Acoustical Soc. Am. 29:523-531, 1957.

28. Halle, M., Hughes, G. W., and Radley, J.-P.A. Acoustic properties of stop consonants. J. Acoustical Soc. Am. 29:107-116, 1957.

29. Harris, Cyril M. A study of the building blocks of speech. J. Acoustical Soc. Am. 25:962-969, 1953.

30. Harris, Katherine S., Hoffman, Howard S., Liberman, Alvin M., and Delattre, Pierre C. Perception of the voiced stop consonants. J. Acoustical Soc. Am. 30:122-126, 1958.

31. Hasse, K. H. and Vilbig, F. Errors in spectrum analysis by a set of narrow-band selecting filters. U. S. Atomic Energy Commission Report TR-56-121 (Air Force, Washington, D. C.), November, 1956.

32. Hughes, George W. and Halle, Morris. Spectral properties of fricative consonants. J. Acoustical Soc. Am. 28:303-310, 1956.

33. Kork, Winston K. The problem of selective voice control. J. Acoustical Soc. Am. 24:625-628, 1952.

34. __________ and Miller, R. L. Dynamic spectrograms of speech. J. Acoustical Soc. Am. 24:783-784, 1952.

35. Liberman, Alvin M., Delattre, Pierre C., Gerstmann, Louis J., and Cooper, Franklin S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Experimental Psych. 52:127-137, 1956.

36. Licklider, J. C. R. The intelligibility of amplitude-dichotomized time-quantized speech waves. J. Acoustical Soc. Am. 22:820-823, 1950.

37. __________. On the process of speech perception. J. Acoustical Soc. Am. 24:590-594, 1952.

38. Miller, George A. and Licklider, J. C. R. The intelligibility of interrupted speech. J. Acoustical Soc. Am. 22:167-173, 1950.

39. Miller, R. L. Nature of the vocal cord wave. J. Acoustical Soc. Am. 31:667-677, 1959.

40. Nakata, Kazuo. Synthesis and perception of nasal consonants. J. Acoustical Soc. Am. 31:661-666, 1959.

41. Peterson, Gordon E. The information-bearing elements of speech. J. Acoustical Soc. Am. 24:629-637, 1952.

42. ________ and Barney, Harold L. Control methods used in a study of the vowels. J. Acoustical Soc. Am. 24:175-184, 1952.

43. Potter, Ralph K., Kopp, George A., and Green, Harriet C. Visible speech. New York, N. Y. D. Van Nostrand Co., Inc. 1947.

44. Olson, Harry F.,and Belar, Herbert. Phonetic typewriter. J. Acoustical Soc. Am. 28:1032-1035, 1956.

45. Rosen, George. Dynamic analog speech synthesizer. J. Acoustical Soc. Am. 30:201-209, 1958.

46. Rudmose, H. W., Clark, K. C., Carlson, F. D., Eisenstein, J. C., and Walker, R. A. Voice measurements with an audio spectrometer. J. Acoustical Soc. Am. 20:503-512, 1948.

47. Sharf, Donald J. Intelligibility of reiterated speech. J. Acoustical Soc. Am. 31:423-427, 1959.

48. Tanner, Wilson P., Jr. Theory of recognition. J. Acoustical Soc. Am. 28:882-888, 1956.

49. Vilbig, F. Method for measuring fine details of differences in the structures of various speech sound spectrums. U. S. Atomic Energy Commission Report TR-56-122 (Air Force, Washington, D.C.). November, 1956.

50. ________ and Hasse, Kurt H. Some systems for speech-band compression. J. Acoustical Soc. Am. 28:573-577, 1956.

51. Wiren, Jacob and Stubbs, Harold L. Electronic binary selection system for phoneme classification. J. Acoustical Soc. Am. 28:1036-1038, 1956.