SOME ANALYSES OF THE SPEECH OF HEARING-IMPAIRED SPEAKERS UTILIZING DIGITAL SIGNAL PROCESSING TECHNIQUES

by

DEBRA JANE BRIERY

B.S., Kansas State University, 1973

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

Harry R. Rainbolt

Major Professor

ACKNOWLEDGEMENTS

I thank Dr. Harry Rainbolt, my advisor for this thesis, who provided the needed advise and support that allowed me to learn maximally from this experience.

Thank you to Dr. Nasir Ahmed for the use of the digital computer programs and data, and the technical assistance provided.

Appreciation is extended to the other members of my graduate committee, Dr. Norma Bunton and Dr. Bruce C. Flanagen.

I also extend my appreciation to the three hearing-impaired speakers, who participated in this study.

TABLE OF CONTENTS

Chapter		page
I	INTRODUCTION	1
	Characteristics of Deaf Speech	2
	Application of Acoustic Analysis Data	4
	Recent Developments in Acoustic Analysis of Speech .	9
	Purpose of This Study	10
II	PROCEDURES	11
	Speakers	11
	Recording Procedures	11
	Digital Spectrograms	12
	Determination of Fundamental Frequency	13
III	RESULTS AND DISCUSSION	14
IV	SUMMARY AND CONCLUSIONS	17
	REFERENCES	19
	APPENDIX A: Pure-tone air conduction audiogram for speaker 1	22
	APPENDIX B: Pure-tone air conduction audiogram for speaker 2	23
	APPENDIX C: Pure-tone air conduction audiogram for speaker 3	24
	TABLE 1: The character combinations for the grey levels utilized in the computer generated spectrograms	12
	FIGURE 1: The cepstrum computer program output for determination of voicing in a block of speech	13
	FIGURE 2: Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 1.	25
	FIGURE 3: Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 2.	26
	FIGURE 4: Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 3.	27
	FIGURE 5: Spectrogram of the sentence "Noon is the sleepy time of day." spoken by a normal hearing speaker	28

TABLE OF CONTENTS (Cont'd.)

Chapter		page
FIGURE 6:	Average fundamental frequency histogram obtained from speaker 1	29
FIGURE 7:	Average fundamental frequency histogram obtained from speaker 2	30
FIGURE 8:	Average fundamental frequency histogram obtained from speaker 3	31

The production of intelligible and natural-sounding speech requires a complex interaction of acoustic, physiological and perceptual mechanisms (Levitt, 1971). Hearing impairments cause deficiencies in this speech communication ability that range from minor to quite severe. The degree of the speech deficiency depends primarily on the severity of impairment and the age of onset (Pickett, 1972).

The speech communication problem encountered by the individual who suffers from a severe early hearing impairment stems from three main factors: 1) poor auditory reception of speech, 2) lack of speech that is intelligible to others, and 3) poor language competence in the sense of normally developed internal grammar and vocabulary. Speechreading and sign language serve for limited communication, but these methods cannot approach speech communication in flexibility and scope (Pickett, 1972).

Speech communication depends on having an adequate knowledge of the language. For pre-lingually, hearing-impaired individuals, this knowledge does not develope sufficiently, no matter how intensively they are educated. Normal language development depends heavily on auditory reception of speech and the consequent oral interaction between a child and other persons (Pickett, 1972).

It is evident that large improvements are needed to aid hearing-impaired individuals in their speech communication. In order to develope better techniques and instruments for improving the speech of hearing-impaired children and adults, it is necessary to understand and have a knowledge of the ways in which the speech of the hearing-impaired deviates from that of normal hearing individuals (Levitt, 1971).

The production of intelligible and natural-sounding speech requires a complex interaction of acoustic, physiological and perceptual mechanisms (Levitt, 1971). Hearing impairments cause deficiencies in this speech communication ability that range from minor to quite severe. The degree of the speech deficiency depends primarily on the severity of impairment and the age of onset (Pickett, 1972).

The speech communication problem encountered by the individual who suffers from a severe early hearing impairment stems from three main factors: 1) poor auditory reception of speech, 2) lack of speech that is intelligible to others, and 3) poor language competence in the sense of normally developed internal grammar and vocabulary. Speechreading and sign language serve for limited communication, but these methods cannot approach speech communication in flexibility and scope (Pickett, 1972).

Speech communication depends on having an adequate knowledge of the language. For pre-lingually, hearing-impaired individuals, this knowledge does not develope sufficiently, no matter how intensively they are educated. Normal language development depends heavily on auditory reception of speech and the consequent oral interaction between a child and other persons (Pickett, 1972).

It is evident that large improvements are needed to aid hearing-impaired individuals in their speech communication. In order to develope better techniques and instruments for improving the speech of hearing-impaired children and adults, it is necessary to understand and have a knowledge of the ways in which the speech of the hearing-impaired deviates from that of normal hearing individuals (Levitt, 1971).

Although considerable information has been gained through the listener's perceptual analysis, one of the major tools that has been useful in gaining this information has been the sound spectrograph, a device for analyzing the spectral characteristics of the speech signal. A number of investigations have attempted to describe the aspects which characterize the speech of the hearing-impaired. In particular there are those kinds of data that spectral analysis has provided.

Characteristics of Deaf Speech

Angelocci, Kopp, and Holbrook (1964), in a spectrographic analysis of the speech of 11 to 14 year old hearing-impaired and normal hearing boys, found that the hearing-impaired had a higher mean fundamental frequency for all vowels. The mean ranges of fundamental frequency and intensity were greater for the normal hearing, and the formants of the hearing-impaired showed significantly less variation. The authors postulated that this result could probably be related to the hearing-impaired child trying to achieve vowel differentiation through varying fundamental frequency and intensity of the voice relatively more than frequency and intensity of the formants. Angelocci, et. al., suggested that physiologically, the hearing-impaired child is probably achieving vowel differentiation by excessive laryngeal variations rather than by adequate control of the articulators.

Another source of error in the speech of the hearing-impaired is the distortion of speaking rate. Calvert (1961) acoustically analyzed the durational characteristics of voiced-voiceless consonants. The hearing-impaired subjects spoke CVCVC syllables in which the final phoneme was held constant, and the middle consonant and succeeding stressed vowel were varied. Measurements were made of the length of the stressed and unstressed vowels, fricative consonants, and the release and closure periods of plosive consonants. The data indicate that hearing-impaired speakers extended all durations except release periods and the relative durational relationships between sounds were not maintained. No systematic pattern in durational distortions were found which correlated with listener judgements of voiced and voiceless phonemes. The author did note that release periods for plosives were not extended by the hearing-impaired but were similar in duration to that of the normals.

Several investigators have explored the perceptually noted lack of blending of successive phonemes in the speech of the hearing-impaired and the effect of expanded duration of vowels and voiced continuents. Rothman (1972) evaluated acoustic and electro-myographic data on consonant-vowel transitions. A constant carrier phrase was used along with a varied key word. The results were similar to those of Angelocci, Kopp, and Holbrook (1964). In hearing-impaired speakers, a restricted range of movements of the formant transitions were evident, as well as a slower rate of movement of the formants. In particular, they observed a longer closure duration for the plosives which would tend to cause discontinuities in the speech and which might allow the hearing-impaired speaker time to produce each phoneme separately. These data point out the obvious lack of coarticulation effects in the speech of the hearing-impaired.

Monsen (1973) has provided further data on the linking problem. He points out that although a phoneme may be correctly articulated, it may be incorrectly coarticulated, and that distorted patterns of coarticulation may be a primary source of poor intelligibility in the speech of the hearing-impaired. Monsen utilized a spectrographic analysis to examine two aspects of coarticulation in hearing-impaired and normal hearing adolescents: 1) the influence of the final consonants (/t/, /s/, /n/, /d/, /z/) on the duration of /i/ and /I/, and 2) the influence of the initial consonants (/b/, /d/, /f/) on the second formant of /i/ and /I/. In normal speech the tense vowel /i/ is relatively longer than the lax vowel /I/, however in the speech of the hearing-impaired, Monsen found that the hearing-impaired speaker did not vary the duration in the normal manner; i.e., did not correctly coarticulate the phonemes. The hearing-impaired speakers appeared to separate the two vowels in what Monsen calls "absolute terms", /i/ was made long and /I/ was short. Monsen suggests that the speaker may rely on duration rather than spectral properties to signal a difference between phonemes. In the measurement of F_2 transitions, Monsen found in agreement with Rothman (1972), that second formant transitions in the hearing-impaired typically appeared reduced. The vowels appear to be very near their target position as soon as they begin.

Applications of Acoustic Analysis Data

Another area in which research in the spectral analysis of speech has been important to the hearing-impaired population is the development of sensory training aids. Through the use of modern speech analyzing techniques, it is possible to extract important information from the speech signal and to display this information by visual presentation. These visual displays can give direct feedback on the articulatory movements and are useful to describe the task to the child (Risberg, 1968).

The first attempt to use an instantaneous visual speech display was made by Alexander Graham Bell in 1874. The current era of visual speech aids for the hearing-impaired began in 1944 at Bell Telephone Laboratories, with the linking of a sound analyzer to instantaneous phosphorescent screens similar to television screens, in order to display speech sound patterns. This device, the visible speech translator (VST), is the best known but most complicated speech analyzing aid. The VST analyzes the frequency pattern of the signal and displays the pattern in time from left to right on a screen. The vertical dimension displays a range of frequencies from low to high. The presence of energy at any frequency is represented by lighting the screen, with intensity displayed by varying brightness (Pickett, 1972). Later versions of this device use a storage screen (Stark, 1968), thus permitting the freezing of the display. This was a very important improvement because it allowed the teacher a better possibility to describe the pattern and define the task before erasing it. In addition, a model pattern can be stored along with the student's production for comparison (Risberg, 1968).

Potter, Kopp, and Green (1947) carried out a series of studies in speech reception with the initial model of the VST. They were able to train viewers to read a large number of words and to converse in simple phrases with careful pronunciation. VST patterns for fluent speech are very complex and it is virtually impossible to communicate totally in this way at the speed of fluent speech. It does appear however, to be useful in supplementing training by concentrating on small segments of speech production.

Stark (1968) reported a pilot study designed to discover what features of speech might be taught to hearing-impaired subjects with the aid of the new version of the VST. Stark found the following speech features were successfully modified in the severely and profoundly hearing-impaired, school-age and adult subjects: 1) abnormally high pitch, 2) incorrect timing, 3) excessive nasalization, and 4) faults of segmental structure.

Several visible speech trainers have been developed which provide somewhat simpler displays than the VST. These spectrum displays present a fairly detailed picture of the speech spectrum in a single display with axes of frequency and intensity. Frequency is displayed along the horizontal axis extending from 200 Hz to 7000 Hz. Intensity is displayed on the vertical axis. One such speech trainer is the LUCIA (Risberg, 1968), designed in Sweden. The display screen is a matrix of blocks, 10 x 20, which are lighted individually according to the frequency pattern of the voice. Frequency is divided into 20 bands and intensity varies in steps of 3 dB; i.e., each block corresponds to a 3 dB change (30 dB span).

A number of experimenters, feeling that the display of complete frequency patterns might be too complex for certain training situations and too expensive for widespread adoption, have designed simpler indicators of frequency patterns using a trace on an oscilloscope. Pickett (1972) describes three sensory training aids that fall in this category.

One such training aid is the voice pitch indicator. Vibrations of the vocal folds are picked up by a contact microphone and the average frequency of the voice is computed over a period of a few milliseconds. Horizontal movement of the oscilloscope trace indicates time. An upward movement of the trace indicates a rise in fundamental frequency and a deflection of the trace indicates a lowering of fundamental frequency (Pickett, 1972).

Mártony (1968) used the voice pitch indicator to modify the overall voice pitch level of severely hard of hearing subjects. Although he obtained substantial improvement, no control group was used and it is not clear whether the measured improvement was due to use of the device or to the additional, specialized training the child received.

Dolaskey, et. al., (1966) used a device quite similar to the voice pitch indicator to give subjects training designed to correct overall pitch control and to establish control of intonation. The results indicated a moderate improvement in pitch quality but no measureable change in control of intonation.

The second indicator of frequency patterns using an oscilloscope display is the consonant-vowel timing indicator. This device splits the speech spectrum into two parts. The high frequency part controls the downward deflection of an oscilloscope trace and the low frequency portion controls the upward deflection. Time is indicated by horizontal movement of the trace (Pickett, 1972).

The third type of oscilloscope display described by Pickett (1972) is the vowel indicator, which plots the average frequency of the spectrum above 1000 Hz against the average frequency of the spectrum below 1000 Hz. The resulting display is a single dot which moves to different positions on the display screen for different vowels which the child may learn to recognize visually. Diphthongs are displayed by movement from one spot to another.

Very simple and inexpensive displays can be provided by the movement of a needle on a meter. Several meter displays have been developed for indicating voice pitch, the quality of sounds such as /s/ and /ʃ/, and nasal intensity (Risberg, 1968).

One such device is the pitch meter which utilizes a contact microphone placed on the throat to pick up vocal fold vibrations. The short term average frequency is calculated and displayed on a meter. Two criterion lights can be set on the instrument which indicate when the average fundamental frequency is either too low or too high (Pickett, 1971).

The second meter device developed by the Risberg group is the s-meter which shows the relative proportion of speech energy in the high frequency region. It is used in training correct production of the /s/ and /ʃ/ phonemes, which have the major portions of their energy in the high region. The greatest deflection on the meter comes from correct production of /s/, and there is very little meter movement for strongly voiced sounds. A criterion light is provided which identifies correct production. The meter has a long time constant, whereas the light has a short time constant to provide more rapid feedback (Pickett, 1971).

Generally, the s-meter appears to be a useful device. Several investigators (Risberg, 1968; Pronovost, 1968; Kringlebotn, 1968) have utilized this device and found it helpful in training the hearing-impaired in the correct production of the /s/ phoneme.

The nasal meter is the third meter device developed by the Risberg group. It uses a contact microphone placed against the soft part of the nose to pick up nasal vibrations. These vibrations are amplified and their relative intensity is displayed by a meter. The device has two criterion lights. One is switched on when the nasal intensity falls below a value corresponding to the meter reading of half-scale. The other is switched on when the intensity exceeds the half-scale deflection. A gain is provided so that the meter indicates a relative measure of nasal intensity (Pickett, 1971).

Over the last 30 years, activity in the development of speech aids for the hearing-impaired has been steadily increasing. Although the trend is changing, relatively few of the devices mentioned previously have undergone adequate evaluative procedures (Levitt, 1972). Two questions can be asked (Levitt & Nye, 1971): "1) what is the device designed to do, and 2) does the device assist in teaching a skill to the hearing-impaired child or improve some performance characteristic, either receptive or expressive." Most of the evaluative work that has been done focuses on the first question. It seems considerable disagreement exists among researchers and educators as to how this problem should be solved.

Recent Developments in Acoustic Analysis of Speech

Computers have been used in speech research since about 1958 (Denes, 1970). Almost all research was carried out in computation centers. Although the large machine offered very powerful computing techniques, there were delays in obtaining the computed results. For many applications this delay was undesirable. For example, in a case where the researcher must adjust his program as the experiment proceeded and be able to evaluate these changes quickly, the delay encountered in the computing center was unacceptable.

The approach that developed to solve this timing problem was time-sharing of large central machines. This approach was unsuccessful and the development of medium and small general purpose computers followed. Currently, practically all computer speech research is done on small on-line machines which can be connected directly to laboratory instruments (Denes, 1970).

On-line computing offers several advantages (Denes, 1970). On-line computers can be connected directly to laboratory devices to provide increased flexibility in presentation of stimuli in speech perception experiments. Until recently most studies on the acoustic characteristics of hearing-impaired speech have used analog filter systems, such as the speech spectrograph (Levitt, 1972). Analog systems lack the precision and flexibility of digital methods. The speech spectrograph allows the choice of two filter bandwidths, whereas the digital system offers considerable flexibility in choice of filter bands. Since one must make a compromise between filter bandwidth and integration time, flexibility is extremely important.

A second advantage of digital techniques is the capacity to analyze large quantities of data. Since differences between hearing-impaired speakers are generally greater than those seen between normal speakers, more data is required to determine the differences between hearing-impaired speakers (Levitt, 1972).

Purpose of This Study

Recently, in the Kansas State University (KSU) Department of Electrical Engineering, an on-line computing system has been developed. Dr. Nasir Ahmed and his students have developed versions of two transforms, the chirp-z and the cepstrum, which are currently being utilized in the acoustic analysis of speech (Levitt, 1972). It was of interest to determine whether joint efforts of the Department of Speech and the Department of Electrical Engineering could work out projects which would benefit students in both departments.

At the time that data were collected for this study, three hearing-impaired students were enrolled in therapy at the KSU Speech and Hearing Center. It was decided that speech samples of these individuals would be obtained and analyzed by digital methods. Two aspects were of principal interest: 1) how do digital spectrograms obtained with the methods available to us, compare with those obtained with conventional analog devices; i.e., are they sufficiently precise to allow their utilization in research projects on the acoustic characteristics of speech, and 2) what other types of data, such as fundamental frequency analysis, do they provide which would be of use to the hearing clinician in preparing therapy procedures.

PROCEDURES

Speakers

Speakers were three college-age males who were receiving aural rehabilitation therapy at the KSU Speech and Hearing Center. Audiograms of each speaker are provided in Appendices A, B, and C.

Recording Procedures

Two sentences were recorded from each speaker on seven consecutive days. These sentences were "Noon is the sleepy time of day." and "I like happy movies better." All recordings were made in a quiet room using an Ampex A6 500 tape recorder and an Electro-Voice Model 664 dynamic microphone. Recording speed was 15 IPS. The intensity level was standardized by fixing the recording level control in position and varying the speaker's distance from the microphone until the major intensity peaks in the speaker's utterance peaked the VU meter near zero VU. This recording was then dubbed on a Wollensak Model 1520 AV tape recorder at a speed of 7.5 IPS.

Since facilities were not available to digitize this tape at KSU, the tape was sent to the Engineering Experiment Station College of Engineering, University of Wisconsin, Madison, Wisconsin to be digitized. There it was digitized by sampling at a rate of approximately 8000 samples per second. The bandwidth of the speech signals is thus 4000 Hz. Subsequently, it was put into a format compatible with the IBM 370/158 computer at the KSU Computing Center.

The digital data was then processed at the computing center using programs which were developed by the KSU Department of Electrical Engineering. These programs were used to accomplish the following: 1) generate digital spectrograms, and 2) compute fundamental frequency histograms associated with each speaker.

Digital Spectrograms

These spectrograms are obtained via the fast Fourier transform algorithm which enables one to compute the Fourier spectrum of successive blocks of digital speech data. Each block consisted of 256 points of digital speech. Successive blocks were overlapped as the spectra were computed.

To construct the spectrograms, the successive spectra were converted to yield a grey scale effect. This is achieved by means of a multiprint technique which uses four characters to give a perceptually linear gradation of eight shades of grey. The character combinations used to achieve the grey level are provided in Table 1 (Stucki, IBM Research Report).

TABLE 1

Grey Level	1	2	3	4	5	6	7	8
1st print		-	=	+	X	X	O	O
2nd print				+	X	X	O	W
3rd print						=	*	M
4th print								*

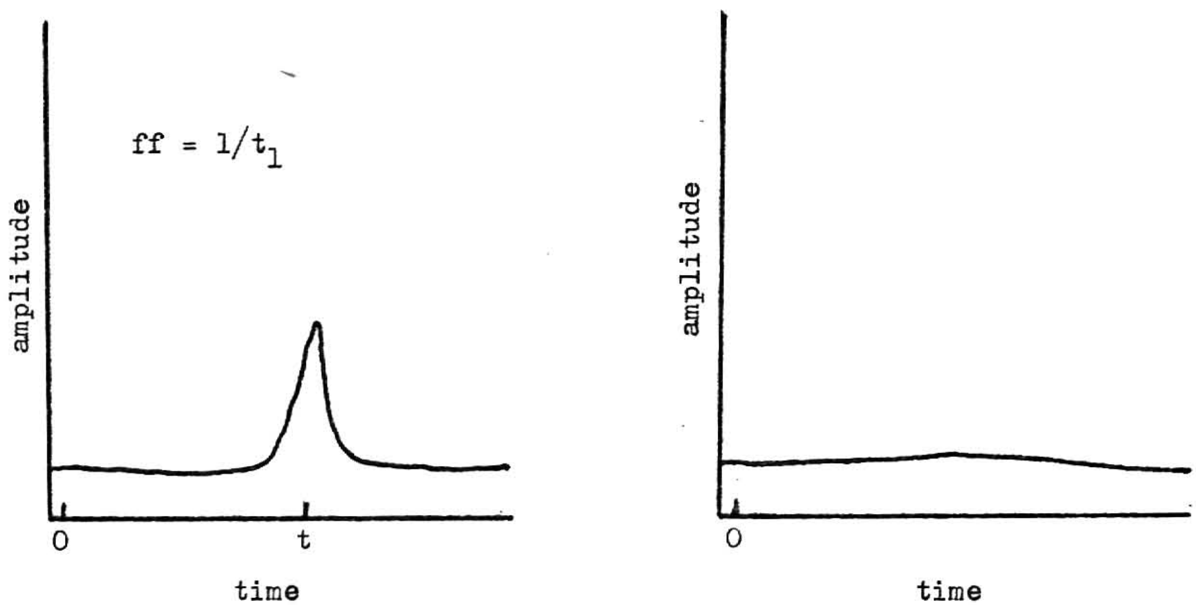
Table 1. The character combinations for the grey levels utilized in the computer generated spectrograms.

While printing the spectrograms, each vertical character line is spaced approximately 30 milliseconds apart. The frequency resolution is 33 Hz; i.e., each horizontal character line is 33 Hz apart, beginning with 0 Hz at the bottom and ending at 4191 Hz at the top, resulting in 128 frequency bands.

Determination of Fundamental Frequency

In order to estimate the fundamental frequency accurately, two stages of data processing are used. The first stage consists of computing the cepstra associated with successive blocks of speech. This process enables one to determine whether a block of speech is voiced or unvoiced, by means of examining the cepstrum* computer program output as illustrated in Figure 1. If a peak is detected then the corresponding block of speech is considered to be voiced, otherwise it is considered to be unvoiced. Voicing is indicated only for those segments which contain 30-60 milliseconds of voicing.

FIGURE 1



(a) voiced case (b) unvoiced case

*The cepstrum is defined as the inverse Fourier transform of the logrithm of the Fourier spectrum of each block of speech data; i.e., the distribution of power with respect to frequency.

In the second stage, the chirp-z transform is used to compute the Fourier spectrum of a block designated as voiced. This computation is carried out in the vicinity of the estimated fundamental frequency given by $1/t_1$ via cepstrum analysis. Such a computation leads to a high resolution power spectrum. The resolution achieved in the present study is $\pm$ 1 Hz.

RESULTS AND DISCUSSION

Figures 2, 3, and 4 display the spectrographic analysis of the sentence "Noon is the sleepy time of day." for each of the three speakers. The upper section of each figure shows the analog spectrogram and the lower section of each figure shows the digital spectrogram. The analog spectrograms were obtained on a Kay Sona-Graph 6061 A spectrum analyzer provided by the Speech Science Laboratory at the University of Kansas, Lawrence, Kansas. A 45 Hz analyzing bandwidth was used. The horizontal dimension represents approximately 2.4 seconds and the vertical dimension gives a range of approximately 3500 Hz. The digital spectrograms provide a frequency range of 4191 Hz and a time span of 4.32 seconds. The analyzing bandwidth was approximately 33 Hz.

Several comments should be noted regarding the analog spectrograms. The baseline was slightly out of adjustment since a mirror image is evident just below the dark continuous line at the bottom of each analog spectrogram. In Figures 2 and 4 it should also be noted that the spectrogram paper was slightly out of position, so that a few tenths of a second of the final portion of the utterance appears at the beginning of the analog spectrogram. The sentence spoken by speaker 2 exceeded the 2.4 second duration. The analog spectrogram shown in Figure 3 was obtained by splicing together two spectrograms which contained the beginning and end of the utterance. It represents a duration of approximately 3.3 seconds. Figure 5 provides an analog spectrogram of the utterance by a normal hearing speaker. This is included to provide a reference for the analog spectrograms. Problems encountered in digitizing the data did not allow for the inclusion of a normal hearing speaker in that group of data.

Figure 2 provides data for the hearing-impaired speaker in this group who had considerably better hearing sensitivity in the lower frequencies then did the other two speakers (see audiograms in Appendices A, B, and C). The temporal pattern and overall intelligibility of speaker 1 tended to be more like that of a normal hearing speaker. It is interesting that for speaker 1, the narrow band analysis gave a fairly accurate representation of the formant structure of the speech sample. This is especially true in the digital spectrogram. In comparing the digital spectrograms of Figures 2, 3, and 4 it is evident that the harmonic structure is more clearly defined for speakers 2 and 3. It seems likely that this result is related to the value of the fundamental frequency. As will be noted later, speaker 1 had a significantly lower fundamental frequency than did the fundamental frequency of speakers 2 and 3. Speaker 2's fundamental frequency shows a very wide variation and close examination of the digital spectrogram indicates some segments where the formant structure is more evident and other segments where the harmonic structure is clearly displayed. It would be these former segments in which the fundamental frequency is lower.

The problem encountered in the digital spectrograms is similar to that encountered when using an analog analysis. A narrow band analysis of a female voice gives an example of the apparent formant structure (Fant, 1968), while the broad band analysis of a high-pitched voice shows a mixture of formant and harmonic structure.

The Fourier transform algorithm used to compute the Fourier spectrum of the blocks of digital speech data looked at the block lengths of 30 milliseconds. Thirty milliseconds was found to be the optimum block length by Oppenheim (1970) for processing normal speech which had a considerably smaller frequency range than did the speech of speakers 2 and 3 in this study. It would be of interest to consider longer block durations in an attempt to find an optimum value for speech signals which cover a fairly large frequency range.

Figures 6, 7, and 8 provide a distribution of the average fundamental frequency of the three hearing-impaired speakers for the first sentence. It is readily apparent that the distribution of speaker 1 is more like what one would expect to see for the normal hearing adult male. However, this result was expected since speaker 1 had about a 30 dB hearing loss at 250 Hz, compared to about a 75 dB loss for speakers 2 and 3. It is unquestionably a result of considerably better low frequency hearing.

The data of interest was obtained from speakers 2 and 3. The average histograms reflect what the experimenter concluded from a perceptual analysis of the tapes of the speakers' voices. Each has a frequency range of approximately 80 to 200 Hz. However, the distribution of speaker 3 has an obvious central tendency at about 170 Hz while speaker 2 has a relatively flat distribution. Perceptually, speaker 2's voice pitch was constantly fluctuating inappropriately. On the other hand, speaker 3 uses a habitual pitch level which is much too high for his age and sex.

For the clinician, these data are quite useful. They provide a clear and concise picture of the speaker's fundamental frequency which gives the clinician a baseline and can be readily used to monitor progress in training more appropriate pitch level. In addition, they provide confirmation of the clinician's perceptual analysis of the voice characteristics.

In working with the young adult and adult age group, these data are particularly useful since the clinician can provide the hearing-impaired speaker with a simple, easy to read visual representation of the pitch of his voice production.

SUMMARY AND CONCLUSIONS

Analog and digital spectrograms of a sentence spoken by three hearing-impaired speakers were generated and compared. Since a critical relation exists between voice fundamental frequency and analyzing filter bandwidth, acoustic analysis of the speech of the hearing-impaired speaker presents a difficult task, due to its extreme variation in fundamental frequency. Visual inspection of the digital spectrograms indicate that the particular duration of the blocks of digitized speech analyzed which is related filter bandwidth, was not an optimum value. Further study which seeks to find an optimum duration for hearing-impaired speakers is warranted.

In addition, data regarding fundamental frequency was obtained in the form of an average fundamental frequency histogram of a single sentence. These data very precisely inform the clinician of the status of the speaker's habitual pitch. For the young adult and adult hearing-impaired speaker, this type of data should be particularly useful in helping them establish more normal pitch patterns.

Information obtained from visual presentations of speech are hard to read, no matter how clear the transform or how extensive the training.

This difficulty is due to the complexity of the speech signal on the phoneme structure of language and the need for speech to be decoded before it can be correctly perceived. The ear contains a readily available processor that decodes the speech signal. There is no equivalent decoder available to the eye. Liberman, et. al. (1968) consider the possibility of presenting speech information to the eye in decoded form. Since much of the speech encoding occurs at the conversion of muscle contraction to the vocal tract, the authors postulate that information concerning articulatory muscle contraction might hold some promise as a useful way of presenting the speech signal to the hearing-impaired (Liberman, et. al., 1968).

REFERENCES

Angelocci, A., Kopp, G., & Holbrook, A. The vowel formants of deaf and normal hearing 11 to 14 year old boys. _Journal of Speech and Hearing Disorders_, 1964, _29_, 156-170.

Calvert, D. R. Some acoustic characteristics of the speech of profoundly deaf individuals. Doctoral dissertation, Stanford University, 1961. Cited by Keiser, J. H. Perceptual and acoustic effects of modifying the speaking rate of deaf speakers. Unpublished paper, University of Iowa, 1974. P. 4-5.

Denes, P. B. On-line computers for speech research. _IEEE Transactions on Audio and Electroacoustics_, 1970, _AU-18_, 418-425.

Dolaskey, L., Karis, C., Phillips, N. & Pronovost, W. Teaching vocal pitch patterns using visual feedback from the instantaneous pitch-period indicator for self-monitoring. VRA Project No. 1907-S, Boston, Mass.: Northeastern University, 1966.

Fant, G. Analysis and synthesis of speech processes. In B. Malmberg (Ed.), _Manual of phonetics_. Amsterdam: North-Holland Publishing Co., 1968.

Keiser, J. H. Perceptual and acoustic effects of modifying the speaking rate of deaf speakers. Unpublished paper, University of Iowa, 1974.

Kringlebotn, M. Experiments with some visual and vibrotactile aids for the deaf. _American Annals for the Deaf_, 1968, _113_, 311-316.

Levitt, H. Speech production and the deaf child. In L. E. Connor (Ed.), _Speech and the deaf child: Knowledge and use_. Washington, D. C.: The Alexander Graham Bell Association for the Deaf, Inc., 1971.

Levitt, H., & Nye, P. W. _Sensory training aids for the hearing-impaired._ Washington, D. C.: National Academy of Engineering, 1971.

Levitt, H. Acoustic analysis of deaf speech using digital processing techniques. IEEE Transactions on Audio and Electroacoustics, 1972, AU-20, 35-41.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Why are speech spectrograms hard to read? American Annals for the Deaf, 113, 1968, 127-133.

Mártony, J. On the correction of voice pitch level for severly hard of hearing subjects. American Annals for the Deaf, 1968, 113, 195-202.

Monsen, R. B. On coarticulation: Vowel duration and f_2 transitions in speech of deaf and normally hearing children. Paper presented at the Acoustical Siciety of America Meeting, Los Angelos, Fall 1973.

Oppenheim, A. V. Speech spectrograms using the fast fourier transform. IEEE Spectrum, 1970, 57-62.

Pickett, J. M. Speech science research and speech communication for the deaf. In L. E. Connor (Ed.), Speech and the deaf child: Knowledge and use. Washington, D. C.: The Alexander Graham Bell Association for the Deaf, Inc., 1971.

Pickett, J. M. Status of speech-analyzing communication aids for the deaf. IEEE Transactions on Audio and Electroacoustics, 1972, AU-20 3-8.

Potter, R. K., Kopp, A. G., & Green, H. C., Visible Speech. New York: D van Nostrand Co., Inc., 1947.

Pronovost W., Yenkin, L., & Anderson, D. C. The voice visualizer. American Annals for the Deaf, 1968, 113, 230-238.

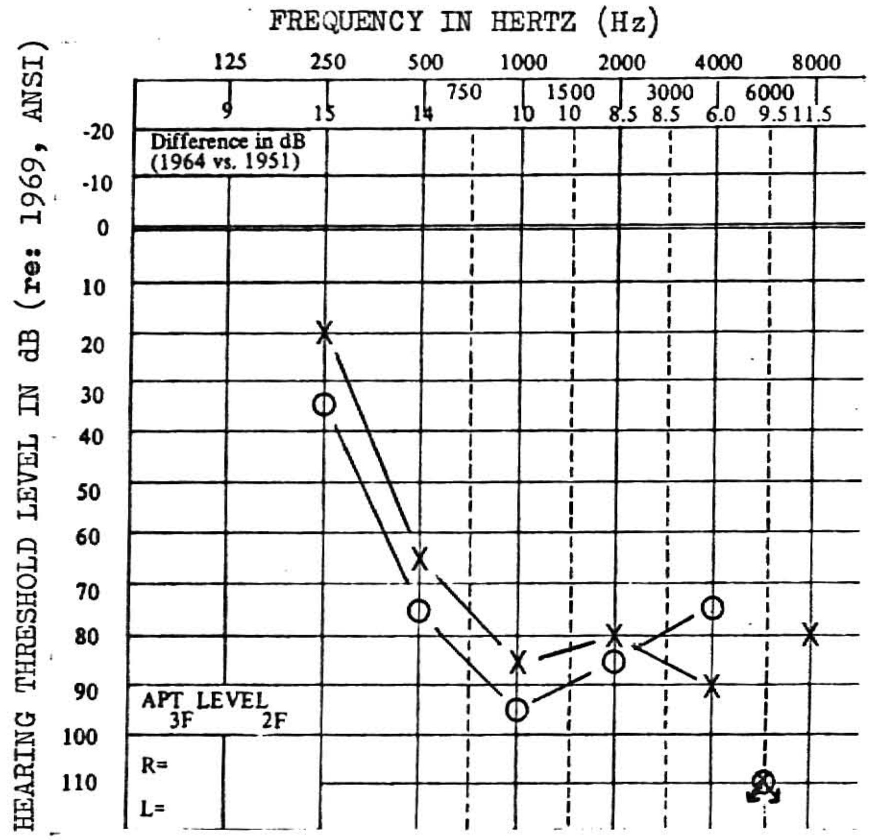
Risberg, A. Visual aids for speech correction. American Annals for the Deaf, 1968, 113, 178-194.

Rothman, H. B. An acoustic and electro-myographic investigation of consonant-vowel transitions in the speech of deaf adults. Pre-publication draft from author, 1972, Cited by Keiser, J. H. Perceptual and acoustic effects of modifying the speaking rate of deaf speakers. Unpublished paper. University of Iowa, 1974. P. 4.

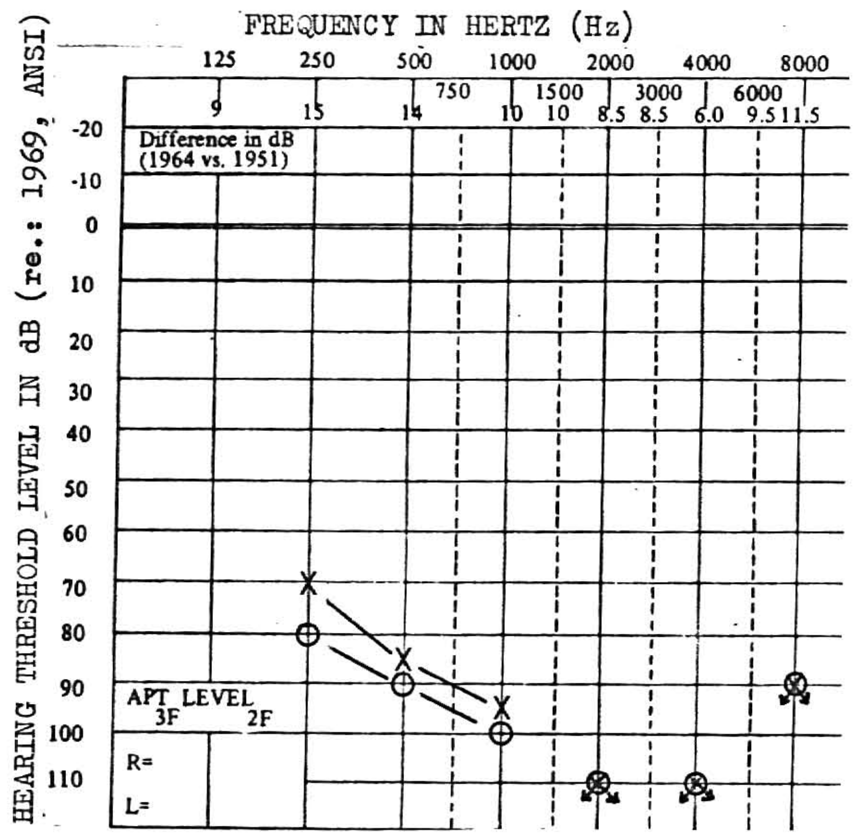
Stark, R. E., Cullen, J. K., & Chase, R. A. Preliminary work with the new bell telephone visible speech translator. American Annals for the Deaf, 1968, 113, 205-214.

Stucki, P. Generation of grey lines by computer for simulation of visual information systems. IBM Research Report, R2-282.

APPENDIX A. Pure-tone air conduction audiogram for speaker 1.



APPENDIX B. Pure-tone air conduction audiogram for speaker 2.



APPENDIX C. Pure-tone air conduction audiogram for speaker 3.

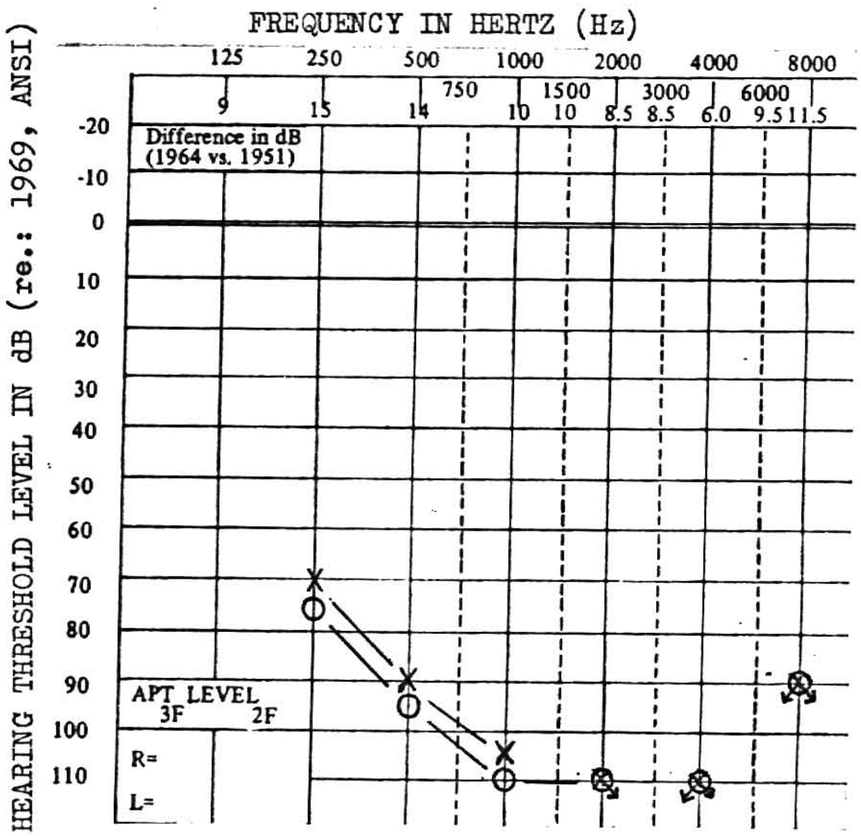


FIGURE 2

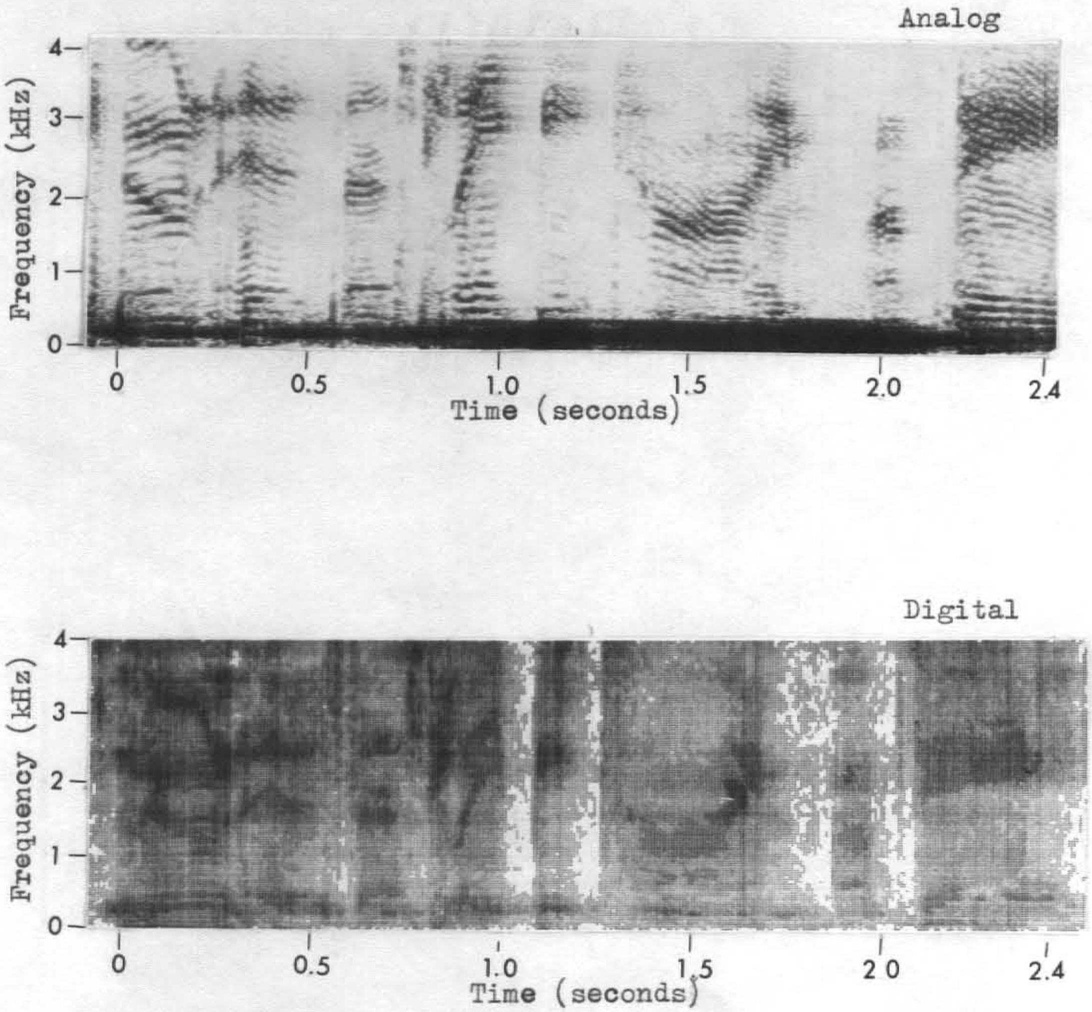


Figure 2. Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 1.

FIGURE 3

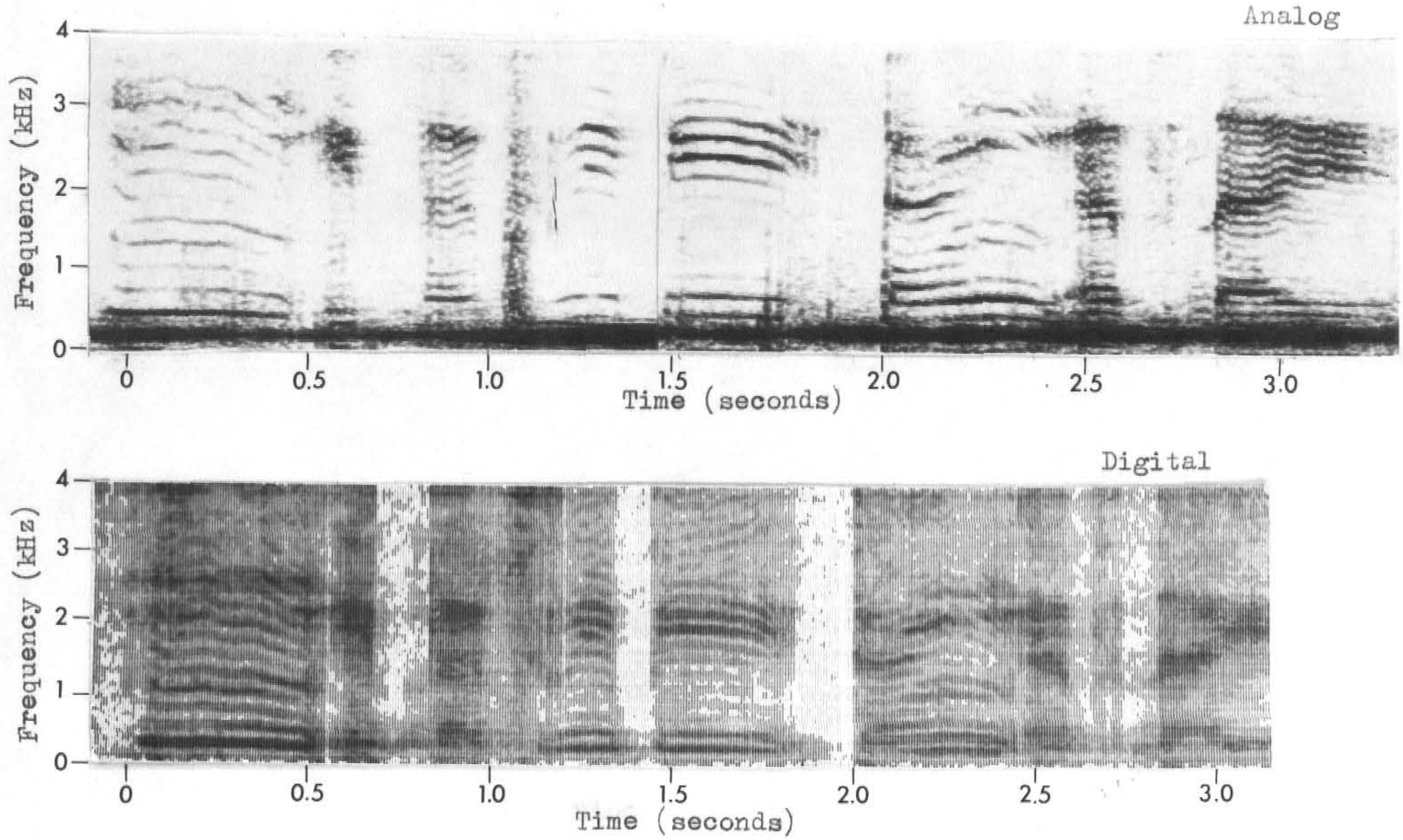


Figure 3. Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 2.

FIGURE 4

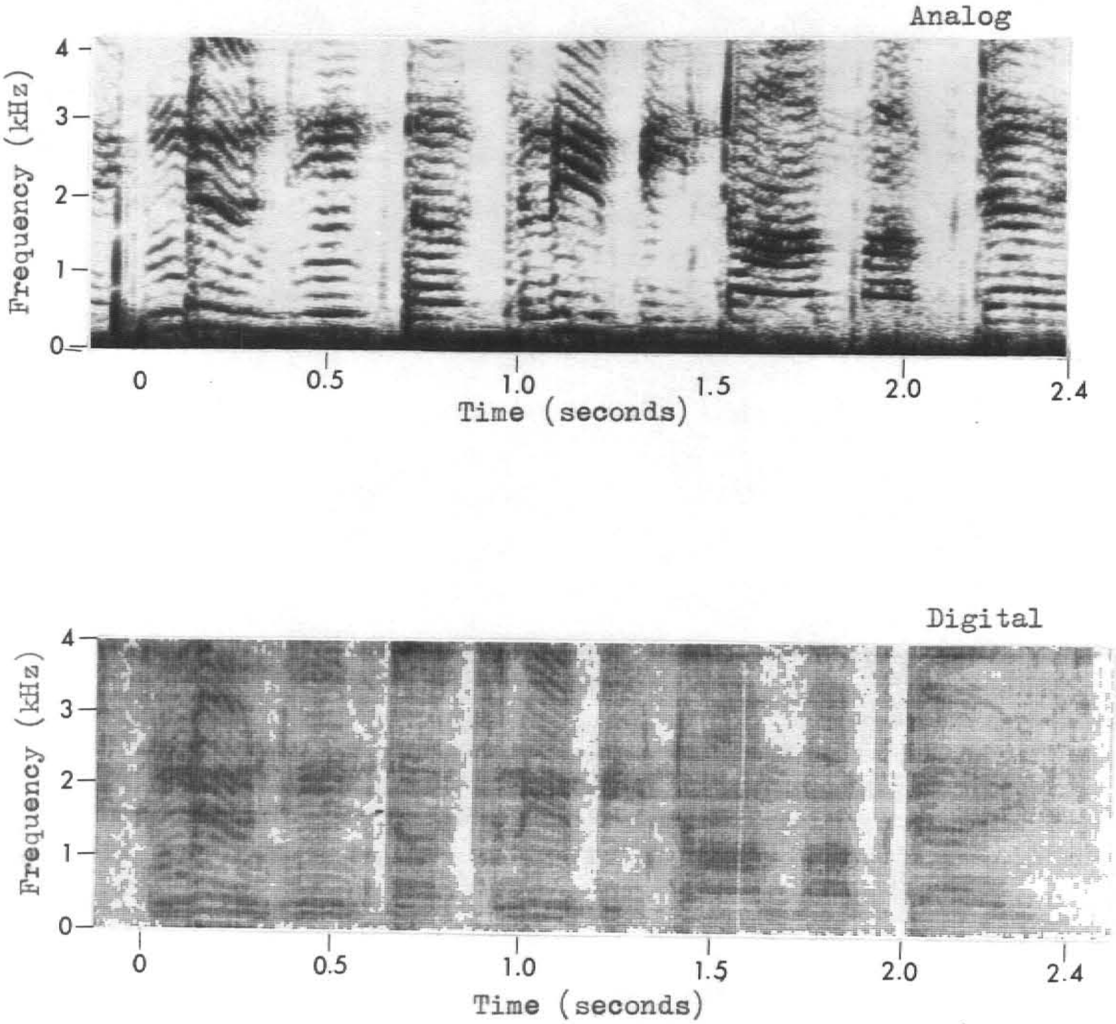


Figure 4. Spectrograms of the sentence "Noon is the sleepy time of day." spoken by speaker 3.

FIGURE 5.

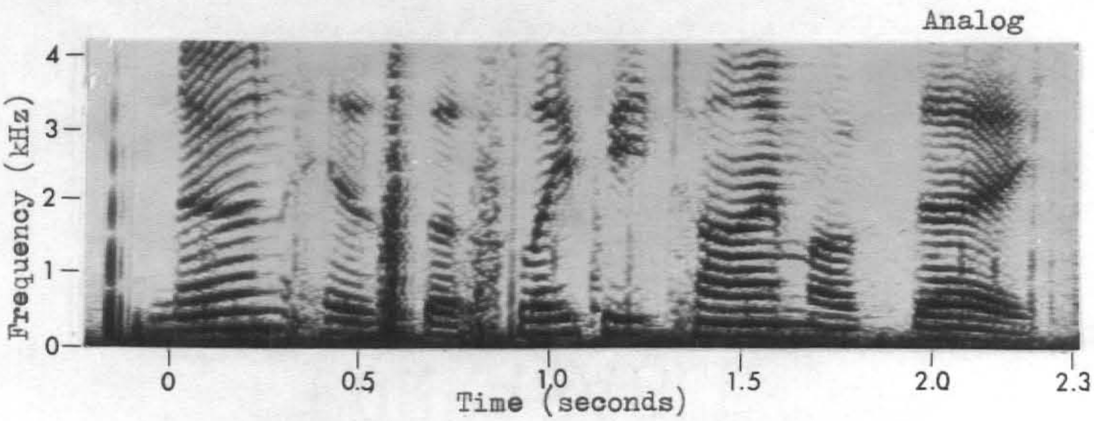


Figure 5. Spectrogram of the sentence "Noon is the sleepy time of day." spoken by a normal hearing speaker.

FIGURE 6

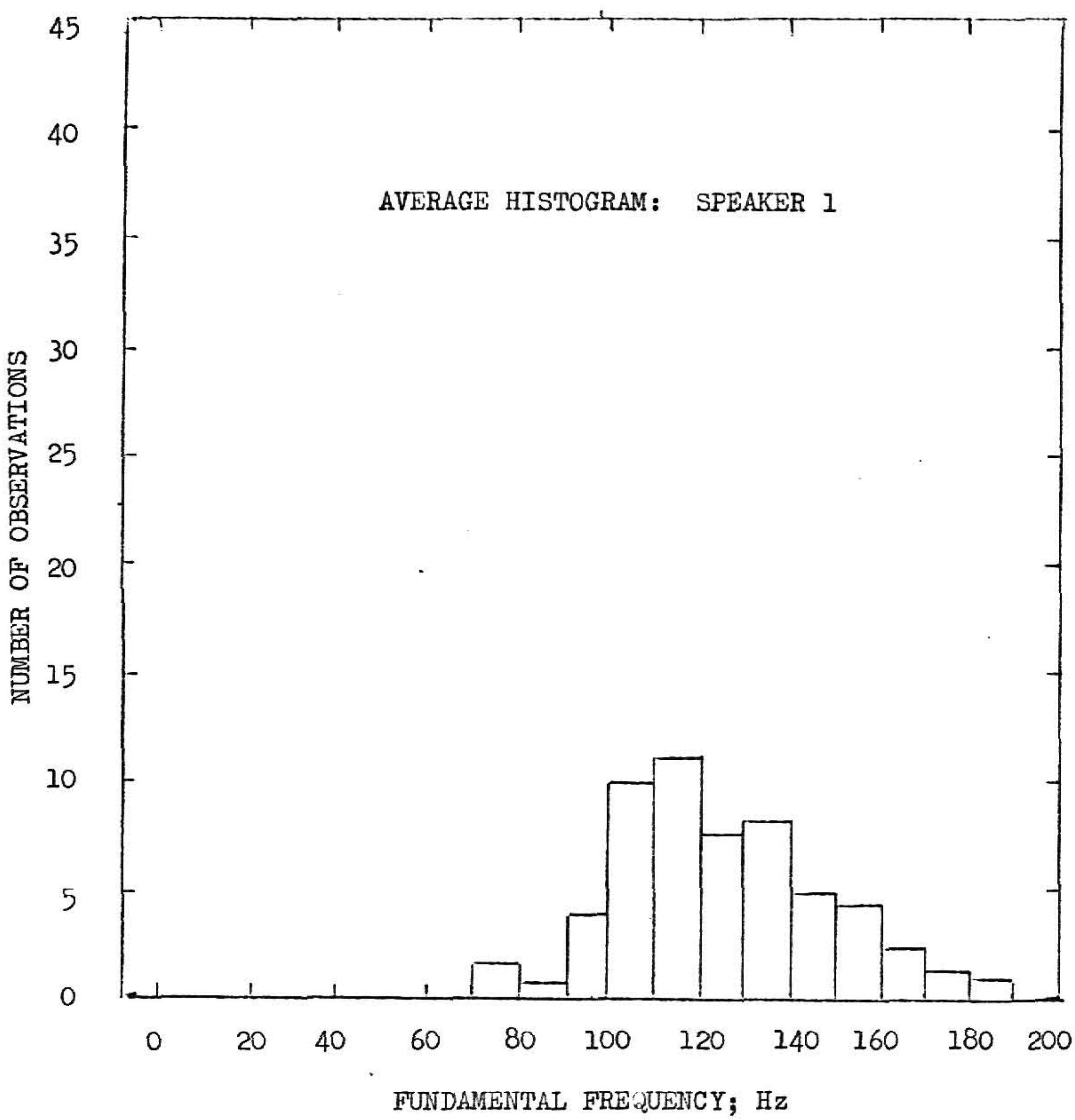


Figure 6. Average fundamental frequency histogram obtained from seven replications of the sentence "Noon is the sleepy time of day."

FIGURE 7

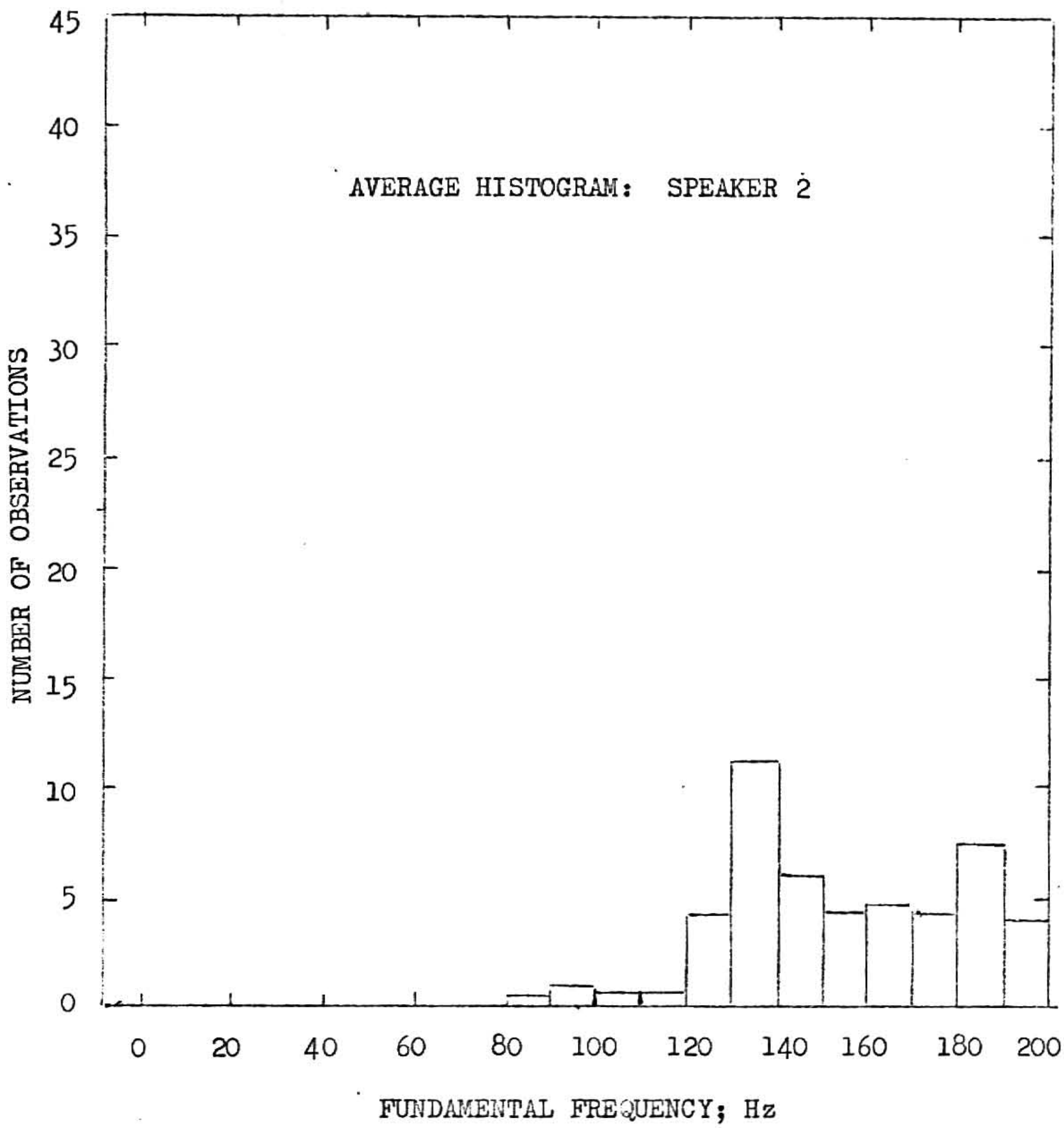


Figure 7. Average fundamental frequency histogram obtained from seven replications of the sentence "Noon is the sleepy time of day."

FIGURE 8

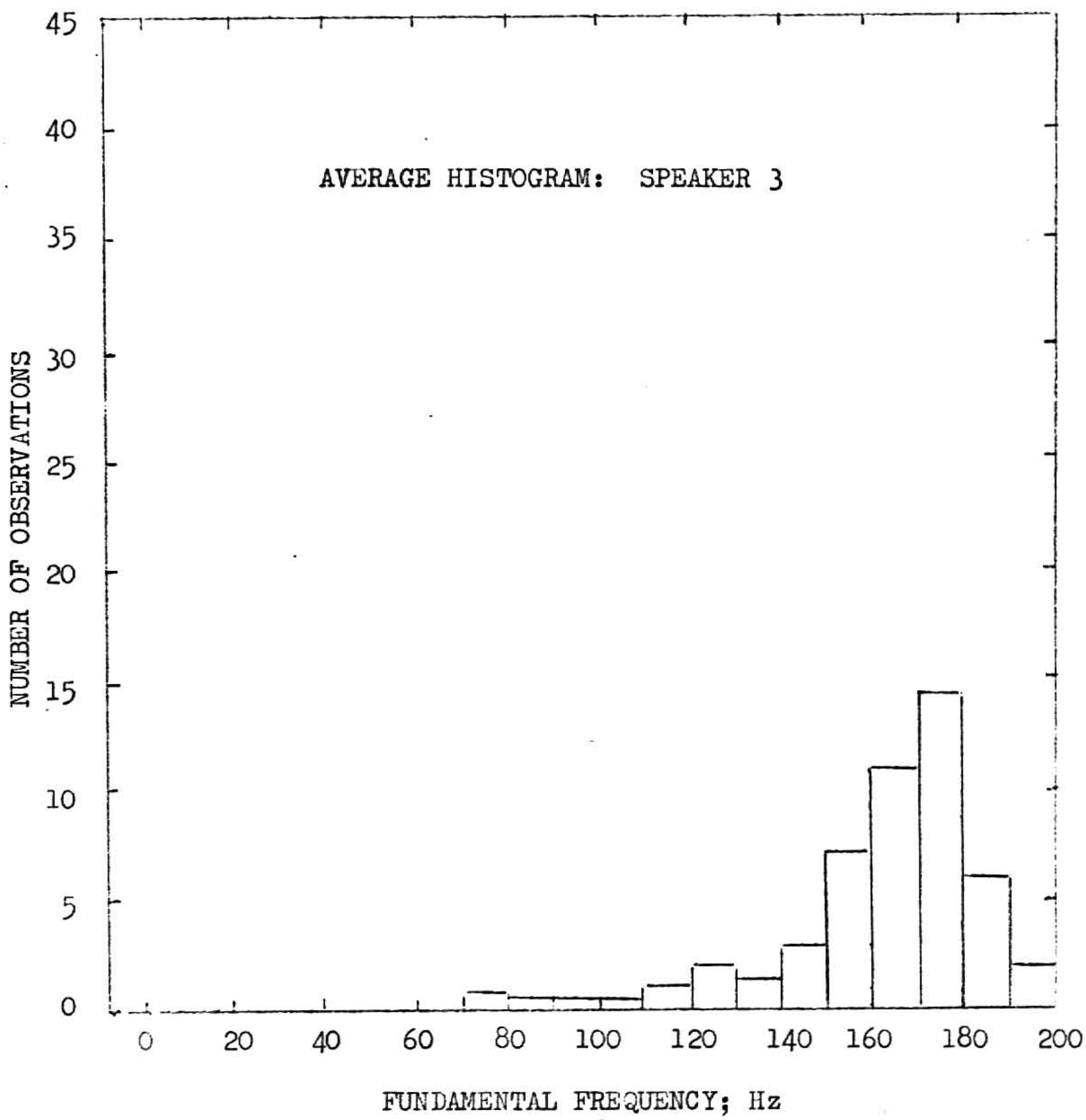


Figure 8. Average fundamental frequency histogram obtained from seven replications of the sentence "Noon is the sleepy time of day."

SOME ANALYSES OF THE SPEECH OF HEARING-IMPAIRED
SPEAKERS UTILIZING DIGITAL SIGNAL PROCESSING TECHNIQUES

by

DEBRA JANE BRIERY

B.S., Kansas State University, 1973

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Large improvements are needed to aid hearing-impaired individuals in their speech communication. Research involving the spectral analysis of speech which attempt to describe the aspects that characterize the speech of the hearing-impaired was reviewed. Spectral analysis research has also been important in the development of sensory training aids. A summary of research activity and corresponding findings of visual speech aids for the hearing-impaired was presented.

Recent developments in the acoustic analysis of speech has involved the use of computers. Currently, practically all computer speech research is done on small on-line machines which can be connected directly to laboratory instruments. The present study generated and compared analog and digital spectrograms of a sentence spoken by three hearing-impaired speakers. Two aspects were of principal interest: 1) how do digital spectrograms obtained with the methods available to us compare with those obtained with conventional analog devices, and 2) what other types of data, such as fundamental frequency analysis, do they provide which would be of use to the hearing clinician in preparing therapy procedures.

Comparison of analog and digital speech analysis methods indicated that further research and refinement in digital computer analysis techniques is warranted. However, since a critical relationship exists between voice fundamental frequency and analyzing filter bandwidth, acoustic analysis of the speech of the hearing-impaired presents a difficult task due to its extreme variation in fundamental frequency. Further study, which seeks to find an optimum duration for analyzing blocks of digitized speech is needed. In addition, data regarding fundamental frequency obtained in the form of an average fundamental frequency histogram provide useful information about the status and progress of the hearing-impaired individual's habitual pitch patterns.