A computer based analysis of the effects of rhythm modification on the intelligibility of the speech of hearing and deaf subjects

Harry Lang

Follow this and additional works at: https://scholarworks.rit.edu/theses

Recommended Citation

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
A COMPUTER BASED ANALYSIS OF THE EFFECTS OF RHYTHM MODIFICATION ON THE INTELLIGIBILITY OF THE SPEECH OF HEARING AND DEAF SUBJECTS

by

Harry G. Lang

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE in

Electrical Engineering

Approved by:

Prof. Harvey E. Rhody
(Thesis Advisor)

Prof. Watson F. Walker

Illegible Signature
(Director, Center for Communications Research)

Illegible Signature
(Technical Staff, CCR)

Prof. Illegible Signature
(Department Head)

DEPARTMENT OF ELECTRICAL ENGINEERING

COLLEGE OF ENGINEERING

ROCHESTER INSTITUTE OF TECHNOLOGY

APRIL, 1975
ACKNOWLEDGEMENTS

My special thanks go to Professor Harvey Rhody, who not only guided me with the technical aspects of the study; but who also "lent me his ear"—literally—to audit the results of my work. I am also indebted to Dr. Robert Houde and Dr. William Newcomb for their enthusiastic support and technical assistance in the project, and to the professional staff in the Communication Center at the National Technical Institute for the Deaf.

I would like also to express my thanks to Dr. Harry Levitt and Dr. June Miller for their assistance in the background literature study, and to the six listeners who volunteered their time to audit the speech: Marv Sachs, Karen Connor, Bev Price, Marilyn Fowler, Paul Peterson and Larry Hill.

And to Bonnie, my wife, for her patience and encouragement.
ABSTRACT

The speech of profoundly deaf persons often exhibits acquired unnatural rhythms, or a random pattern of rhythms. Inappropriate pause-time and speech-time durations are common in their speech. Specific rhythm deficiencies include abnormal rate of syllable utterance, improper grouping, poor timing and phrasing of syllables and unnatural stress for accent and emphasis. Assuming that temporal features are fundamental to the naturalness of spoken language, these abnormal timing patterns are often detractive. They may even be important factors in the decreased intelligibility of the speech.

This thesis explores the significance of temporal cues in the rhythmic patterns of speech. An analysis-synthesis approach was employed based on the encoding and decoding of speech by a tandem chain of digital computer operations. Rhythm as a factor in the speech intelligibility of deaf and normal-hearing subjects was investigated.

The results of this study support the general hypothesis that rhythm and rhythmic intuition are important to the perception of speech.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>...........................................................</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>...........................................................</td>
<td>vi</td>
</tr>
</tbody>
</table>

## I. INTRODUCTION

A. Statement of the Problem .............................................. 1
B. The Approach .......................................................... 5
C. Other Attempts at Timing Modification ............................. 9
D. Computer Applications to the Rhythm Problem ................. 13

## II. REVIEW OF THE LITERATURE

A. The Significance of Timing in Speech Production ............ 17
B. Temporal Features in the Speech of the Deaf ................. 21
C. Voice Quality Characteristics of Deaf Speakers ............. 26

## III. DESCRIPTION OF THE ANALYSIS-SYNTHESIS SYSTEM

A. Digital Computer Operations in Speech Processing .......... 31
B. The Analysis Process .................................................. 33
C. Rhythm Analysis ....................................................... 46
   1. Segmentation of Utterances ...................................... 35
   2. Duration Measurements ........................................... 46
   3. Reallignment of Short-time Spectra ............................ 47
D. The Synthesis Process .................................................. 53
IV. EXPERIMENTAL PROCEDURE

A. Preliminary System Test ........................................... 54.
B. Criteria for Selection of the Subject ......................... 55
C. Sentence Inventory .................................................. 59
D. Description of the Tape Auditions .............................. 65

V. RESULTS

A. Deaf Subject .......................................................... 67
B. Normal-hearing Subject ............................................ 77

VI. DISCUSSION

A. General Discussion of the Results ............................... 86
B. Some Factors Influencing the Data ............................... 88
C. Additional Findings .................................................. 91

VII. CONCLUSIONS

A. Summary .............................................................. 94
B. Implications for Further Research ............................... 96

VIII. APPENDICES

A. Sonagrams of Test Sentences in the Inventory ............... 98
B. A Brief Review of Literature on Segmentation ............... 109
C. A Brief Review of Literature on Rhythm Feedback .......... 113
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computer-generated Spectrogram (deaf speaker)</td>
</tr>
<tr>
<td>2. Computer-generated Spectrogram (hearing speaker)</td>
</tr>
<tr>
<td>3. Computer-generated Spectrogram</td>
</tr>
<tr>
<td>4. Kay Sonagram (deaf speaker)</td>
</tr>
<tr>
<td>5. Kay Sonagram (hearing speaker)</td>
</tr>
<tr>
<td>6. Block Diagram of the Analysis System</td>
</tr>
<tr>
<td>7. Block Diagram of the Synthesis System</td>
</tr>
<tr>
<td>8. Filter Response Curve</td>
</tr>
<tr>
<td>9. Segmentation Algorithm Printout</td>
</tr>
<tr>
<td>10. Spectral Slope Change</td>
</tr>
<tr>
<td>11. Spectral Slope Change as a Function of Time</td>
</tr>
<tr>
<td>12. Vocal Tract Resonance Display</td>
</tr>
<tr>
<td>13. Computer Construction of a New Speech Segment</td>
</tr>
<tr>
<td>14. Computer-generated Spectrograms of Rhythm-Modified Speech</td>
</tr>
<tr>
<td>15. Rhythm Modification (Expansion of a Speech Segment)</td>
</tr>
<tr>
<td>16. Rhythm Modification (Regrouping of a Phrase)</td>
</tr>
<tr>
<td>17. Audiogram of the Deaf Subject</td>
</tr>
<tr>
<td>18. Spectrogram of Deaf Subject for Durational Comparison</td>
</tr>
<tr>
<td>19. Spectrogram of Normal-hearing Subject for Durational Comparison</td>
</tr>
<tr>
<td>20. Spectrogram Showing Temporal Distortion</td>
</tr>
<tr>
<td>21. Instructions to the Listeners</td>
</tr>
<tr>
<td>22-31. Sonagrams of Test Sentences</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Digital Computer Algorithms and Their Functions</td>
<td>31</td>
</tr>
<tr>
<td>2.</td>
<td>Communication Profile for the Deaf Subject</td>
<td>59</td>
</tr>
<tr>
<td>3.</td>
<td>Inventory of Test Sentences</td>
<td>60</td>
</tr>
<tr>
<td>4-7.</td>
<td>Duration Measurements</td>
<td>61-64</td>
</tr>
<tr>
<td>8.</td>
<td>Key Contextual Words in the Inventory</td>
<td>70</td>
</tr>
<tr>
<td>9-18.</td>
<td>Intelligibility Analysis Data (Deaf Subject)</td>
<td>71-75</td>
</tr>
<tr>
<td>19.</td>
<td>Summary Table of Analyses (Deaf Subject)</td>
<td>76</td>
</tr>
<tr>
<td>20-29.</td>
<td>Intelligibility Analysis Data (Hearing Subject)</td>
<td>80-84</td>
</tr>
<tr>
<td>30.</td>
<td>Summary Table of Analyses (Hearing Subject)</td>
<td>85</td>
</tr>
</tbody>
</table>
INTRODUCTION

Statement of the Problem

One of the most obvious acoustic characteristics found in the speech of profoundly deaf is the distortion of temporal patterns. As demonstrated by the literature on this subject, many writers\(^1\) believe that the faulty timing found in the speech of the profoundly deaf shows significant correlations to both intelligibility and naturalness.

Vowels are frequently distorted, often beyond identification. Phonation time is often two to four times longer than the corresponding speech rate of normal-hearing speakers. Abnormal pause intervals also add to the excessive total duration of phrasing.

The discontinuity of connected speech caused by speaking each syllable or word as a separate entity contributes to a loss of intelligibility. Furthermore, when the temporal framework of the speech of the deaf is distorted, it is difficult for a listener to attend to the context.

---

\(^1\)See pages 16 to 25 for a survey of the literature.
Levitt\textsuperscript{2} asserts that the prolongation of vowels and continuants is sometimes so exaggerated that the listener may not understand a deaf speaker under these conditions, even if all the phonemes are correctly identified.

The significance of duration as one of the physical variables of stress may explain, in part, the existence of two more deficiencies found in the speech of the deaf: stress and intonation.

Experiments reported by Denes and Fry\textsuperscript{3} principally involved perceptual tests with synthesized syllables. A series of experiments was performed with different intensities, fundamental frequencies and durations. The results indicated that although all three were significant cues in the perception of stress, \textit{duration} and fundamental frequency were generally of more importance. They conclude that perception of stress in natural speech depends on the interaction of a number of cues. Vowel quality, amplitude and \textit{duration} all seem to be factors in the perception of intonation.

The absence of auditory feedback may explain this faulty stress in deaf speech. Duration-type stress may, however, be the most apparent stress-type in speechreading.


\footnotesize{\textsuperscript{3}P. Denes and D. Fry, "Effect of Duration on the Perception of Voicing," \textit{J.A.S.A.}, 27 (1955), 761-764.}
Stress is not easily read on the lips. The emphasis on certain signs in manual communication doesn't give much information as to its acoustic counterpart either, especially to prelingually deaf, who have little or no aural recollection. Since duration-type stress is more apparent on the lips than intensity- and pitch-type stress, it may be that some profoundly deaf individuals assume duration-type stress as part of their manner of speaking. This may even be a factor in the perception of their speech.

In terms of perceiving stress through the visual modality, this author, who is himself profoundly and postlingually deaf, finds the reinforcement of aural recollection invaluable to lipreading. It is unfortunate that there is a total absence of stress indicators in printed English.

Speech therapists themselves at times sacrifice correct duration to emphasize the proper positioning of the oral articulators when teaching a deaf person to speak a particular word. The Ewings share a similar view:

Attempting to make word patterns clearer, teachers may open their mouthes too widely by dropping the lower jaw and mouthing words. Children then begin to imitate the mouthed patterns of words that they see and in so doing acquire unnatural and laboured habits of speech that militate against its fluency and rhythm.  

All of this leads to the conclusion that speech reading is an insufficient replacement for acoustical monitoring of the stress and intonation parameters.

It is interesting to add to this discussion that the profoundly deaf are not isolated from localisms and dialects of our nation. This is especially true with the durational and glide features associated with the vowels.

Faulty timing features in the speech of the profoundly deaf can be associated with dialectal origins, therefore, as well as with the coordination of the articulators, central nervous system disorders and aerodynamic constraints, which are discussed later in this paper.

The purpose of this study is to investigate the rhythm characteristics of a profoundly deaf speaker using the digital computer as an analysis-synthesis tool. The specific questions are these:

1. Can the rhythm characteristics of a profoundly deaf speaker be isolated to allow a diagnostic evaluation by a computerized tempo modification of the speech wave?
2. Will listeners be able to perceive a higher percentage of a speech sample spoken by a profoundly deaf person if the rhythm is realigned to be more natural?
3. To what extent is the rhythmic intuition of the listeners important in speech perception?
The Approach

Although many writers have consistently classified malphrasing and incorrect rate of production as important factors in intelligibility ratings of deaf speech, few have attempted to alter the rhythmic patterns through applied research to test the perceptual judgments of the tempo-modified speech signal.

In this investigation, the durations of speech and pause segments were uniformly and consistently expanded or condensed via digital computer techniques in order to realign the temporal framework of speech samples spoken by a profoundly deaf student at the National Technical Institute for the Deaf.

The timing is modified in this way: Short-time spectra are printed for every ten milliseconds of speech in a computer-generated, three-dimensional spectrogram, as shown in Figures 1 and 2.

Figure 1  A computer-generated spectrogram of the word "A" spoken by the deaf subject. Only twenty-three filter outputs shown (20 Hz - 3220 Hz)

Figure 2  A computer-generated spectrogram of the word "MAN" spoken by the normal-hearing subject. Twenty-three filter outputs shown (20 Hz - 3220 Hz)
The computer-generated spectrogram is produced on 15 inch by 11 inch (38 cm by 28 cm) printout sheets, as illustrated in Figure 3:

In the computer-generated spectrogram, frequency is displayed vertically, time horizontally, and power by printed numerical values.
A Kay sonagram was made for the words "A MAN" spoken by the deaf subject. The sonagram is shown in Figure 4.

The reader should compare the portion of the sonagram for the word "A" with the computer-generated spectrogram in Figure 1. The excessive nasality in the word "MAN" is apparent. A sonagram of the word "MAN" as spoken by the normal-hearing subject is shown in Figure 5.
The relative duration and amount of nasality are clearly different.

In either representation of the speech wave, abnormality of vocal quality and, in particular, speech rhythmic patterns, can be ascertained by distortions in the spectral display produced by changes in the intensity-frequency-time pattern.

Simultaneous with the production of the computer spectra, the digital data is stored on magnetic tapes.

Each ten millisecond short-time spectrum can be duplicated or deleted to expand or condense the speech segments, respectively. The process is done consistently, working from the center of the speech segment toward the more critical boundaries.

The analysis-synthesis approach\(^5\) used in this study has one significant dimension: deletion of spectral energy involves no direct changes in the articulatory or breathing apparatus. The formants are still located in the same positions. The relative lengths of the speech and pause segments are altered; but the fundamental frequency data, voice/voicelessness information and overall power trends are not modified significantly. Other advantages (such as no change in the speaker's emotional state) are also apparent. Thus, this approach provides a unique way of observing the effects of change in rhythm as a function of change in syllable segment length, while all other factors remain unaltered.

---

\(^5\)See pages 31 to 53 for a detailed description of the Analysis-Synthesis System.
Other Attempts at Timing Modification

Much research has been done on measurements of speaking rate and the effects of rate differences on intelligibility. Speech rate, the number of events divided by the total speaking time, can be measured accurately using various instruments, as discussed by Drisko. 6

These techniques have been used to measure the speaking rates of both normal-hearing and deaf speakers. In many reports on the speaking rate of deaf persons, distortions were frequently demonstrated in relative phonation time and in inappropriately varied duration as a function of phonetic environment.

The correlations found between slow speaking rate (poor rhythm ratings) and low intelligibility have prompted many investigators to suggest more intensive training for the deaf in this area. 7

Surprisingly, there has been very little controlled research in rhythm training of deaf speakers through the visual, vibro-tactile or kinesthetic modalities.

Writers who suggest the metronome or other pulsing devices in speech training fail to realize that this precludes use of reading materials. 8

---


7 M. Drisko, p.13.

8 M. Drisko, p.17.
Azrin, Jones and Flye have experimented with a transitorized vibro-tactile wrist bone oscillator, and have found the reinforcement valuable in improving the rhythmic characteristics of stutterers.\(^9\)

Drisko\(^10\) applied this kind of stimulus to the deaf, concluding that it is an effective method for training subjects to increase their speaking rate. Results also indicated that each child had individual patterns of acquisition and generalization, and what a child's training program should be specific to his own reading rate.

Two controlled studies reported in the literature involving actual training in rhythm characteristics show varied results. John and Howarth noted an improvement in intelligibility of phrases spoken by twenty-nine deaf children after a few minutes of therapy directed toward durational aspects.\(^11\)

The training was brief, with most of the phrases having three to eight words. Feedback was largely through use of whatever residual hearing the child had. A 56\% improvement in word intelligibility, and an even greater improvement in the listeners' ability to recognize complete syntactic patterns, was found.

---


\(^10\) M. Drisko, 48-49.

On the other hand, Houde found that the intelligibility of two of his four deaf subjects decreased after six hours of training. In this investigation, the deaf subjects were given training in "...(a) the control of duration and intensity suitable to represent two degrees of syllable stress and (b) the elimination of inappropriate pauses between syllables." The feedback in Houde's study was through visual display.

The results of this investigation were surprising. There was a loss of intelligibility for the two subjects who showed most improvement in the trained rhythmic aspects of speech. Houde elucidates:

One plausible explanation of this loss in intelligibility is that increased attention to the rhythmic aspects of speech is accompanied by a relaxation of attention to the articulatory aspects.

An analysis of the findings of John and Howarth, and those of Houde, poses a question as to whether or not a short-term therapy experience aimed at improving durational errors can be evaluated effectively with only a few subjects.

Duration training may be more successful if combined with training in articulation and breathing.


14 R. Houde, p.6.
This writer believes that short-term therapy, even over a few weeks to a month, may produce temporary modifications in the breathing and articulatory apparatus, with consequent changes in the intelligibility. Perhaps the deaf speaker should be given more time to adapt to his new speaking habits before the effects of training can be evaluated. Anyone who has studied French realizes the difficulty in mastering the nasal "r" in "professeur", for example, and only through repetition and practice does he feel comfortable that his version fits appropriately into a sentence.

Calvert mentions that one of the most crucial errors for the intelligibility of the speech of the deaf was the "surd-sonant" or voiced-voiceless error. He further points out that since the difference in durations of speech sounds such as in the (p) - (b), or (t) - (d) distinctions have been found to affect their perception, then it seems possible that durational distortion may be a contributing factor in this surd-sonant error of the deaf.15

If this were true, then any attempt at modification of the timing of the profoundly deaf person's speech through training would most likely introduce new errors which might directly influence perception. For a short therapeutic session, then, it would probably make no difference whether segmental or suprasegmental timing errors are attended to; new breathing and articulatory habits are being acquired; and a new acoustic representation of the sound is being produced.

---

Computer Applications to the Rhythm Problem

Recent advances in digital computer technology have paved the way for applications to speech pathology. Levitt, in particular, has opened up new avenues in the field of research into the characteristics of the speech of the deaf. His discussions of the potential of analysis-by-synthesis of deaf speech are especially enlightening.

Using the technique designed by C. G. Bell, et al., Levitt simulated the speech of deaf children in order to study the speech problems quantitatively. The results of his work are exciting:

This result indicates that synthetic normal speech sounded more natural than real deaf speech. In fact, all of the synthesized versions were judged more natural than the real speech of the very poor deaf child. The above results imply that, given a simulated version of a poor deaf speaker, it should be possible to improve the quality of the simulated speech to a level where it is judged to be more natural than the original deaf speech. The changes required could be extracted directly from the parameters used in synthesizing a good deaf or normal speaker...in principle, it would seem that the technique for simulating deaf speech is a feasible way of finding out what speech improvements are needed, at least in terms of the parameters of the speech synthesis model.\[16\]

---


In one study aimed at investigating temporal features, Levitt found a reduction in the intelligibility of speech as weighted deviant duration increased. Excessive pause time was combined with excessive prolongation time, using the digital computer.

An overall measure of deviant duration was first derived in inverse proportion to the variation in duration for corresponding segments in the speech of hearing children. Levitt assumed that damage would be most severe if excessive prolongation occurred to those speech and pause segments in a sentence that showed the least variation between normal-hearing children. The data from this investigation indicates a clear trend towards a reduction in intelligibility as the weighted deviant duration increases. Thus, Levitt provides additional quantitative data supporting the view that the prolongation of speech sounds by the deaf is a contributing factor to the intelligibility of the speech.

The merits of the computer in analyzing the speech deficiencies of the deaf are just becoming manifest, however. The question as to the extent to which rhythm in deaf speech is significant has not been completely and satisfactorily answered, and only when further investigations are completed will the problem be resolved.

---


It is difficult to compare the results of this investigation with those of Houde, John and Howarth and Levitt. In the first two studies, training was involved. In Levitt's study, an analysis of the data was done with dissimilar strategies and objectives.

The investigation described in this paper is unique, and the results provide substantial evidence that speech rhythm is important to intelligibility.

Seven out of ten sentences spoken by the deaf subject in this investigation showed an equal or improved intelligibility rating for key contextual words when the rhythm was modified to resemble that of the normal-hearing subject. The improvement in the rhythm patterns of the deaf speech appears to have allowed greater attention to the context.

In addition, when the rhythmic patterns of the normal-hearing subject were modified by computer to resemble those of the deaf subject, the overall intelligibility of key contextual words decreased by 12 per cent.
REVIEW OF THE LITERATURE

The Significance of Timing in Speech Production

Time is a fundamental dimension in speech. Our perception of speech is partially based on our cognitive recollections of the spoken utterances. We expect to hear typical durations, relative durations and natural rates of change in the time dimension of speech. We acquire a rhythmic intuition in terms of these rates. When the normal temporal framework is altered, the speech is likely to be called "unnatural" or "abnormal". This may indicate a lack of perception of this time-distorted speech due to a reaction of this rhythmic intuition. Many writers have investigated the influence of segmental and supra-segmental temporal features on the perception of speech.

On the segmental level, the temporal features of concatenated phonemes have been found experimentally to be significant determining factors in their perception, whether in isolation or in strings. For example, if the brief gap between fricative and vowel in a recording of a word "sag" is increased by 20 or 30 milliseconds, the word sounds clearly like "stag". The difference in one test was just a 10 millisecond increment.


Lisker also found that the durations of some of the silences are crucial. He concludes that although there are several acoustic differences corresponding to the (p) - (b) distinction in the words "rapid" and "rabid", for example, the difference in closure duration appears to be one of the important perceptual cues.\(^{23}\) The difference in the durations was found to be a mere 35 milliseconds.

The durational properties of syllable nuclei and the influence of adjacent segmental sounds (i.e., consonants) upon their duration were studied in another investigation.\(^{24}\) The writers conclude that the durations of all syllable nuclei in English are significantly affected by their environment. Durations of syllable nuclei as a function of adjacent consonant phonemes were plotted and discussed in this paper.

Although Lindblom\(^{25}\) and Rapp\(^{26}\) worked with Swedish syllables, one would expect their findings relative to control of timing in speech to be true for many languages. For example, Lindblom found that the


acoustic duration of an unstressed syllabic segment is approximately constant, but is longer in a final position in a word than in initial or medial positions. His plots show how the segment durations of a syllable initial consonant and phonologically long and short vowels vary as a function of the word length. The vowel and consonant segments decrease in duration as the number of syllables per word increases.

Lindblom used a Swedish nonsense vocabulary with words having one to five syllables. Each word had one main stress and for each word length, the stress occurred in all possible positions. The effects described above were found for all subjects.²⁷

Segmental and suprasegmental timing structures do not operate as separate entities. Newcomb writes that:

The simultaneity of segmental phonemes and prosodemes is possible because, phonetically, the identification of both vowels and consonants is largely a function of the quality or color of the vowel, and thus the pitch, the length and loudness parameters remain to act as the phonetic material of the prosodemes. The development of prosodic theory necessarily followed the development of segmental theory, because it was not until rigorous techniques had been developed and perfected on the segmental level that they could be extended to other systems.²⁸

In this investigation, Newcomb provides linguistic evidence of an influence of sentence context on tempo variation.

²⁷Lindblom, 2-3.

Peterson and Lehiste suggest that when speakers vary their tempo, the durations of stressed syllables are affected. The writers found that duration of stressed words was decreased by a factor of 1.5 when the rate of utterance was increased by a factor of two.29

Thus, the duration of prosodemes, which are suprasegmental (or intonational) phenomena, depends partly on the durational variations of their segmental building blocks and on the physiological and articulatory factors which influence them.

Physiological duration constraints imposed upon larger groups of words, phrases and sentences are primarily respiratory ones. DiCarlo writes:

Speech consists of a series of rapid, highly skilled movements of the breathing and articulatory muscles. The syllabic stream of normal speech tends to be emitted swiftly. The units tend to cluster about the maximum physiological limit. Speech movements of the chest muscles produce the syllable pulse - the fundamental unit of speech - the larger abdominal muscles support the action of the chest musculature producing a series of syllables. These syllables are then fused into a single breath-group, or phrase, on the expiration phase of respiration. The syllables are grouped into rhythmic units (or feet) which, in turn, are grouped into a larger unit, the phrase, by the action of the abdominal muscles. All synergic sequences need timers. So does speech.30

29G. Peterson and I. Lehiste, 699.

Psychosocial conditioning also plays a role in speech timing. Definite durational differences can be seen on the speech spectrograms produced by Stevens and Williams. Wideband spectrograms of the cluster "For God's sake" spoken in five emotional situations were made, and acoustical correlates were found.

What is being shown is that speaking rate may be a function of linguistics, physiology and/or phonology, but, equally important, speech entails psychological and sociological dimensions. DeVito says that the psychological context is perhaps the most difficult to isolate in communication, and involves, for example, "the various attitudinal and motivational states of the communicators as well as their past histories, experiences, and knowledge".

It is appropriate at this point to mention that duration also serves as a cue for semantic disjuncture. The break between words may give different meanings to the spoken sentence, even without the emotional stress described above. An example of semantic influence employed by many writers is given below:

She is a lighthouse keeper.
She is a light housekeeper.

---


In these sentences, the semantic differential is apparent. Oettinger\textsuperscript{33} discusses the "semantic wall" in depth, and gives an interesting analysis of the problem.

The references cited in this discussion illustrate the significance of many factors affecting the time dimension of speech, and consequently, its perception.

\textbf{Temporal Features in the Speech of the Deaf:}

The temporal manifestations in speech production have been recognized by pathologists specializing in the speech of the deaf as important to intelligibility, especially on the suprasegmental level.

In an early discussion of the problem, Stetson wrote that "rhythm has a vital influence on details of pronunciation."\textsuperscript{34} Rhythm at high speed, he pointed out, determines whether syllables will be slurred or pronounced correctly.

Alexander Graham Bell was aware of the importance of rhythm in speech. In 1914, Bell remarked that visitors preferred the "imperfect


gabble" of some semi-mute pupils with fairly natural rhythm to the "elocutionary" speech he had taught other congenitally deaf children. The latter group articulated perfectly but slowly and monotonously. 35

Hudgins and Numbers pointed out that Bell was actually inferring that the phonetic sounds spoken with accuracy are incompatible with the dynamic processes of grouping, accentuation, and subordination of syllables. They further conclude that less emphasis should be placed on the "Elements Method". In this method of teaching speech, sounds are first mastered individually and then combined to form words and phrases. Each sound must have a relative fixed durational qualitative and intensity value before being combined. The writers say that this procedure overlooks the fact that individual sounds must accommodate themselves to the demands of the rhythmic grouping of the syllables in which they occur. 36

Many writers agree with the viewpoint of Hudgins and Numbers on teaching speech. Peterson, for example, believes that typical deaf speech is missing a "vital overlap": when words are spoken as separate entities with excessive pauses between them, there is a lack of smooth transition. The acoustic information of a given phoneme, then, is not


helpful to the recognition of adjacent speech sounds.\textsuperscript{37}

Hudgins and Numbers contend that rhythmical or non-rhythmical utterances affect speech intelligibility as adversely as errors of articulation per se.\textsuperscript{38} The authors assert that the labored, mal-phrased and badly-accented sentences found in the speech of the deaf is one of the most serious defects: "Speech rhythm, as shown by the data, becomes an important factor in speech intelligibility."\textsuperscript{39}

In their study, the speech intelligibility of 192 deaf pupils 8 to 20 years old was evaluated quantitatively with an inventory of 1900 test sentences. The sentences were classified into three types of rhythm categories: those spoken with normal rhythm, those with abnormal rhythm, and those non-rhythmically spoken. The authors' conclusions pertinent to this study are that the rhythmically-correct sentences spoken by deaf pupils have a 3.5 to 1 advantage of being understood over those spoken with incorrect rhythm. Of all of the sentences spoken by the 192 pupils, the 45\% classified as having normal rhythm accounted for 74\% of all the sentences understood by the auditors.

Later investigations supported the findings of Hudgins and Numbers. Calvert did a detailed study of durational characteristics of the speech of the profoundly deaf. He concludes that durational distortion


\textsuperscript{38}C. Hudgins and F. Numbers, 289-392.

\textsuperscript{39}C. Hudgins and F. Numbers, 356.
may contribute to the unnatural sound of deaf speech which leads to the identification of speakers as being deaf and that these distortions may interfere with the intelligibility of deaf speech.  

Considerations of temporal factors in deaf speech must include phoneme duration as well as rate and duration of the total utterance. What has been regarded as characteristic 'voice quality' of the deaf must be a composite of several acoustic cues, including durational distortion of phonemes. This distortion may also affect the intelligibility of deaf speech.

Calvert also noted that normal vowel-consonant ratios are not found in deaf speech. Vowels are extended and the duration values for phonemes in the plosive and fricative classes are consistently extended. For plosives, he found consistent extensions from four to five times the normal duration.

Hood made measurements on the rhythm of deaf speakers, finding that the durations of their sentences were from two to four times longer than sentences spoken by speakers with normal hearing. Listeners rated the speech rhythm of deaf subjects well below the ratings for the normal-hearing speakers. Hood and Dixon write that until a minimum level of rhythm proficiency is attained, intelligibility might improve. The rhythm does not have to be near-normal to be associated with high intelligibility, however. They also conclude that improving a deaf

---

40 D.R. Calvert, p.91.

41 D.R. Calvert, p.94.

person's speech rhythm may make his speech easier and more pleasant, allowing the listener to attend to the context of the speech with less distraction by abnormal rhythm.\footnote{R. Hood and R. Dixon, "Physical Characteristics of Speech Rhythm of Deaf and Normal-Hearing Speakers," \textit{Journal of Communication Disorders}, 2 (1969), 20-28.}

Clarissa Smith, in a recent study with forty congenitally deaf children, also concluded that rate of utterance, as one of the supra-segmental errors observed most consistently, showed significant correlations with intelligibility. No clear pattern emerged in the investigation, however.\footnote{C.R. Smith, "Residual Hearing and Speech Production in Deaf Children," \textit{Communications Sciences Laboratory Research Report Number 4}, (April 15, 1973), 33.}

In still another study, M. J. Osberger produced simulated deaf speech with the typical errors present in isolation or combination. A normal-hearing person mimicked the speech of five deaf children using a masking technique. The general conclusion is that the fricative errors have the greatest effect on intelligibility. She did find some effects on timing: "The data for another child with timing errors showed a slight decrement in intelligibility when the errors in timing were present."\footnote{M.J. Osberger, "Simulation of Deaf Speech as a Diagnostic Tool," Paper presented to the 87th Meeting of the Acoustical Society (April, 1974).}
Voice Quality: A Psychosocial and Psychophysical Dilemma

Many authors write that human communication encompasses complex entities not easily understood. The psychophysical and psychosocial processes by which human beings "communicate" are integrations of attitude, behavior, linguistics and symbolism.

The speech process, just one of the many communication modes, is a form of behavior which employs a symbol system. This symbol system is not a completely vocal one; it is made up of behavior stimuli and responses which include voice and articulation, visible symbols, language and the various cognitive processes.

In this way, we see that speech communication science is an interdisciplinary melange of semantics and linguistics, phonetics and acoustics, psychology and physiology.

The oral speech code is a combination of articulation, vocalization and gesture. Through voice, a speaker reveals how he feels about what he is saying. His intonation and stress suggest thoughts which are not readily apparent when these inflectional changes are absent.

Moreover, articulation and gestures are physical mannerisms which, with vocalization, provide man with the ability to "communicate" with his neighbor, and to live in a happy co-existence with those who "speak his language". Consequently, a person deficient in the oral speech code will find himself isolated from his neighbor; and the isolation is a lonely one indeed.

The speech chain is like any communication channel and has, in its operative state, an encoder-decoder pair and a medium. Render one of the terminal pair inefficient, and the entire system is inefficient. Analogously, this is the psychophysical dilemma of the profoundly deaf. The isolation is even more intense, however, than that of a deficient speaker, since the deaf are unable to receive the acoustic speech signal and, thus, are deprived of auditory feedback advantages.

Some of the oral speech deficiencies of the profoundly deaf other than faulty timing should be mentioned briefly at this point.

Voice quality of profoundly deaf individuals is distinctive. Many writers agree that it is unique from that of individuals with normal hearing, and characterized by a lack of control of the vocal cords and surrounding musculature. Silverman notes that this unusual quality is associated with a hearing loss of 45 decibels or greater. Other characteristics are incorrect positioning of the articulators, excessive nasality, breathiness, poor pitch control and poor loudness coordination.

Colton and Cooker assert that the most frequent quality deviation is excessive nasality, and attribute the nasalization to lack of velar control. 48

Breathiness, attributed to the large flow of air through the vocal cords held too wide during phonation, is a significant contributing factor to the duration problems of deaf speakers. Hudgins found that the length of the phrases uttered in one breath is shortened as the rate of breath expended increases. 49

Uncontrollable pitch variations, falsetto voice, and monotone pitch are three characteristics related to improper coordination of the larynx. 50 A deaf speaker often attempts to raise his pitch by varying the loudness of his voice. Thus, both pitch and loudness control are not independently achieved without difficulty, and the necessary laryngeal modifications are often substituted with respiratory ones.

Malarticulation of consonant and fricative sounds is an especially serious problem. Dynamic features of consonant production are often distorted. The transitions from consonant articulatory configurations to vowels, for example, may be incorrect. Topological features are also abnormal (i.e., incorrect articulatory configurations).


Harris, et al, employing the technique of surface electromyography (suction electrodes), conclude that the intelligibility of the hearing-impaired talkers cannot be accounted for on a topological basis only. Topological speech habits (positioning of articulators) were not found to be inherently unsatisfactory on the basis of articulatory models observed visually:

Third, deaf speakers do acquire regular, distinctive habits of articulation, even with respect to tongue posture, though the habits may be 'wrong'. From the therapeutic point of view, we believe that this means that the topological characteristics of deaf articulation are probably improvable with supplementary visual information about tongue position...only if we assume that 'visibility' is the important reason for the apparent superiority of the lip aspects of articulation. Another possibility is that there is somehow a hierarchy of difficulty of various aspects of enunciation, and that the tongue gestures are inherently more complex. 51

The authors checked this by examining in a similar manner the articulation of a group of dysarthric speakers, and found a sharp contrast. The dysarthric subjects produced unintelligible consonantal utterances (poor motor control), whereas the vowels were little impaired, if at all. 52

From this discovery the writers assert that the deaf produce perceptually and myographically acceptable consonants with respect to lip movements and stereotyped, though inappropriate, vowels. The authors


52 K. Harris, et al, 145.
wondered why deaf speakers produce such poor vowels, hypothesizing that normal-hearing subjects rely more heavily on auditory feedback for consonants.\textsuperscript{53}

The voice quality characteristics of the deaf often deter them from making communicative overtures with the "hearing world". Even the postlingually deaf find themselves sometimes discouraged by the manifestation of these distinctive quality characteristics inherent in deaf speech.

There is no doubt that any attempt to improve this quality through applied research will be one step forward in ameliorating the attitudinal and motivational burdens of the deaf.

\textsuperscript{53}K. Harris, et al, 146.
DESCRIPTION OF THE ANALYSIS-SYNTHESIS SYSTEM

Speech Processing: A Tandem Chain of Operations

The speech signal processing in this investigation utilizes a tandem chain of digital computer operations to analyze and synthesize human utterances. Four computer algorithms make up the heart of the system (Table 1). All are run on a Xerox Sigma Six computer at the Rochester Institute of Technology.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECTRUM</td>
<td>Computes Fast Fourier transforms, produces 10 millisecond short-time spectra</td>
</tr>
<tr>
<td>SEGMENT</td>
<td>Assists in locating syllable boundaries, measures durations of speech and pause segments</td>
</tr>
<tr>
<td>TEMPO</td>
<td>Realigns the rhythm of a sentence by expanding or condensing individual segments</td>
</tr>
<tr>
<td>ENGSYNTH</td>
<td>Transforms spectral data back to digital data (inverse Fourier transforms)</td>
</tr>
</tbody>
</table>

TABLE 1 Digital Computer Algorithms used in the investigation and their functions.

SPECTRUM and ENGSYNTH were originally written by engineers at the Center for Communications Research, 50 W. Main St., Rochester, New York, and later modified for use in this project. SEGMENT and TEMPO were written by the author.
Several additional computer programs were supplied by the department of Electrical Engineering at Rochester Institute of Technology to run the analog-to-digital and digital-to-analog processes on a Hewlett-Packard 2116-B mini-computer.

Block diagrams of the analysis and synthesis systems are shown in Figures 6 and 7.

---

**Figure 6** THE ANALYSIS SYSTEM
Speech was first recorded on an analog tape. These speech samples were carefully identified as to speaker and the exact phrase recorded. The recording was done in a chamber with suitable acoustics.

The speech was then passed through passband filters with cutoff frequencies of 20 Hz and 4000 Hz, and an analog-to-digital converter. The spectral response was 20 db down at 5000 Hz.

The A/D process was done in real time on an HP-2116B mini-computer. The speech was digitized with a rate of 10000 samples per second and quantized to 10 bits of accuracy. A quantizing operation was employed which produced a discrete set of amplitude levels. Any sample falling
between the allowed levels was converted into a sample of amplitude nearest the actual value of the signal. Besides sampling and quantizing, the analog-to-digital conversion involved binary encoding. The samples were stored on a magnetic tape in the digital mode.

Fast Fourier transform scientific subroutine packages HARM and RHARM were then used to compute a sequence of short-time spectra. The spectra contained all the information in the original wave except for phase relationships, and this was not a serious loss assuming the ear is rather insensitive to phase. A single spectrum was computed for each 10 milliseconds of speech and a total of 1024 samples were produced in the frequency domain. Seven adjacent 20 Hz samples were combined to produce thirty-six channel outputs, each 140 Hz wide. The speech wave, in effect, was processed through a bank of thirty-six bandpass filters.

The spectrum data from the filtering program was recorded on another magnetic tape for the final speech signal analysis. This data appeared on a computer-generated spectrogram very much like a sonagram, but with the relative power given by printed values rather than by the darkness of the spectral traces. Thus, a measure of relative power could be made if desired. In addition to the short-time spectra, information on fundamental frequency, voicing and overall power was printed for each 10 milliseconds of speech (Figures 1, 2 and 3 on pages 5-6).
Rhythm Analysis

Three main processes were involved in the rhythm analysis: Segmentation of the speech, duration measurements of the segments, and tempo modification.

Segmentation of the speech of both the deaf and normal-hearing subjects involved a careful examination of the three-dimensional speech spectrum. The purpose of segmenting the sounds was to measure the duration of each syllable and compare this data with corresponding measurements for both subjects. Pauses were also measured and compared. With these timing measurements, the relative durations of the spoken utterances were computed, and the data was used in the third process of tempo modification. For each sentence in the inventory, the entire suprasegmental feature of timing was rebuilt by a computer algorithm, which retimed each individual syllable or pause segment.

These three processes are described in depth on the following pages.

Segmentation - The First Process

The speech continuum was segmented into sub-syllable-length segments. Originally, it was planned to segment on the syllable level; but this was found to be inadequate. The process of retiming by deleting certain short-time spectral segments in proportion to deviant duration of the syllable gave equal weight to the consonantal and vowel portions of the utterance.
When syllable segmentation was actually attempted, the CVC type segments produced after the deletion were frequently deficient in consonantal information near the boundaries.

This was especially the case for plosives, more than for fricatives which usually have a larger temporal framework. Characteristically, plosives are sudden bursts of energy throughout the frequency domain in a very short time interval. They often show up as "spikes" on speech spectrograms.

Consequently, if one of the short-time spectra chosen for deletion by the computer contained a spike, practically all of the plosive would be eliminated. The resultant speech wave, when heard by a listener, would seem largely vowel in character.

This problem was greatly improved by segmenting on a subsyllable level so that each consonant and each vowel were retimed in proportion to their durational deviance.

Two methods were used to segment the speech wave. Segmentation of the computer-generated spectra employed an analysis of several acoustic parameters and their trends. This data was compared to manual measurements done directly on Kay Electric Company sonagrams.

Since the results of two experimental methods were being compared, the percentage difference was computed. This permitted greater insight into the precision of each method. The percentage difference was in terms of segment lengths.

The segment boundaries were chosen visually, using wideband spectrograms, which have better time resolution than narrowband spectrograms. The percentage difference was usually less than 5 percent.
for segments with plosive type boundaries. These types of boundaries required the greatest accuracy since they were extremely short in duration.

The difference was never greater than 10 percent for segments with fricative, vowel, nasal or glide characteristics at the boundaries; but the errors were negligible since these speech sounds had a large temporal framework.

The two most apparent factors influencing the segmentation accuracy for the fricative, vowel, nasal and glide sounds were the intensity of the spark from the Kay Sonagraph power source and, in the case of the computer-generated spectrograms, the scaling factor which was used. In both displays, the intensity of the spectral "trace" tapered off toward the boundaries where there existed regions of vague discontinuity, especially when a segment was adjacent to interior or intra word silence.

In the computer segmentation, one parameter used (a voicing decision) was based on the detection of the presence of a larynx tone. Different codes were used for unvoiced speech and silence. Thus, a relatively accurate estimate of a boundary could be made using this voicing criteria alone.

The greatest difference between the segmentation data for the two techniques occurred when frication was extended or when the deaf speaker's nasality was prolonged. These discrepancies in the data can be attributed to the four kilohertz cutoff frequency. The filter response being down 20 db at 5 kilohertz precluded high resolution in
the upper frequency region (Figure 8).

![Frequency Response Curve](image)

Figure 8  Frequency Response Curve
Thus, a compromise was made between signal aliasing and the spectral resolution of fricative-like speech. With a Nyquist rate of 10,000 samples per second, aliasing was negligible.

The effect of the compromise was apparent in the duration measurements of syllables having fricative boundaries. These syllables were usually longer when measured on the sonagrams.

Difficulty was also encountered when segmenting at the center of a glide such as the "l" sound in "umbrella" since the acoustic cues gave no clear indication for the glide segmentation.
Because the sounds which are characteristically long in time are those with greater differences in segmentation boundary measurements, the impact is negligible. The discrepancies will not have a marked effect on the intelligibility during the rhythm modification.

In addition to these sources of difference, the deaf subject's speech wave was distorted by aspiration and "swallowing sounds" at some of the boundaries.

It is difficult to recognize clearly the acoustic cues of many speech sounds spoken by the profoundly deaf subject. In the actual speech events, the discreteness of the spectral segments on the phoneme level is blurred or totally obliterated by breathiness, nasalization, malarticulation and guttural sounds; so that it is often difficult to isolate the segments and their features. Accurate segmentation on a sub-syllable level is a formidable task which still eludes the most capable engineers working with the speech of normal-hearing subjects. It is widely recognized that there is little one-to-one correspondence between speech segment features and particular physical events. Rather, descriptions of these features are abstract representations of classes of events.

In one study reported in the literature, Martony found similar results of segmentation. Using the method of Fant and Lindblom,


Martony segmented the speech of deaf subjects in an attempt to specify the deviation from the speech of normal-hearing subjects on an objective level. The comparison, using sonagrams, revealed the main difficulty in phonation and the lack of synchronization between articulation and phonation. Nasalization and denasalization errors were observed. The "super-stationary" form of vowels, absent or too-short transitions of formants, indicated the abruptness of articulatory movements. Errors in speech rhythm were not studied in this investigation.

The spectral distortion of the speech of the deaf makes many segmentation schemes difficult to utilize. Much sophistication has been incorporated into segmenting speech, as shown by the literature. In the present study, however, all that is necessary is to determine visually the segment lengths with the help of the data from computer measurements of the acoustic parameters.

In the computer-based segmentation scheme used in this study, two kinds of decisions were performed on the acoustic data. One type of decision was an examination of the properties found in a single 10 millisecond spectral segment. For each short-time spectrum, the analysis was done over the entire frequency domain with a range of 4000 Hz. This range is sufficient since although speech extends from 100 Hz to 8000 Hz, most of the energy is found below 3000 Hz. This

---

57 See appendix B for a brief description of some of the significant segmentation attempts.
decision process was performed on each of the short-time segments. Initial and final components of a larger utterance were specified by identification indices.

A second kind of decision process required several short-time segments or knowledge of a previous decision performed on the data. This was used for comparison of two or more short-time spectra, such as when searching for spectral change.

The parameters used to identify the segment boundaries were the following:

1) Overall power
2) Spectral slope change
3) Formant transitions
4) Pitch change
5) Voice/voicelessness information

A short discussion of these parameters follows:

**Overall Power**

As a first step in the segmentation, the general region of a boundary was automatically detected using a routine to search the overall power trends. As these regions were located, data on pitch change, voicing change and a rough estimate of the duration of the last segment were output with the spectra as shown in Figure 9.
The overall power trend is useful in two ways. First, it can be used to single out the approximate region of a boundary, often to within 10 milliseconds. Secondly, the power trend may sometimes be used as a criterion for deciding on the best of two possible locations for a boundary, especially when the remaining parameters fail.
Spectral Slope Change

When a single filter output in a time $t_2$ is greater than the two adjacent power values for the same filter for times $t_1$ and $t_3$ as shown in Figure 10, the power value is called a "peak". The sum of all of the peaks in one 10 millisecond spectrum is an indication of the amount of spectral energy change occurring at that instant corresponding to transients in the speech apparatus. It is not a spectral change function (e.g., $|t_2 - t_1|$ for all filter values), but only a measure of positive-to-negative slope change for each 10 milliseconds.

Figure 10 Spectral Change. The syllable "bird" spoken by the deaf subject. The number of peaks in time $t_2$ equals nine.
Figure 11 Spectral Change as a function of time. The word "twenty" spoken by a normal-hearing person.

Figure 11 shows the number of filter output peaks as a function of time for a speech sample spoken by a normal-hearing subject. This data on spectral slope change was printed for inspection. The data was helpful in choosing one of several short-time spectra as a segment boundary at times when the selection was vital (e.g., keeping the entire initial plosive consonant within a segment).
Formant Transitions

Transitions in F1 have been correlated with syllable boundaries. F1 usually dips and subsequently rises at a syllable junction. To study transitions in the formants, a transparent template was used with the sonagrams. The transitions were also examined on the computer-generated spectrograms although there was less frequency resolution. Formant transitions were especially helpful in segmenting syllable type segments containing vowels or glides.

For deaf speakers, however, formants are often flat due to uncoordinated articulation and nasality associated with steady-state resonance. Figure 12 illustrates an output of part of the program SEGMENT used to study vocal tract resonance.

---

**Fundamental Frequency**

The vocal pitch often changes abruptly when a new speech sound is produced (e.g., in the word "/s/a/t/"). With pitch problems being common in the speech of the prelingually deaf, however, this particular parameter was not a dependable criterion.

**Voicing**

As mentioned previously, the presence or absence of a larynx tone was represented by a code (phonation=1, unvoiced speech=2, silence=0). This data was extremely important in the selection of the segment junctions. The voice/voicelessness decision was based on the concentration of energy in the short-time spectral segment and the location of this energy in the frequency domain.

**Duration Measurements – The Second Process**

In the second process of the tempo-modification scheme, the duration of each segment was automatically measured by computer. Besides measuring the individual segment lengths for each subject, two quantities were derived from the data.

The deviant duration of each syllable was computed by subtracting the lengths of the segments. A positive number indicates prolonged speech by the deaf subject as was the case 90% of the time.
In addition, a prolongation factor was computed as a ratio of the length of the segment spoken by the deaf subject divided by the length of the same segment spoken by the normal-hearing subject. The prolongation factor presented a clear picture of the distortions of speech and pause segments in the time domain by the profoundly deaf subject.

Both the deviant duration and the prolongation factor were used to determine the output durations in the rhythm modification algorithm.

Duration data is shown in Tables 4-7 on pages 61 to 64.

The duration of each syllable was also measured on sonagrams and a comparison was made with this data and that produced by the computer.

Rhythm Modification - The Third Process

The process of rhythm modification involved the actual compression and expansion of speech and pause segments to new durations. The deviant duration and prolongation data obtained by segmentation and timing measurements determined whether a segment was expanded or condensed in the time domain.

The main point of investigation in this process was one of acoustical energy deletion and its significance to perception of individual sounds.

The retiming process must be consistent in segmenting similar speech sounds. The quality of the retimed speech appeared to be wholly dependent on the segment type. Consonant-Vowel-Consonant (CVC) type segments, for example, were not processed as efficiently as phonemic-type segments.
The method of energy deletion chosen was one based on the relative lengths of the segments spoken by the normal-hearing and profoundly deaf subjects. Working from the center of the segment outward in both directions toward the segment boundaries, ten millisecond short-time spectra were deleted. The number of deletions was proportional to the prolongation factor.

A uniform deletion of energy on both sides of the segment midpoint was first performed.

Any remaining number of short-time spectra to be deleted after the first pass was also done by working from the center of the segment.

In this way, the acoustical parameter trends in the segment are not distorted significantly. Uniform deletion of the spectral energy will maintain the formant structure, the rising and falling trends in fundamental frequency and the relative power throughout the segment.

To illustrate the tempo-modification process, the segment "A" in sentence # 563 is selected from the data for discussion. The segment spoken by the deaf subject is 480 milliseconds long, and is to be shortened to 130 milliseconds, the duration of the same syllable as spoken by the normal-hearing subject.

The prolongation factor is 3.69 and the deviant duration is 350 milliseconds. Thirty-five, 10 millisecond short-time spectra would be deleted from the digital tape. A new printout would be generated showing the tempo-modified speech which represents the newly arranged spectral data.
The identification indices are 1832 and 1879 for the first and last short-time spectra, respectively (Figure 13). The center of the segment is at 1855.

Figure 13. Computer Construction of a New Speech Segment.

There are approximately 24 spectra on each side of the center in the original utterance. The algorithm selects, on the first pass, the first, last and center spectra as a base for building the new spectral profile, as shown in Figure 13.

On the second pass, every fourth spectra is selected since the prolongation factor is almost four. Energy is selected uniformly on
both sides of the center segment until the desired duration is accumulated in the new profile bank.

For this particular example, exactly five spectra were selected on both sides, in addition to the three spectra chosen on the first pass. In the event that there is a fractional remainder after the second scanning, the computer calculates a new incremener and a third pass balances the spectral profile to the desired duration. The computer-generated spectrograms for this segment are shown in Figure 14.

Figure 14 Computer-generated spectrograms illustrating the retiming of the syllable "A". Original segment is at left.
The prolongation factor was usually between 1.00 and 3.00 for speech segments. Inter- and intra-word pauses were also tempo-modified using this algorithm.

The lack of sensitivity of the retiming process to segment features required a much more delicate segmentation scheme. When CVC type segments were processed, for example, much of the important acoustic information was lost because of failure to discriminate phonemic boundaries within syllables. The best results were obtained by processing individual sounds.

One may question whether it is possible to accurately segment individual sounds as described in this paper. As explained in the section on segmentation, many acoustic parameters were used to assist in the boundary selections. The care taken in segmenting rendered the effects of errors on intelligibility negligible.

The segmentation, duration measurement and retiming processes, working together, accurately produced new segment lengths which were representative of the general acoustic parameter trends in the original utterances.

Although the trends were accurately represented in the tempo-modified segments, the effects of energy deletion on intelligibility were possibly of significance. This is presented in depth in the discussion of the results on pages 86-91.
Figures 15 and 16 also illustrate how a speech segment is expanded and condensed by the tempo-modification algorithm. The identification indices of the short-time spectral components are at the top. In the output sample in Figure 16, the deleted spectra are apparent by the missing indices.

**Figure 15** Rhythm Modification. The syllable "brown" expanded.

**Figure 16** The tempo-modified version of the phrase "in a big". The original total duration for this phrase was twice as long as the new version.
Synthesis

In the synthesis stage of the system, the short-time spectral information was transformed back into digital data, run through a digital-to-analog converter and played through a loudspeaker. Tape recordings of the retimed speech were made and sonagrams were produced for comparison and evaluation.
EXPERIMENTAL PROCEDURE

Preliminary System Test

A preliminary run was performed with normal speech to examine the effects produced by the modification of spectral data. A test sentence was selected from an inventory designed for intelligibility tests.\(^{59}\)

TEST SENTENCE: A MAN PUT SOME BREAD IN A BIG BROWN BAG.

Recordings were made of this sentence using a Sony high fidelity tape recorder and low noise magnetic tape. The sentence was spoken by a professional linguist, first with normal rhythm patterns and then with time distortions.

The synthesized versions were compared to the originals and the quality, intelligibility and naturalness were evaluated. It was during this pilot study that a need for sub-syllable segmentation was demonstrated.

The computer program TEMPO was found to uniformly and consistently retime the segments in proportion to the prolongation factor for each utterance.

Selection of the Speaker

As shown by the audiogram in Figure 17, the subject chosen for the study has a profound bilateral hearing loss. The twenty-year old male student was born deaf, cause unknown, and seldom wears a hearing aid.

Figure 17 Audiogram of deaf subject
His parents have no hearing impairment and speak General American English in the home.

The important criterion for selection of this subject was his obviously inappropriate speaking rate and abnormal rhythm characteristics. In one study of rate production, the deaf subject's rate was 1.94 times longer than that of a normal hearing subject. His average syllable rate was 1.31 syllables per second. The rate of the normal-hearing speaker was 2.55 syllables per second.

When speaking the "Rainbow Passage", the subject exhibited an average rate of production of 1.47 syllables per second. The total time of recitation was 86 seconds. A normal-hearing speaker performed the same task in 53 seconds with an average rate of 2.39 syllables per second.

---

A computer spectrogram of the word "the" spoken by the deaf subject is shown in Figure 18. The word spoken in the same sentence by a normal-hearing subject is shown in Figure 19 for a durational comparison.

Figure 18 First twenty-eight filter outputs (20 Hz - 3920 Hz) for the word "the" spoken by the profoundly deaf subject.
Figure 20  Temporal distortion of the word "AT" spoken by the deaf subject. The interword pause of 30–40 milliseconds separates the prolonged vowel from the plosive terminal phoneme.

Figure 20 illustrates temporal distortion of another word by the deaf subject.
The deaf subject's speech and hearing communication profiles are described below in Table 2:

1) Rating of I on Hearing Discrimination (I to V scale):

Profile descriptor: The student is able to recognize only 0-15 percent of the items in a same-difference test with two-syllable words.

2) Rating of III of the Speech Profile (I to V scale):

Profile descriptor: Speech is difficult to understand; however, the gist of the content can be understood.

Table 2 Communication Profile for the Deaf Subject

This information was categorized by professional speech pathologists at the National Technical Institute for the Deaf's Communication Center.

Sentence Inventory

The deaf speaker was asked to read ten sentences, each having ten syllables. The sentences were selected from an inventory designed by the Clarke School for the Deaf. The subject was permitted to practice each sentence at least once to familiarize himself with the words and to minimize reading errors. The recordings were done in a quiet chamber.

---

61 M. Magne, "A Speech Intelligibility Test for Deaf Children".
The sentences in the inventory are given in Table 3:

561  A BROWN BIRD SAT ON A FENCE NEAR A HOUSE.
562  TOM PUT SOME OLD SHOES IN A WASTEBASKET.
563  A GIRL LOST AN UMBRELLA AT THE ZOO.
564  A MAN TOOK SOME BOOKS TO THE HOSPITAL.
565  A POLICEMAN WALKED TO THE SCHOOL PICNIC.
566  A COWBOY WENT SHOPPING IN THE COUNTRY.
567  A BABY PUT A BALL IN A BIG SHOE.
568  A DOG PLAYED WITH A WHITE LAMB IN THE PARK.
569  MISS BROWN WANTED SOME TEA AT THE MEETING.
570  TWO GIRLS PLAYED NEAR THE RIVER ONE HOT DAY.

Table 3: Sentences selected from the Clarke School Inventory

These sentences were also recorded for a normal-hearing male. The duration measurements for speech and pause times of the normal-hearing subject were then used as reference lengths in the process of rhythm modification.

The syllable lengths for both speakers are shown in Tables 4 through 7.

The deviant duration and prolongation factors are also shown for each syllable. A positive number indicates prolongation and a negative number represents a syllable duration less than the reference length.
<table>
<thead>
<tr>
<th>RUN</th>
<th>SYLLABLE LENGTH</th>
<th>SYLLABLE LENGTH</th>
<th>REFERENCE DURATION</th>
<th>DEVIANT DURATION</th>
<th>PROLONGATION FACTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DEAF SPEAKER</td>
<td>HEARING SPEAKER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>510</td>
<td>110</td>
<td>400</td>
<td>-30</td>
<td>4.64</td>
</tr>
<tr>
<td>PAUSE</td>
<td>.70</td>
<td></td>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td>BROWN</td>
<td>920</td>
<td>460</td>
<td>460</td>
<td></td>
<td>7.00</td>
</tr>
<tr>
<td>PAUSE</td>
<td>210</td>
<td>120</td>
<td>90</td>
<td></td>
<td>1.75</td>
</tr>
<tr>
<td>BIR0</td>
<td>570</td>
<td>390</td>
<td>180</td>
<td></td>
<td>1.46</td>
</tr>
<tr>
<td>PAUSE</td>
<td>340</td>
<td>220</td>
<td>120</td>
<td></td>
<td>1.55</td>
</tr>
<tr>
<td>SAT</td>
<td>640</td>
<td>560</td>
<td>60</td>
<td></td>
<td>1.84</td>
</tr>
<tr>
<td>PAUSE</td>
<td>150</td>
<td>290</td>
<td>-140</td>
<td></td>
<td>1.52</td>
</tr>
<tr>
<td>DN</td>
<td>310</td>
<td>210</td>
<td>100</td>
<td></td>
<td>1.48</td>
</tr>
<tr>
<td>PAUSE</td>
<td>20</td>
<td>40</td>
<td>-20</td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>A</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>PAUSE</td>
<td>20</td>
<td>180</td>
<td>-160</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>FENCE</td>
<td>480</td>
<td>470</td>
<td>10</td>
<td></td>
<td>1.02</td>
</tr>
<tr>
<td>PAUSE</td>
<td>1280</td>
<td>220</td>
<td>1060</td>
<td></td>
<td>5.62</td>
</tr>
<tr>
<td>NEAR</td>
<td>530</td>
<td>350</td>
<td>180</td>
<td></td>
<td>1.51</td>
</tr>
<tr>
<td>PAUSE</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td>A</td>
<td>410</td>
<td>100</td>
<td>310</td>
<td></td>
<td>4.10</td>
</tr>
<tr>
<td>PAUSE</td>
<td>60</td>
<td>20</td>
<td>40</td>
<td></td>
<td>3.00</td>
</tr>
<tr>
<td>HOUSE</td>
<td>560</td>
<td>670</td>
<td>-110</td>
<td></td>
<td>.84</td>
</tr>
</tbody>
</table>

| TCM  | 640             | 470             | 170               |                 | 1.36                |
| PAUSE | 420            | 110             | 310               |                 | 3.82                |
| PUT  | 340             | 170             | 170               |                 | 2.00                |
| PAUSE | 300            | 130             | 170               |                 | 2.31                |
| SOME | 680             | 310             | 370               |                 | 2.49                |
| PAUSE | 220            | 30              | 190               |                 | 1.53                |
| OLD  | 660             | 370             | 290               |                 | 1.78                |
| PAUSE | 820            | 730             | -10               |                 | 27.33               |
| SHOES| 730             | 730             | -10               |                 | .99                 |
| PAUSE | 580            | 120             | 460               |                 | 4.83                |
| IN   | 160             | 160             | 0                 |                 | 1.00                |
| PAUSE | 20             | 20              | 0                 |                 | 1.00                |
| A    | 520             | 140             | 380               |                 | 3.71                |
| PAUSE | 290            | 30              | 280               |                 | 9.67                |
| WASTE | 540            | 320             | 220               |                 | 4.69                |
| PAUSE | 470            | 470             | 360               |                 | 4.27                |
| BAS  | 370             | 330             | 40                |                 | 4.12                |
| PAUSE | 80             | 70              | 10                |                 | 1.14                |
| KET  | 430             | 80              | 350               |                 | 5.38                |

| A    | 480             | 130             | 350               |                 | 3.69                |
| PAUSE | 190            | 90              | 100               |                 | 2.11                |
| GIRL | 680             | 450             | 230               |                 | 1.51                |
| PAUSE | 490            | 20              | 470               |                 | 24.50               |
| LOST | 840             | 610             | 230               |                 | 1.38                |
| PAUSE | 510            | 70              | 440               |                 | 7.29                |
| AN   | 510             | 80              | 430               |                 | 6.38                |
| PAUSE | 360            | 20              | 340               |                 | 18.00               |
| UM   | 430             | 450             | 260               |                 | 4.96                |
| PAUSE | 90             | 70              | 20                |                 | 1.29                |
| BRE  | 360             | 260             | 100               |                 | 1.38                |
| LA   | 340             | 150             | 190               |                 | 2.27                |
| PAUSE | 690            | 410             | 260               |                 | 1.68                |
| AT   | 330             | 140             | 190               |                 | 2.36                |
| PAUSE | 120            | 150             | -30               |                 | .80                 |
| THE  | 510             | 130             | 380               |                 | 3.92                |
| PAUSE | 400            | 10              | 390               |                 | 40.00               |
| ZOO  | 720             | 480             | 240               |                 | 1.50                |

Table 4 Duration Measurements for Test Sentences 561, 562 and 563
<table>
<thead>
<tr>
<th>Word</th>
<th>Duration</th>
<th>Measure</th>
<th>Duration</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>430</td>
<td>140</td>
<td>290</td>
<td>3.07</td>
</tr>
<tr>
<td>PAUSE</td>
<td>130</td>
<td>110</td>
<td>20</td>
<td>1.18</td>
</tr>
<tr>
<td>THE</td>
<td>180</td>
<td>230</td>
<td>-50</td>
<td>.78</td>
</tr>
<tr>
<td>LICE</td>
<td>550</td>
<td>330</td>
<td>220</td>
<td>1.67</td>
</tr>
<tr>
<td>MAN</td>
<td>480</td>
<td>340</td>
<td>-190</td>
<td>1.41</td>
</tr>
<tr>
<td>PAUSE</td>
<td>310</td>
<td>500</td>
<td>-190</td>
<td>-62</td>
</tr>
<tr>
<td>WALKED</td>
<td>470</td>
<td>330</td>
<td>140</td>
<td>1.42</td>
</tr>
<tr>
<td>THE</td>
<td>890</td>
<td>260</td>
<td>630</td>
<td>3.43</td>
</tr>
<tr>
<td>PAUSE</td>
<td>330</td>
<td>160</td>
<td>170</td>
<td>2.06</td>
</tr>
<tr>
<td>TAL</td>
<td>60</td>
<td>10</td>
<td>50</td>
<td>6.00</td>
</tr>
<tr>
<td>PAUSE</td>
<td>430</td>
<td>140</td>
<td>290</td>
<td>3.07</td>
</tr>
<tr>
<td>THE</td>
<td>1100</td>
<td>10</td>
<td>1090</td>
<td>110.00</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>780</td>
<td>590</td>
<td>190</td>
<td>1.32</td>
</tr>
<tr>
<td>PAUSE</td>
<td>550</td>
<td>130</td>
<td>420</td>
<td>4.23</td>
</tr>
<tr>
<td>PIC</td>
<td>350</td>
<td>140</td>
<td>210</td>
<td>2.20</td>
</tr>
<tr>
<td>PAUSE</td>
<td>70</td>
<td>170</td>
<td>-100</td>
<td>2.41</td>
</tr>
<tr>
<td>NIC</td>
<td>180</td>
<td>130</td>
<td>50</td>
<td>1.38</td>
</tr>
</tbody>
</table>

Table 5 Duration Measurements for Test Sentences 564 and 565
<table>
<thead>
<tr>
<th>RUN SYLLABLE LENGTH</th>
<th>SYLLABLE DEAF SPEAKER</th>
<th>REFERENCE SYLLABLE DURATION</th>
<th>DEVIANT SYLLABLE DURATION</th>
<th>PROLONGATION FACTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A PAUSE</td>
<td>400</td>
<td>60</td>
<td>340</td>
<td>6.67</td>
</tr>
<tr>
<td>COW</td>
<td>550</td>
<td>290</td>
<td>260</td>
<td>1.90</td>
</tr>
<tr>
<td>PAUSE</td>
<td>50</td>
<td>40</td>
<td>40</td>
<td>2.36</td>
</tr>
<tr>
<td>BGY</td>
<td>560</td>
<td>390</td>
<td>170</td>
<td>1.64</td>
</tr>
<tr>
<td>PAUSE</td>
<td>1560</td>
<td>640</td>
<td>920</td>
<td>2.44</td>
</tr>
<tr>
<td>WENT</td>
<td>300</td>
<td>130</td>
<td>170</td>
<td>2.31</td>
</tr>
<tr>
<td>PAUSE</td>
<td>960</td>
<td>60</td>
<td>900</td>
<td>16.00</td>
</tr>
<tr>
<td>SHOP</td>
<td>680</td>
<td>400</td>
<td>260</td>
<td>1.70</td>
</tr>
<tr>
<td>PAUSE</td>
<td>20</td>
<td>90</td>
<td>-70</td>
<td>2.2</td>
</tr>
<tr>
<td>PING</td>
<td>430</td>
<td>450</td>
<td>-20</td>
<td>2.6</td>
</tr>
<tr>
<td>PAUSE</td>
<td>470</td>
<td>470</td>
<td>0</td>
<td>1.90</td>
</tr>
<tr>
<td>IN</td>
<td>290</td>
<td>190</td>
<td>100</td>
<td>1.53</td>
</tr>
<tr>
<td>PAUSE</td>
<td>110</td>
<td>40</td>
<td>70</td>
<td>2.75</td>
</tr>
<tr>
<td>THE</td>
<td>500</td>
<td>80</td>
<td>420</td>
<td>6.25</td>
</tr>
<tr>
<td>PAUSE</td>
<td>160</td>
<td>80</td>
<td>80</td>
<td>2.00</td>
</tr>
<tr>
<td>COUN</td>
<td>600</td>
<td>220</td>
<td>380</td>
<td>2.73</td>
</tr>
<tr>
<td>PAUSE</td>
<td>70</td>
<td>50</td>
<td>20</td>
<td>1.40</td>
</tr>
<tr>
<td>TRY</td>
<td>600</td>
<td>200</td>
<td>400</td>
<td>3.00</td>
</tr>
</tbody>
</table>

| A PAUSE              | 440                    | 160                         | 280                       | 2.75                |
| BA                  | 240                    | 100                         | 140                       | 2.40                |
| BY                  | 430                    | 260                         | 170                       | 1.65                |
| PAUSE                | 150                    | 50                          | 100                       | 2.90                |
| BY                  | 330                    | 190                         | 140                       | 1.74                |
| PAUSE                | 310                    | 500                         | -190                      | 1.62                |
| PUT                 | 290                    | 210                         | 80                        | 1.38                |
| PAUSE                | 360                    | 20                          | 340                       | 18.00               |
| A                   | 440                    | 70                          | 370                       | 6.29                |
| PAUSE                | 130                    | 150                         | -20                       | 0.87                |
| BALL                | 770                    | 430                         | 340                       | 1.79                |
| PAUSE                | 280                    | 650                         | -40                       | 3.38                |
| IN                  | 340                    | 90                          | 250                       | 3.78                |
| PAUSE                | 60                     | 20                          | 40                        | 3.00                |
| A                   | 290                    | 290                         | 0                         | 1.00                |
| PAUSE                | 170                    | 170                         | 0                         | 1.00                |
| BIG                  | 230                    | 140                         | 90                        | 1.64                |
| PAUSE                | 470                    | 140                         | 330                       | 3.36                |
| SHOE                | 790                    | 560                         | 230                       | 1.41                |

Table 6 Duration Measurements for Test Sentences 566, 567 and 568
<table>
<thead>
<tr>
<th></th>
<th>MISS</th>
<th>PAUSE</th>
<th>BROWN</th>
<th>WANT</th>
<th>PAUSE</th>
<th>ED</th>
<th>PAUSE</th>
<th>SOME</th>
<th>PAUSE</th>
<th>TEA</th>
<th>PAUSE</th>
<th>RT</th>
<th>PAUSE</th>
<th>THE</th>
<th>PAUSE</th>
<th>MEET</th>
<th>JING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>440</td>
<td>230</td>
<td>210</td>
<td>1.91</td>
<td>250</td>
<td>130</td>
<td>120</td>
<td>1.92</td>
<td>850</td>
<td>540</td>
<td>310</td>
<td>1.57</td>
<td>390</td>
<td>500</td>
<td>-110</td>
<td>1.78</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>20</td>
<td>160</td>
<td>9.00</td>
<td>60</td>
<td>130</td>
<td>-70</td>
<td>4.46</td>
<td>690</td>
<td>130</td>
<td>560</td>
<td>5.31</td>
<td>340</td>
<td>140</td>
<td>200</td>
<td>2.43</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>850</td>
<td>540</td>
<td>310</td>
<td>1.57</td>
<td>390</td>
<td>500</td>
<td>-110</td>
<td>1.78</td>
<td>430</td>
<td>390</td>
<td>40</td>
<td>1.10</td>
<td>180</td>
<td>20</td>
<td>160</td>
<td>9.00</td>
<td>690</td>
</tr>
<tr>
<td></td>
<td>340</td>
<td>140</td>
<td>200</td>
<td>2.43</td>
<td>410</td>
<td>440</td>
<td>-30</td>
<td>0.93</td>
<td>350</td>
<td>620</td>
<td>-270</td>
<td>0.56</td>
<td>270</td>
<td>70</td>
<td>200</td>
<td>3.8</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>520</td>
<td>80</td>
<td>240</td>
<td>6.50</td>
<td>250</td>
<td>20</td>
<td>230</td>
<td>12.50</td>
<td>560</td>
<td>300</td>
<td>260</td>
<td>1.87</td>
<td>590</td>
<td>90</td>
<td>500</td>
<td>6.56</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TWO</th>
<th>PAUSE</th>
<th>GIRLS</th>
<th>PAUSE</th>
<th>PLAYED</th>
<th>PAUSE</th>
<th>NEAR</th>
<th>PAUSE</th>
<th>THE</th>
<th>PAUSE</th>
<th>RIV</th>
<th>ER</th>
<th>PAUSE</th>
<th>ONE</th>
<th>PAUSE</th>
<th>HOT</th>
<th>PAUSE</th>
<th>DAY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>450</td>
<td>300</td>
<td>150</td>
<td>1.50</td>
<td>50</td>
<td>1.63</td>
<td>450</td>
<td>125</td>
<td>450</td>
<td>400</td>
<td>250</td>
<td>1.46</td>
<td>450</td>
<td>130</td>
<td>320</td>
<td>1.94</td>
<td>450</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>80</td>
<td>50</td>
<td>1.63</td>
<td>450</td>
<td>125</td>
<td>90</td>
<td>1.50</td>
<td>450</td>
<td>400</td>
<td>250</td>
<td>1.46</td>
<td>450</td>
<td>130</td>
<td>320</td>
<td>1.94</td>
<td>450</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>740</td>
<td>590</td>
<td>150</td>
<td>1.25</td>
<td>450</td>
<td>125</td>
<td>450</td>
<td>125</td>
<td>450</td>
<td>400</td>
<td>250</td>
<td>1.46</td>
<td>450</td>
<td>130</td>
<td>320</td>
<td>1.94</td>
<td>450</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>650</td>
<td>400</td>
<td>300</td>
<td>1.25</td>
<td>390</td>
<td>400</td>
<td>250</td>
<td>1.25</td>
<td>390</td>
<td>400</td>
<td>250</td>
<td>1.25</td>
<td>390</td>
<td>400</td>
<td>250</td>
<td>1.25</td>
<td>390</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>450</td>
<td>130</td>
<td>320</td>
<td>1.68</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>640</td>
<td>380</td>
<td>260</td>
<td>1.68</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
<td>220</td>
<td>1.50</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>440</td>
<td>440</td>
<td>240</td>
<td>4.14</td>
<td>50</td>
<td>440</td>
<td>150</td>
<td>1.87</td>
<td>50</td>
<td>440</td>
<td>150</td>
<td>1.87</td>
<td>50</td>
<td>440</td>
<td>150</td>
<td>1.87</td>
<td>50</td>
<td>440</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>2.50</td>
<td>320</td>
<td>260</td>
<td>120</td>
<td>0.92</td>
<td>320</td>
<td>260</td>
<td>120</td>
<td>0.92</td>
<td>320</td>
<td>260</td>
<td>120</td>
<td>0.92</td>
<td>320</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>390</td>
<td>10</td>
<td>380</td>
<td>39.00</td>
<td>250</td>
<td>380</td>
<td>10</td>
<td>39.00</td>
<td>250</td>
<td>380</td>
<td>10</td>
<td>39.00</td>
<td>250</td>
<td>380</td>
<td>10</td>
<td>39.00</td>
<td>250</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>2.50</td>
<td>220</td>
<td>240</td>
<td>30</td>
<td>1.13</td>
<td>220</td>
<td>240</td>
<td>30</td>
<td>1.13</td>
<td>220</td>
<td>240</td>
<td>30</td>
<td>1.13</td>
<td>220</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>380</td>
<td>130</td>
<td>130</td>
<td>1.52</td>
<td>470</td>
<td>110</td>
<td>130</td>
<td>1.52</td>
<td>470</td>
<td>110</td>
<td>130</td>
<td>1.52</td>
<td>470</td>
<td>110</td>
<td>130</td>
<td>1.52</td>
<td>470</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>490</td>
<td>70</td>
<td>70</td>
<td>2.00</td>
<td>360</td>
<td>360</td>
<td>30</td>
<td>1.13</td>
<td>360</td>
<td>360</td>
<td>30</td>
<td>1.13</td>
<td>360</td>
<td>360</td>
<td>30</td>
<td>1.13</td>
<td>360</td>
<td>360</td>
</tr>
</tbody>
</table>

Table 7 Duration Measurements for Test Sentences 569 and 570
In 90% of the segments compared, the deaf subject extended the utterance. Prolongation of pauses occurred in 72% of the cases.

Sonagrams of the original versions of all ten sentences are shown in Appendix A for both subjects.

**Listening Sessions**

A group of six semi-naive listeners was asked to evaluate the speech samples. Five listening sessions, each separated by a week, were conducted. Ten sentences were given in each session with all four rhythmic conditions represented:

1) Deaf speaker with malphrased and prolonged speech
2) Deaf speaker with tempo-modified speech
3) Normal-hearing speaker with natural rhythmic patterns
4) Normal-hearing speaker with tempo-modified speech

Prior to each session, five practice sentences were given for the purpose of having the listeners adjust their earphones to a comfortable level and familiarize themselves with the general quality of the speech. The instruction sheet given to the listeners is shown in Figure 21.
You will listen to a variety of taped sentences spoken by ten male and female persons. These sentences vary in clarity. In addition, a new set of the same sentences have been experimentally modified with the addition of various noise and distortion. The object of the experiment is to determine how well you can understand these sentences. This will be measured by the number of words you identify correctly. You will listen to each set and attempt to write in words according to the following procedures:

PART I. PRAXIS
The first five sentences are for practice only, to allow you to become accustomed to the experiment. You will not write in your answer book.

1. Before each sentence in this part, the examiner will call out the SENTENCE NUMBER. Locate the appropriate numbered line on your answer sheet.

2. LISTEN to the sentence. You will only hear it once. WRITE your answer to the sentence in the correct numbered line on your answer sheet. These answers will be used to grade your score. GRADE your answers carefully as this is your first attempt at a test.

3. WRITE all your answers for this part as neatly as you can. Do not erase anything, and rewrite each sentence carefully before going on. Write only words or names from the paragraph to which each instruction applies.

PART II. TEST
The "practice" procedure which has just described will not be used for the TEST sentences.

1. Use the answer sheet marked "TEST" for this part of the test session.

PLEASE
Since you will be testing for additional listening practice, do not hesitate to ask if you do not understand any of the instructions. There is no way to answer these TEST items for your own test. Figure 21. Instructions to the Listeners

After the four sessions were completed, a fifth session was held to test the effects of recall on the intelligibility ratings. In this fifth session, the sentences from the first tape were repeated and the results were compared.

The listeners were not informed about the sources of the sentences in the tape sessions, the order of the sentences or their intelligibility. The listeners were not told that the same sentences would appear again, spoken by a different subject, or retimed by computer. It was not possible for the listeners to determine the order of the sentences, nor which subject would be speaking next.

The purpose of the study was explained after the final listening session.
RESULTS

The data from the listening sessions was analyzed in three ways. First, the general word intelligibility was assessed. Next, a word-by-word analysis was attempted to study the specific change of intelligibility of each individual word in the pre- and post-rhythm modification versions of the test sentences. The third analysis was one of key contextual words in the sentences. The reader should refer to Table 8 for a listing of the key contextual words which were selected on the basis of subject-verb-object groupings. Tables 9 to 18 give the data for all three analyses.

Two engineers with phonetic training evaluated the tapes to identify certain words which were distorted by the computer. If the distortion was significant enough to mask the words, the words were deleted from the data. The number of words per sentence, therefore, varied at times. Even when the data was analyzed without consideration of distortion by computer, the results were very similar to those discussed on the following pages.

Deaf Speaker - Individual Word Analysis

Twenty-eight (28) per cent of the words were identified correctly in both the pre- and post- rhythm modification versions. Eleven (11) percent of the words were identified correctly after rhythm modification which were not identified in the original malphrased versions.
Fifteen (15) percent of the words were identified correctly in the malphrased versions but were not identified correctly in the retimed sentences. Forty-six (46) percent of the words were unintelligible in both versions. This data is shown in summary Table 19.

Considering the relative locations of the words in the listening sessions, the data does not show a significant trend toward an increment or decrement in intelligibility of individual words for the deaf subject. His overall intelligibility was very low, as is shown by the data.

**Deaf Speaker - General Word Intelligibility**

The general word intelligibility was calculated using Tables 9-18 by dividing the number of words correctly identified by the total possible score. Forty (40) percent of the words were identified correctly in the malphrased versions and thirty-eight (38) percent of the words were identified in the retimed sentences. Again the data shows no significant trend toward a change in intelligibility for the deaf speech. Some factors which may have influenced the data are presented in the discussion.

**Deaf Speaker - Key Contextual Words**

Table 19 shows that seventy (70) percent of the test sentences in the inventory had an equal or improved intelligibility rating for key
contextual words. Four sentences showed an improvement in the intelligibility of key contextual words when the deaf speech was retimed to the rhythmic patterns of the normal-hearing subject. Three sentences showed an equal intelligibility rating before and after the retiming. Of the three sentences which decreased in intelligibility, one decreased by only a single context word, one by two context words and a third by four context words. Context words were selected on the basis of subject-verb-object groupings (see Table 8).

The engineers who audited the tapes agreed that several sentences which showed a decrease in intelligibility appeared to them to be more intelligible, but their relative locations early in the listening sessions might have had some significance. A possible explanation is that the retimed versions of the deaf speech early in the study had an impact on the rhythmic expectations of the listeners. This was quite obvious when watching their reactions through an observation window. After their initial reaction to deaf speech with more temporal fluidity, the listeners appeared to have realized that they would be hearing variations in the time domain. Consequently, the conjecture may be made that some important data was lost during this initial reaction.

The apparent attention given to key contextual words is important. This writer hypothesizes that the listeners have trained themselves to identify key contextual words, disregarding the relatively unimportant adjectives, conjunctions and articles. It appears to be a typical and desirable receptive skill of educators of the profoundly deaf. This study presented a challenge to this skill, however, since the listeners
had no reinforcement via total communication (i.e., lip movements, and manual signs simultaneous with speech).

Since the listeners were all educators of the deaf, applications of the findings of this analysis to a typical man-on-the-street encounter between a deaf person and someone naive to the deaf speaker's vocal traits must be done cautiously.

The results for the deaf subject suggest the hypothesis that when the temporal framework is more compatible to the listeners rhythmic intuition, more attention may be paid to sentence context. This is demonstrated by 30 percent of the sentences having at least equal intelligibility, and 40 percent showing some improvement in the key contextual word ratings, while the general word intelligibility for all words in the sentences is comparatively unaltered by rhythm modification.

561 A BROWN BIRD SAT ON A FENCE NEAR A HOUSE
562 TOM PUT SOME OLD SHOES IN A WASTEBASKET
563 A GIRL LOST AN UMBRELLA AT THE ZOO
564 A MAN TOOK SOME BOOKS TO THE HOSPITAL
565 A POLICEMAN WALKED TO THE SCHOOL PICNIC
566 A COWBOY WENT SHOPPING IN THE COUNTRY
567 A BABY PUT A BALL IN A BIG SHOE
568 A DOG PLAYED WITH A WHITE LAMB IN THE PARK
569 MISS BROWN WANTED SOME TEA AT THE MEETING
570 TWO GIRLS PLAYED NEAR THE RIVER ONE HOT DAY

Table 8  Key Contextual Words
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (19%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (8%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (8%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (65%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (24%) CONTEXT WORDS
AFTER RETIMING (27%)

BEFORE RETIMING (11%)
AFTER RETIMING (17%)

Table 9 SENTENCE # 561 DEAF SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (40%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (10%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (23%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (27%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (50%)
AFTER RETIMING (52%)

CONTEXT WORDS
BEFORE RETIMING (58%)
AFTER RETIMING (56%)

Table 10 SENTENCE # 562 DEAF SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (37%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (27%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (7%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (30%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (44%) AFTER RETIMING (43%)

CONTEXT WORDS

BEFORE RETIMING (29%) AFTER RETIMING (42%)

Table 11 SENTENCE # 563 DEAF SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (92%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (3%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (5%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (89%) AFTER RETIMING (97%)

CONTEXT WORDS

BEFORE RETIMING (83%) AFTER RETIMING (91%)

Table 12 SENTENCE # 564 DEAF SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (22%)

PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (28%)

PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (8%)

PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (42%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (43%) BEFORE RETIMING (17%)

AFTER RETIMING (31%) AFTER RETIMING (17%)

Table 13 SENTENCE # 565 DEAF SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (30%)

PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (0%)

PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (37%)

PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (33%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (22%) BEFORE RETIMING (6%)

AFTER RETIMING (67%) AFTER RETIMING (67%)

Table 14 SENTENCE # 566 DEAF SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (33%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (30%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (7%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (30%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING   (50%)  BEFORE RETIMING   (50%)
AFTER RETIMING    (40%)  AFTER RETIMING    (28%)

Table 15  SENTENCE # 567  DEAF SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (17%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (17%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (13%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (54%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING   (33%)  BEFORE RETIMING   (21%)
AFTER RETIMING    (29%)  AFTER RETIMING    (21%)

Table 16  SENTENCE # 568  DEAF SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (17%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (29%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (54%)

GENERAL WORD INTELLIGIBILITY

<table>
<thead>
<tr>
<th></th>
<th>BEFORE RETIMING</th>
<th>AFTER RETIMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTEXT WORDS</td>
<td>(53%)</td>
<td>(17%)</td>
</tr>
<tr>
<td>GENERAL WORD INTELLIGIBILITY</td>
<td>(53%)</td>
<td>(17%)</td>
</tr>
</tbody>
</table>

Table 17  SENTENCE # 569  DEAF SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (0%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (4%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (2%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (94%)

GENERAL WORD INTELLIGIBILITY

<table>
<thead>
<tr>
<th></th>
<th>BEFORE RETIMING</th>
<th>AFTER RETIMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTEXT WORDS</td>
<td>(4%)</td>
<td>(2%)</td>
</tr>
<tr>
<td>GENERAL WORD INTELLIGIBILITY</td>
<td>(4%)</td>
<td>(2%)</td>
</tr>
</tbody>
</table>

Table 18  SENTENCE # 570  DEAF SPEAKER
GENERAL WORD INTELLIGIBILITY

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE</td>
<td>40%</td>
</tr>
<tr>
<td>POST</td>
<td>38%</td>
</tr>
</tbody>
</table>

INDIVIDUAL WORD ANALYSIS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE AND POST</td>
<td>28%</td>
</tr>
<tr>
<td>PRE ONLY</td>
<td>15%</td>
</tr>
<tr>
<td>POST ONLY</td>
<td>11%</td>
</tr>
<tr>
<td>NEITHER PRE NOR POST</td>
<td>46%</td>
</tr>
</tbody>
</table>

KEY CONTEXTUAL WORDS - INTELLIGIBILITY TRENDS

<table>
<thead>
<tr>
<th>SENTENCE #</th>
<th>PERCENTAGE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>561</td>
<td>+ 6%</td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>- 2%</td>
<td></td>
</tr>
<tr>
<td>563</td>
<td>+13%</td>
<td></td>
</tr>
<tr>
<td>564</td>
<td>+ 8%</td>
<td>(High intelligibility for both versions)</td>
</tr>
<tr>
<td>565</td>
<td>NO CHANGE</td>
<td></td>
</tr>
<tr>
<td>566</td>
<td>+61%</td>
<td></td>
</tr>
<tr>
<td>567</td>
<td>-22%</td>
<td>(appeared early in the study)</td>
</tr>
<tr>
<td>568</td>
<td>NO CHANGE</td>
<td></td>
</tr>
<tr>
<td>569</td>
<td>-28%</td>
<td>(appeared early in the study)</td>
</tr>
<tr>
<td>570</td>
<td>NO CHANGE</td>
<td>(both versions had very low intelligibility)</td>
</tr>
</tbody>
</table>

Table 19 Summary of Intelligibility Analyses
Profoundly deaf subject
Normal-hearing Speaker - Individual Word Analysis

The results of the data analysis for the normal-hearing speaker substantiate the general hypothesis that rhythm is important to the intelligibility of speech. Table 30 shows that ninety-one (91) percent of the words were identified correctly in both the pre- and post-rhythm modification versions.

When the speech of the normal-hearing subject was retimed to give it the rhythmic patterns of the profoundly deaf subject, eight (8) percent of the words were identified correctly in the pre-rhythm modification versions but were not identified correctly after the retiming process. One (1) percent of the words were identified correctly after retiming, but were not in the original versions.

Although the trends have shown a general decrease in intelligibility, some of the data may have been influenced by segmentation and distortion by computer.

Normal-hearing Speaker - General Word Intelligibility

There was measured a seven (7) percent decrement in intelligibility of general words when the rhythm of the normal-hearing subject was modified to resemble that of the deaf subject. In the original sentences, there was a ninety-nine (99) percent intelligibility rating (Table 30), and ninety-two (92) percent of the words were identified after the rhythm modification. The results indicate that rhythm is
important to the perception of general words spoken by the normal-hearing person.

**Normal-hearing Speaker - Key Contextual Words**

The data shows that when the rhythm was modified, the number of key contextual words correctly identified decremented twelve (12) percent. Almost one-hundred (100) percent of the key contextual words were correctly identified in the original sentences. Table 30 shows that eighty-eight (88) percent of the key contextual words were identified correctly in the post-rhythm modification versions. Most of the sentences showed some decrease in key contextual word intelligibility (Tables 20-29).

The trend toward a decrease in intelligibility, and, especially, the sharp decrease which occurred for the first few sentences with malphrased patterns may be attributed to the rhythmic intuition of the listeners. In this case, the listeners were not prepared for such deviation from the durational characteristics of General American English as spoken by a normal-hearing adult.

Even after the listeners mentally prepared themselves to meet some rhythmic variety, the intelligibility ratings for key contextual words and for general words in the sentences were consistently less than the ratings for sentences with normal rhythmic patterns of the same speaker.

The writer concludes from the data that abnormal rhythm distracts the listener from the context, whether he is listening to speech spoken
by a deaf or normal-hearing individual. If the psychological reaction of the listeners to the first few sentences spoken by both subjects was considered, the data trends provide more consistent evidence of the importance of rhythm to the perception of speech.
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (100%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (0%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (100%)
AFTER RETIMING (100%)

CONTEXT WORDS

BEFORE RETIMING (100%)
AFTER RETIMING (100%)

Table 20  SENTENCE # 561  NORMAL-HEARING SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (98%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (0%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (2%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (98%)
AFTER RETIMING (100%)

CONTEXT WORDS

BEFORE RETIMING (94%)
AFTER RETIMING (100%)

Table 21  SENTENCE # 562  NORMAL-HEARING SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (100%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (0%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING  (100%)  
AFTER RETIMING  (100%)

CONTEXT WORDS

BEFORE RETIMING  (100%)  
AFTER RETIMING  (100%)

Table 22  SENTENCE # 563  NORMAL-HEARING SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (97%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (3%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING  (100%)  
AFTER RETIMING  (97%)

CONTEXT WORDS

BEFORE RETIMING  (100%)  
AFTER RETIMING  (91%)

Table 23  SENTENCE # 564  NORMAL-HEARING SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (100%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (0%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (100%) BEFORE RETIMING (100%)
AFTER RETIMING (100%) AFTER RETIMING (100%)

Table 24 SENTENCE # 565 NORMAL-HEARING SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (93%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (7%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING (100%) BEFORE RETIMING (100%)
AFTER RETIMING (93%) AFTER RETIMING (88%)

Table 25 SENTENCE # 566 NORMAL-HEARING SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (55%)

PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (43%)

PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (0%)

PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (2%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING  (98%)  
AFTER RETIMING  (52%)  
CONTEXT WORDS

BEFORE RETIMING  (100%)  
AFTER RETIMING  (54%)

Table 26  SENTENCE # 567  NORMAL-HEARING SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING  (88%)

PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY  (5%)

PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY  (7%)

PERCENTAGE NOT IDENTIFIED IN EITHER VERSION  (0%)

GENERAL WORD INTELLIGIBILITY

BEFORE RETIMING  (94%)  
AFTER RETIMING  (96%)  
CONTEXT WORDS

BEFORE RETIMING  (100%)  
AFTER RETIMING  (96%)

Table 27  SENTENCE # 568  NORMAL-HEARING SPEAKER
INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (83%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (17%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

| BEFORE RETIMING (100%) | BEFORE RETIMING (100%) |
| AFTER RETIMING (83%) | AFTER RETIMING (79%) |

Table 28  SENTENCE # 569  NORMAL-HEARING SPEAKER

INDIVIDUAL WORD ANALYSIS

PERCENTAGE IDENTIFIED CORRECTLY BEFORE AND AFTER RETIMING (96%)
PERCENTAGE IDENTIFIED CORRECTLY BEFORE RETIMING ONLY (4%)
PERCENTAGE IDENTIFIED CORRECTLY AFTER RETIMING ONLY (0%)
PERCENTAGE NOT IDENTIFIED IN EITHER VERSION (0%)

GENERAL WORD INTELLIGIBILITY

| BEFORE RETIMING (100%) | BEFORE RETIMING (100%) |
| AFTER RETIMING (96%) | AFTER RETIMING (96%) |

Table 29  SENTENCE # 570  NORMAL-HEARING SPEAKER
GENERAL WORD INTELLIGIBILITY

PRE 99%
POST 92%

INDIVIDUAL WORD ANALYSIS

PRE AND POST 91%
PRE ONLY 8%
POST ONLY 1%
NEITHER PRE NOR POST 0%

KEY CONTEXTUAL WORDS – INTELLIGIBILITY TRENDS

<table>
<thead>
<tr>
<th>SENTENCE #</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>561</td>
<td>NO CHANGE</td>
</tr>
<tr>
<td>562</td>
<td>+ 6%</td>
</tr>
<tr>
<td>563</td>
<td>NO CHANGE</td>
</tr>
<tr>
<td>564</td>
<td>- 9%</td>
</tr>
<tr>
<td>565</td>
<td>NO CHANGE</td>
</tr>
<tr>
<td>566</td>
<td>-12%</td>
</tr>
<tr>
<td>567</td>
<td>-46%</td>
</tr>
<tr>
<td>568</td>
<td>- 4%</td>
</tr>
<tr>
<td>569</td>
<td>-21%</td>
</tr>
<tr>
<td>570</td>
<td>- 4%</td>
</tr>
</tbody>
</table>

Table 30 Summary of Intelligibility Analyses
Normal-hearing subject
DISCUSSION

General Discussion of the Results

The rhythmic patterns in speech have been found to be important to perception. For the profoundly deaf subject, seventy (70) percent of the test sentences exhibited an equal or improved intelligibility of key contextual words when the rhythm was modified to resemble that of a subject with normal hearing.

A decrease in intelligibility of both general words and key contextual words spoken by a normal-hearing subject was measured when his rhythm was modified to resemble that of the profoundly deaf subject.

This writer hypothesizes that the rhythmic intuition of listeners which appears to be vital in this investigation may be just as important in oral communication between deaf and hearing people in the greater community.

Each deaf person has his own unique voice quality. Rhythmic patterns and articulation are two characteristics which vary greatly from deaf person to deaf person. A teacher of the deaf must familiarize himself with the speech patterns of an individual before he is able to perceive the speech (without simultaneous communication of signs and lipreading).

English is a stress-timed language. The time intervals between successive major stresses are thought to be roughly equal. Speakers
of the English language with no hearing impairment normally sense their rhythm intuitively.

Considering a man-on-the-street encounter, then, between a profoundly deaf person with poor speech rhythm and a person naive to the general speech characteristics of the deaf, it is logical to conclude that the rhythm patterns, per se, are detractive, and that they act in conjunction with poor articulation, breathiness, nasality and other speech problems to produce low speech perception. If educators of the deaf must train themselves to identify context words in the malphrased speech of the deaf, then the average, or "naive" listener would most likely have even more difficulty.

Other writers have found that rhythm has a vital influence on speech intelligibility. The Clarke School for the Deaf, for example, found an 0.57 correlation between intelligibility and normality of rhythm.\textsuperscript{62}

The Ewings rated recordings of deaf children showing lack of continuity of utterance which appeared to be related to the inability of the children to retain images and ideas simultaneously in the mind.\textsuperscript{63}

The problem of cognition may work in varying degrees with the psychophysical and psychosocial constraints discussed earlier in this study.


\textsuperscript{63}Ewing and Ewing, p.70.
Some Possible Factors Influencing The Data

1) **Overfortis Speech**

The sonagrams clearly display the excessive effort put into articulation by the deaf subject. Known as overfortis, this characteristic is thought to be caused by an overemphasis in producing articulatory strictures and/or excessive breath pressure. The speech resulting from overfortis contains long plosive, fricative and nasal sounds as well as excessive aspiration and affrication of stops. The excessive breath pressure is partly responsible for the pitch fluctuations heard in some cases.

The spectral energy of the deaf subject often exceeded several times the energy of a normal-hearing speaker for identical words.

A word was selected at random for power comparison. The deaf subject, when speaking the word "A", had an overall power of 1480 units over 430 milliseconds as measured on the computer-generated spectra. The same word in the same test sentence spoken by the normal-hearing subject had an overall power measure of 522 units over 130 milliseconds.

The ratio of the overall powers was found to be 2.8, clearly indicating the relative energy in their speech. The result of shortening overfortis speech was an apparently choppy quality which may have distracted the listeners to some degree.
2) **Intelligibility of the Deaf Subject**

The deaf subject recommended for the study had a low articulation profile. When articulation is very poor, the durations of the sounds may possibly be helpful in their identification. The conjecture may be made that if a deaf subject with better articulation were used in this study, the data may have shown a more definitive correlation.

3) **Mechanics of Rhythm Modification**

There was some question as to what extent the mechanics of the retiming process influenced the intelligibility as an inherent factor. Individual speech and pause segments were tempo-modified according to the acoustic parameter values at boundaries or at times within a glide transition. When the deaf speaker's utterance was phonemically divided into unnatural, temporally-isolated sounds, an attempt to accurately produce a similar spectral sequence may have resulted in decreased intelligibility of the normal speech.

When a speech segment spoken originally by a normal-hearing subject was temporally expanded to several times the initial duration, the otherwise smooth pitch trend becomes jittery due to duplication of each short-time spectral component.

A similar effect occurred with the deaf speech when shortened to a new duration. Since the computer algorithm used to synthesize the speech determined the fundamental frequency for each ten millisecond short-time spectral segment, newly alligned segments previously separated by twenty or thirty milliseconds had pitch values differing by
five or ten Hertz. The result was a jitter of the fundamental frequency.

4) **Quality of the Simulated Speech**

The quality of the synthetic speech was not very different from the speech originally recorded on audio tapes. A sacrifice was made in a small amount of fricative energy (over 5000 Hz) to avoid aliasing of low frequency signals. The quality of the retimed speech of the deaf subject, however, is questionable with respect to how representative it is of a profoundly deaf person with good rhythm.

There is a sharp difference between improving the rhythm of a deaf subject by a cut-and-dried computer process and in emphasizing rhythmic aspects in speech training simultaneously with articulation, breathing, etc.

The choppiness previously described is definitely a product of the mechanics of the rhythm modification process. The pitch jitter is another source of unnaturalness. In summary, the writer does not believe that the computerized deaf speech with good rhythm was accurately representative of similar speech acquired through training. 64

5) **Listener Recall**

The data shows that although listener recall was an

---

64 For example, the rhythm-modified speech produced in R. Houde's investigation.
influencing factor, it does not detract from the general findings of the investigation.

A fifth listening session was held to test the effects of recall on the data. For the speech of the normal-hearing subject, only one sentence showed a significant increase, and this was the poor rhythmic version which initially scored a very low rating. An increase for this sentence was inevitable.

There was an apparent learning effect shown in the data for the deaf speech, especially for the sentences with poor rhythm. For the sentences with retimed rhythm, there was no increase in the number of words identified correctly in the fifth session. These findings do not change the general trends observed in the data.

For the five practice sentences given before each of the sessions, nearly 80% of the sentences showed no increase from session to session.

Additional Findings

The effect of inputting syllable-type segments into the tempo-modification algorithm which was insensitive to delicate spectral boundaries was to produce speech deficient in consonantal sounds. Deletion of plosives left the synthetic speech highly vowel in character. This problem was solved by segmenting on a sub-syllable basis and retiming each speech sound separately.

The greatest amount of acoustic energy deletion was performed on the vowel portions of the syllables. These are characteristically
time-distorted the most by deaf speakers.

In terms of the analysis-synthesis process, several important results were found. Having a cutoff frequency of 4 kilohertz to avoid aliasing deleted some of the high frequency fricative energy, especially in the unvoiced fricative "s" sound.

This resulted in a greater difference than normally found between the speech sound durations as measured on the sonagrams and computer-generated spectrograms. This error had no influence on the perception ratings of these sounds, however, since they are long in time.

It was also found that extreme care had to be taken when digitizing the speech with the HP2116-B mini-computer. Sixty-cycle hum should be grounded out and recordings should be made in an anechoic chamber, or one with suitable acoustics. Any low frequency noise whatever may produce small peaks in the power spectrum from which the pitch tracker derives the fundamental frequency.

Pitch values characteristically do not change more than 10 Hertz in 20 milliseconds except at the onset of voicing or other times of great spectral change. If the "Peak Picker" calculates a fundamental frequency based on a peak resulting from low-frequency noise, the result will be large pitch fluctuations of 10 millisecond durations and the speech will have an undesirable quality. Care must also be taken to ground out 60 cycle hum from the tape equipment used for the listening sessions.

The pitch tracker described above may easily be revised, however, to ignore unusual jumps in the fundamental frequency by searching the
trend over a period of time.

It should be mentioned that since all other parameters were held constant in this study, their presence in the rhythm-modified speech wave had as vital a role in perception as in the original signal. Malarticulation, breathiness and nasality were especially apparent.

A semantic differential was evident in the data at least once when several listeners identified the word "MISS" as "MRS". Spoken by the normal-hearing subject, the retimed version was longer in fricative sound, giving an apparent approximation to "MRS". All listeners identified "MISS" correctly in the original sentence. Although the differential is not significant in terms of sentence context, it does support the earlier discussion of the effects of time on the features of speech, in this case a semantic feature.  

---

65 Voicing the entire fricative in this word would surely solve the whole problem.
CONCLUSIONS

**Summary**

The objective of this investigation was to study the durational characteristics of a profoundly deaf subject and a normal-hearing subject operating under the premise that rhythm modification by digital computer techniques will make the speech more natural to the listener. This would supposedly allow the listener to attend to the context without the typical distraction from distortion of the temporal framework.

The results of this study support this premise. Although the speech was distorted to some degree by removal of some important acoustic information, especially near the segment boundaries, seven of the ten tempo-modified sentences spoken by the deaf subject showed an equal or improved intelligibility rating for key contextual words.

The speech was evaluated by semi-naive listeners, and the results substantiate the hypothesis that rhythm is of importance to the intelligibility of both speakers.

Native speakers of the language develop a feel for rhythm intuitively. Not only does this rhythmic intuition react when different patterns are perceived, but decreased perception of the listener looking for natural rhythm patterns is likely to occur. This intuitive reaction was shown in the data for two subjects; one with normal hearing and the other deaf. A decrease in both key contextual word intelligibility and general word intelligibility was measured when the
rhythm of the hearing speaker was made to resemble that of the deaf subject. An increase in key contextual word intelligibility was measured when the deaf subject's speech was modified to have the rhythmic patterns of the normal-hearing speaker.

The results of this study indicated that with all speech errors unaltered except rhythm, the speech intelligibility of a profoundly deaf subject is improved when the temporal framework of the speech sounds is compatible with the rhythm of the listener's dialect.
Implications for Further Research

The digital computer is a powerful tool which can be applied to obtaining objective information on the acoustical/phonetical characteristics of the speech of the deaf. In this study alone, for example, the computer was used as an investigative tool through an analysis-synthesis scheme by a neophyte in speech research who is himself deaf.

Valuable quantitative measurements can be obtained quickly and accurately. The potential for developing instantaneous feedback through the visual and tactual modalities via digital computer is awesome.

The data from this investigation gives encouragement for further study of the time domain in the speech of the deaf. This writer plans to pursue the problem with a larger sentence inventory and variety of subjects. More focus will be placed on the correlation between articulation and rhythmic qualities.

The synthetic quality of the speech will be improved by an analysis of the speech signal processing algorithms. An attempt will be made to adjust the overall power after retiming by calculating a scaling factor based on overfortis energy in the deaf speech. In this way the choppiness and pitch jitter produced by realigning segments will be reduced significantly.

A smoothing of filter outputs will be attempted as another strategy to produce an easier transition of energy over the time domain for newly alligned short-time spectra. The results will be compared to those of the present investigation.
More concentrated research is suggested by this writer in the following areas:

1. The effect of continuous speech-rate training combined with articulatory training in various proportions should be studied in a controlled research environment. Programs for deaf children in Eurhythmics which teach rhythm in speech and action should be encouraged.

2. The development of inexpensive instruments for visual and/or tactual feedback of speech rate should be intensified. For example, the potential of the conventional television set as a visual feedback device for spontaneous displays should be investigated. With such a device accessible, willing parents might tutor a profoundly deaf child in appropriate durational aspects of speaking while the child is still young. This could be done in the home at a relatively inexpensive cost.

3. The potential of the digital computer for teaching speech in a time-sharing environment and for further exploring the accoustical deficiencies of deaf speakers should also be investigated. Analysis – synthesis techniques should be pathologically applied to the speech of the deaf.

4. Further development of complex instrumentation for speech pathology should be increased. Applications of new findings in electronic instrumentation should be continuously investigated.
Sonagrams of ten test sentences

Figures 22-31 illustrate the relative durations of speech and pause segments of the normal-hearing and profoundly deaf subjects.

In each figure, the sonagrams at the top of the page illustrate the speech of the normal-hearing subject. The second version of the sentence shows the abnormal rhythmic patterns of the deaf subject.
A BROWN BIRD SAT ON A FENCE NEAR A HOUSE

Figure 22  Test Sentence #561
Figure 23

Test Sentence #562

TOM PUT SOME OLD SHOES IN A WASTE BAS KET

TOM PUT SOME OLD

SHOES IN A WASTE BAS KET
A man took some books to the hospital.
A POLICE MAN WALKED TO THE SCHOOL PICNIC

A POLICE MAN WALKED TO THE

SCHOOL PICNIC

Figure 26  Test Sentence #565
A DOG PLAYED WITH A WHITE LAMB IN THE PARK

A DOG PLAYED WITH A WHITE LAMB IN THE PARK

Figure 29 Test Sentence #568
MISS BROWN WANTED SOME TEA AT THE MEETING

MISS BROWN WANTED SOME

TEA AT THE MEETING

Figure 30 Test Sentence #569
TWO GIRLS PLAYED NEAR THE RIVER ONE HOT DAY

TWO GIRLS PLAYED NEAR THE RIVER

ONE HOT DAY

Figure 31 Test Sentence #570
A Brief Description of Some Attempts At Segmentation

A segmentation of the speech continuum into successive time segments may be done in various ways. Temporal segments may be phonemes, phoneme dyads, syllable nuclei and margins, half-syllables, syllables, syllable dyads or words. Segment inventories for speech synthesis, and estimates of the number of segments required for each type were compiled by Peterson and Sivertsen. Estimates of the number of prosodic conditions for each type were also made available.

The most common belief at present is that segment boundaries should always be chosen at points of relatively sustained positions within the speech, or at points of natural discontinuity.

Peterson, Wang and Sivertson describe a method known as dyad segmentation, in which the segments are characterized as containing parts of two phones with the mutual influence in the middle of the segment, and beginning and ending at the phonetically most stable position of each phone.

---


Their method was relatively crude in that they cut and spliced tape with corresponding clicks at the tape junctions and spikes on the broad band spectrograms. Muma and Brown also attempted tape-splicing segmentation, as did Harris.

Vowel-consonant diphones were segmented in isolation by Estes et al. The segments were digitized and stored in a computer library with pitch and timing information, and the control signals corresponding to the sub-word segments of speech were later used for synthesis. Reddy describes his primary segmentation using variation of stability of sound intensity levels. Zero crossing counts were used as an aid in resolving ambiguities and in error correction.

Newcomb, Larkin and Houde did a study on speech segmentation

---


69 C.M. Harris, "A Study of the Building Blocks of Speech," JASA, 25 (1953), 962-969.


using a computer-generated spectrum and employing a spectral change function to calculate the relatively sustained positions. Segmentation boundaries were found at either points of maximum spectral change (e.g., immediately before or after a plosive), or points of minimum change (e.g., at the center of a glide transition). They also studied time-smoothing, finding that they received best results when two adjacent short-time spectra were averaged to achieve a 20 millisecond spectrum. 300 Hertz filters were spaced every 150 Hertz (overlapping), and every 300 Hertz (contiguously) with no significant difference in segmentation consistency.

Liljencrantz also employed a spectral change function but did not normalize in amplitude. 73

Cohen and Hart used steady-state segments, with many details of formant information neglected, to test the contribution of the time parameter to the perception of speech. 74 Employing a variable gating circuit, they were able to vary the time envelope of the individual segments.

In addition, Landercy 75 studied the spectral distortions due to

---


exponential and trapezoidal envelopes. This study is helpful in seeing the distortion introduced by a spectral gating function, and the fundamental role this distortion may play in perception, if we accept that the ear carries out a spectral analysis of the presented sounds.

Segment-type features determine segment boundaries by focusing on discontinuities, and segment pattern features specify the contents of the speech segments beyond the categorization of the type features. Known as manner of production and place of articulation respectively, the features were classified and described thoroughly by the Speech Transmission Laboratory.  

Rhody proposes even further development in the sophisticated approaches to segment coding in terms of acoustically/phonetically defined speech segments.

Newcomb discusses the role of perception as a basis for determining syllable boundaries.

These are just a few of the many attempts at segmentation of speech sounds. Additional references can be found in the bibliographies of these papers.

---


APPENDIX C

A Brief Review of the Literature on Rhythm Feedback Devices

Visual Displays

Rhythm can be very easily displayed visually as a time-intensity pattern. Risberg describes a basic rhythm indicator which displays the difference between voiced and unvoiced sounds as deflections above and below the zero-intensity axis. The intensity of the speech sound determines the amount of deflection.

Pickett discusses various time-swept contour indicators for displaying intonation and rhythm. Simultaneous display of other prosodic features is also described in this paper.

Pronovost and Goldberg have done thorough investigations of


various visual displays for feedback of speech information and the development of these systems.

Nickerson and Stevens\(^8^3\) describe a recently-designed computer based display of speech signals. A computer assisted instruction (CAI) scheme enables the deaf subject to help monitor his own voice.

An extremely efficient speech display device is presently being used at the National Technical Institute for the Deaf. Designed by Stewart, Larkin and Houde,\(^8^4\) the ingenious device known as the Visible Speech Training Aid (VSTA) displays the frequency-intensity-time parameters on a moving electronic "Strip-chart". The teacher's representation of the utterance may be overlaid on the subject's attempt for direct comparison. The versions are differentiated by their shades.

**Tactual Devices**

Tactual vocoders have been built and tested by many students of audition and speech acoustics.\(^8^5\)

Discrimination of vowels and consonants, for example, have been

\(^8^3\)R. Nickerson and K. Stevens, 445-455.


studied in depth. 86 Frequency transposition from high frequency sounds to low frequency sounds has enabled the fricative consonants to also be presented to the deaf through the tactile modality.

Whether or not there is frequency transposition, the tactile vocoder is recognized as a potential tool for rhythm training. Voiced sounds generate distinctive tactile patterns. Lower energy unvoiced sounds contrast sharply with the voiced sounds approaching pauses in the display.

The tactile vocoder presents the sound pattern to the fingertips. Other vibrotactile pulsers have been designed for duration and rhythm feedback through the wristbone.

Increases of oral reading rate of deaf subjects, for example, have been measured by Drisko87 using this technique.

Whether the tactile or visual sensory receptors are used for feedback of the acoustical signal may depend on the training situation. If reading material is being used, the subject will be unable to attend to visual stimuli effectively while reading orally.


87 M. Drisko, 1-63.
In any case, therapists are often at odds as to the degree of emphasis which should be placed on rhythm in speech training and the effort to which should be spent on training rhythm in proportion to articulation training. One reason for this problem may be the technical difficulty involved in inefficient feedback-training procedures.

Ainsworth, W. A. "First Formant Transitions and the Perception of Synthetic Semivowels." JASA, 44(1968), 689-694.


Dudley, H. "The Vocoder." Bell Labs, Record 17(1939), 122-126.


Fry, D. B. "Duration and Intensity as Physical Correlates of Linguistic Stress." *JASA,* 27(1955), 765-768.


Harris, C. M. "A Study of the Building Blocks in Speech." JASA, 25(1953), 962-969.


Lindblom, B. "Vowel Duration and a Model of Lip Mandible Coordination." *STL-QPSR*, 4(1967), 1-29.


Peterson, G. "Frequency Detection and Speech Formants." JASA, 23(1951), 668-674.


