

Negative microfilm
received from university

65-5315

HANNE, John Ridge, 1936-
FORMANT ANALYSIS.

The University of Michigan, Ph.D., 1964
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

FORMANT ANALYSIS

by

John Ridge Hanne

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in The
University of Michigan
1964

Doctoral Committee:

Professor Gordon E. Peterson, Chairman
Doctor Hugh K. Dunn
Professor Harvey L. Garner
Associate Professor Henry H. Swain

Acknowledgements

I wish to express my appreciation to Professor G. E. Peterson for his help in all phases of the research reported herein. He was the teacher who introduced me to the topic, taught me to understand the full nature of the problem, and guided me toward the solution. In addition, it was his laboratory which provided the facilities for this research. His guidance ranged all the way from editorial comments to highly technical discussions.

I would like to thank Doctor H. K. Dunn for the many discussions we have had regarding the problem of formant analysis. In addition, I am indebted to him for a very careful perusal of the final manuscript. I would also like to acknowledge the many discussions which I have had with Professor H. L. Garner regarding some of the electrical engineering problems associated with the development of the systems used in this research. Professor H. H. Swain made a number of very helpful suggestions to improve the clarity of presentation of the dissertation.

The final copy of the manuscript is the result of the efforts of several members of the staff of the Communication Sciences Laboratory. Miss Sharon Bauerle typed the entire manuscript, Miss Mary Morey proofread the final version of the dissertation, and Mr. Carter Rose served as draftsman for the figures.

Special appreciation is extended to Patricia Hanne for her encouragement and understanding throughout the duration of my graduate study. I am also indebted to her for proofreading several versions of the

dissertation.

I should like to thank the Bendix Corporation for their assistance with tuition during the period of my graduate work in which I was employed by the Bendix Systems Division. I am also grateful to the General Electric Foundation for a tuition scholarship which was awarded to me through the Program in Communication Sciences for the academic year 1961-62. I owe special thanks to Mr. Richard Wilcox and to Mr. Gordon Goldstein of the Information Systems Branch of the Office of Naval Research. Without their support under Contract Nonr 1224(22), NR 049-122, this research could not have been conducted.

Table of Contents

	<u>Page</u>
Acknowledgements.....	11
List of Tables.....	vii
List of Illustrations.....	viii
List of Appendices.....	x
1. The Formant Tracking Problem.....	1
2. Representation and Analysis of Speech Signals.....	7
1. Information Content of Speech.....	8
2. Techniques Used in Speech Analysis.....	12
a. Spectrographic analysis.....	13
b. Filter banks.....	14
c. Zero crossing measurements.....	15
d. Diversity approaches.....	17
e. Spectral moment calculation.....	19
f. High energy harmonic trackers.....	20
g. Phase techniques.....	21
h. Analysis by synthesis.....	22
i. Pitch synchronous techniques.....	22
j. Inverse filtering.....	24
3. Discussion.....	25
3. Formant Theory.....	26
1. The Resonator Output in the Periodic Case.....	28
a. Exponential input pulses.....	31
b. Critically damped input pulses.....	35
2. The Simple Resonator and Speech.....	39
3. Zero Crossing Frequency as a Measure of Formant Frequency.....	47
4. Highest Energy Harmonic as a Formant Frequency Estimator.....	57
5. Comparison of Errors in the Two Harmonic Trackers.....	61
a. Magnitude errors.....	61
b. Percent error of the input formant frequency.....	62
4. A Formant Estimation Technique.....	68
1. Isolation Technique.....	70
2. Processing Technique.....	74
3. Summary.....	76

Table of Contents (Continued)

	<u>Page</u>
5. Equipment Used to Implement the Formant Tracking Method....	78
1. Introduction.....	78
2. Audio Tape Recorder.....	79
3. Filter Bank.....	81
4. Multiplexer.....	82
5. Analog-to-Digital Converter.....	86
6. Tape Control.....	86
7. Digital Tape Unit.....	87
8. Digital Computer.....	87
9. Summary.....	88
6. Computer Program for First Formant Frequency Estimation....	89
1. Basic Machine Operation.....	89
2. Initialization of Machine for Current Record.....	92
3. Compute Filter Output Parameter Tables.....	94
4. Determination of Period Boundaries.....	102
5. Compute Accurate Zero Crossings and Make Estimates.....	106
6. Record Estimates.....	108
7. Formant Estimate Based on All Filters.....	111
7. Experimental Study of System Accuracy.....	120
1. Introduction.....	120
2. Variation as a Function of Fundamental Frequency.....	120
3. Accuracy of Formant Frequency Measurements.....	124
4. Summary.....	132
8. Results of Speech Formant Tracking.....	133
1. Analysis of Isolated Vowels.....	134
2. Analysis of an Isolated Diphthong.....	136
3. Analysis of CNC Words.....	139
4. Analysis of Speech Recorded in Normal Laboratory Noise.	143

Table of Contents (Continued)

	<u>Page</u>
9. Summary and Discussion.....	145
Appendices.....	150
Bibliography.....	188

List of Tables

<u>Table</u>		<u>Page</u>
3-1	Value of $2\pi kT$ as a function of f_0 and $k/2\pi$	48
3-2	Maximum error in cps for each particular harmonic for the zero crossing and the highest harmonic estimates of the input formant frequency.....	63
3-3	Maximum percent error in estimating the formant frequency as a function of harmonic number for the zero crossing counter and the highest energy harmonic tracker.....	67
6-1	ETAB values as a function of the preceding and following values.....	100
6-2	Conditions associated with each technique for estimating period boundaries.....	105
7-1	Formant frequency estimates as a function of fundamental frequency for a fixed formant frequency of 460 cps.....	123
7-2	Formant frequency calibration.....	127

List of Illustrations

<u>Figure</u>		<u>Page</u>
3-1	Simple Resonator.....	27
3-2	Periodically Recurring Sequences of Pulses.....	30
3-3	Vowel Formant Frequency-Bandwidth Data.....	41
3-4	Graph of $e^{-k\omega T}$	49
3-5	Steady State Waveform as a Function of T for $v = 500$ cps and $k = 50$ cps.....	53
3-6	Steady State Zero Crossing Frequency of a Single Formant as a Function of Period.....	56
3-7	Formant Frequency Estimated by a Highest Energy Harmonic Tracker as a Function of Fundamental Frequency.....	60
4-1	A Single Formant from Actual Speech Isolated by a Band-pass Filter.....	75
5-1	System Diagram for the Formant Tracker.....	80
5-2	Amplitude Spectrum of a Krohn-Hite Filter.....	83
5-3	Amplitude Response of Filter 2.....	84
5-4	Amplitude Response of Filter 3.....	84
5-5	Amplitude Response of Filter 4.....	85
5-6	Amplitude Response of Filter 5.....	85
6-1	Basic Program Flow Chart.....	90
6-2	Graph of the Six Input Channels.....	95
6-3	Speech Formant (Human) Showing ETAB Table Values at Axis Crossings.....	99
6-4	Computer Output Corresponding to the Analysis of a Single Filter.....	109
6-5	Summary Printout of Analysis of Each Individual Filter.....	112
6-6	List of First Formant Frequency Estimates.....	116

List of Illustrations (Continued)

<u>Figure</u>		<u>Page</u>
6-7	Graph of the First Formant Frequency [ɔɪ].....	118
6-8	Graph of the Fundamental Frequency [ɔɪ].....	119
7-1	System Diagram of Method Used to Measure Fundamental Frequency.....	121
7-2	Estimated First Formant Frequency as a Function of Fundamental Frequency.....	125
7-3	Formant Frequency Calibration: 225-375 cps.....	128
7-4	Formant Frequency Calibration: 375-525 cps.....	129
7-5	Formant Frequency Calibration: 525-675 cps.....	130
7-6	Formant Frequency Calibration: 675-825 cps.....	131
8-1	Graph and Spectrogram for an Isolated [ɪ].....	135
8-2	Graph and Spectrogram for an Isolated [U].....	137
8-3	Graph and Spectrogram for an Isolated [ɪ].....	138
8-4	Graph and Spectrogram for the Diphthong [aɪ].....	140
8-5	Graph and Spectrogram for the Word <u>Jug</u>	141
8-6	Graph and Spectrogram for the Word <u>Toad</u>	142
8-7	Graph and Spectrogram for the Word <u>Well</u>	144
A-1	Simple Resonant Circuit.....	151
A-2	Geometric Representation of the Relationships Between the Frequencies of a Simple Resonator when $x > 2k^2$...	154
A-3	Resonator Amplitude Response for Fixed Resonance Frequency and Three Different Conditions of Damping.....	156
A-4	Basic Input Pulses Investigated.....	159
A-5	Recurrent Sequences of Pulses.....	165

List of Appendices

<u>Appendix</u>		<u>Page</u>
A	A Simple Resonator.....	150
	1. Frequencies Associated with a Simple Resonator.	150
	2. Response of a Simple Resonator to Three Different Types of Pulses.....	158
	3. Response of a Simple Resonator to a Recurrent Sequence of Pulses.....	164
B	Mathematical Representations of Periodically Recurring Pulses.....	169
	1. Interval Representations.....	169
	a. Periodically recurring exponential pulses...	170
	b. Periodically recurring critically damped pulses.....	171
	c. Periodically recurring damped sinusoids.....	173
	2. Interval Steady-State Representation.....	178
	a. Periodically recurring exponential pulses...	179
	b. Periodically recurring critically damped pulses.....	179
	c. Periodically recurring damped sinusoids.....	180
	3. Representations on the Interval $(0, \infty)$	182
	a. Periodically recurring exponential pulses...	183
	b. Periodically recurring critically damped pulses.....	184
	c. Periodically recurring damped sinusoids.....	185
	4. Summary of Representations.....	187

Chapter 1

The Formant Tracking Problem

In 1895, Hermann [53] defined a formant as a natural mode of resonance of the acoustical cavity associated with the vocal tract. By 1906 Scripture [100] published his Researches in Experimental Phonetics which, among other things, included a description of the speech wave in terms of damped sinusoids. Scripture not only included equations for describing the speech wave in terms of damped sinusoids, but also presented and illustrated a graphical technique which he used for deriving the formant frequencies independently of the Fourier components. Scripture's emphasis on time domain analysis and his obvious use of stored information in the analysis has an interesting precedence over the present study. More recently, great advances in understanding the physics of speech have been made by Fant [30]. In his book, The Acoustic Theory of Speech Production, the relation between the physiological formations and the resulting acoustical signals is excellently portrayed.

Briefly, the physiological basis for a voiced oral vowel is that the glottis acts as a high impedance source and the vocal tract acts as an acoustical cavity. The main scientific advantages of discussing a voiced oral vowel is that the natural modes of resonance of the vocal tract are not modified by anti-resonances [33, 115]. In other types of

speech it is quite common to find anti-resonances as well as resonances. In an acoustical system, such as the average male vocal tract, there are infinitely many natural modes of resonance. Each of these has a bandwidth and resonance frequency associated with it. The frequency is the formant frequency and the bandwidth is the formant bandwidth. There is a natural ordering placed on the formants by their resonance frequency: the natural mode of resonance with the lowest frequency is called the first formant, the natural mode of resonance with the second lowest frequency is called the second formant, etc.

In analyzing voiced speech there is a strong temptation to use electronic equipment to resolve the harmonics associated with the rate of laryngeal excitation. The use of the harmonic structure of speech is quite convenient for some purposes: say, for estimating the fundamental laryngeal frequency. The line spectrum is not, however, the easiest representation to use in determining the formant frequencies. The reason is, of course, that the resonance frequencies associated with the vocal cavities are generally inharmonic to the fundamental frequency at which the true vocal folds are vibrating.

The problem in analyzing the formants is that speech is a transient phenomenon: during speech the fundamental laryngeal frequency and the acoustical configuration of the vocal tract are both undergoing considerable change. This creates a major analysis problem when one is trying to estimate the natural modes of resonance associated with the vocal cavities. For most adults in a normal speech situation, the fundamental laryngeal frequency is lower than the first formant frequency. In addition, the formants, being associated with the vocal

cavity, tend to move fairly slowly compared to the fundamental laryngeal frequency. Many different techniques of formant frequency analysis have been investigated by the use of filters, digital computers, or some combination of the two. The problem of formant frequency analysis for quasi-periodic speech waves continues to be a problem of major importance in speech research. Various techniques have now been developed, but certain problems or limitations are associated with all of them. Because of these problems, many of the schemes for formant frequency analysis do not provide particularly accurate measurements.

The ultimate accuracy required in formant frequency measurement, of course, depends upon the intended use of the data. Formant frequency analysis is currently one of the major problems in speech research. For instance, high quality formant tracking is needed for work on:

1. acoustic phonetics,
2. automatic speech recognition, and
3. formant encoded vocoders.

It seems clear that with better formant frequency analysis one will be able to obtain better acoustic phonetic specifications of spoken language and also be able to determine better the phonetic value of the input signals. The problem of poor formant tracking in formant encoded vocoders is sufficiently well documented in the literature that there is no need to belabor the point further.

This paper is concerned with the basic properties and with the analysis of formants in speech. The paper is organized according to three main areas: over-all perspective, theoretical considerations, and a scheme for measuring voiced speech formants.

Human speech is a very complex subject. Restricted models are employed to investigate speech, as in all studies of natural systems. The model used here is very much a function of the particular aspect of speech being investigated. To provide a broad perspective of the problem of speech analysis, a general discussion of some of the models used to study speech is presented in Chapter 2. To show the nature of the models, a discussion of the information content of the signals resulting from some of the models is included. This provides a numerical measure which shows the nature of the restrictions placed upon speech by the various models. Thereafter a survey is made of the previous techniques used in analyzing and/or representing speech formants.

In the frequency domain, the theoretical model for a single formant is a pair of conjugate poles. The model commonly used in studying a single formant is a simple R-L-C resonator where the output is across the capacitance. Appendix A of this study is a summary of the main properties of R-L-C resonators. In addition, several general equations are derived relating to simple resonators. The material in Appendix A is brief and is compiled in one place for convenient reference. Appendix B is a consideration of several equations which are used in the study of resonators. The material in Appendix B is included because it does not seem to be in the literature and because it is very necessary to the particular form of analysis which is discussed in this paper.

The actual analysis of the output of a simple resonator is treated

in Chapter 3. The analysis presented here differs from most speech analyses in that it is synchronous with the laryngeal pulses. Attention is then centered upon the parameter values associated with speech and with the resulting implications for the simplification of the previous analysis. Next, two forms of speech analysis are discussed in terms of the model. A theoretical comparison of the errors introduced by each of the systems is included.

Finally, a feasibility study is included for a fixed filter, pitch synchronous form of formant estimation. Chapter 4 consists of a heuristic description of the physical basis for the estimation technique. The feasibility study reported herein examined the possibility of automatically detecting the fundamental period and tracking the first formant frequency. The technique employed was based on the assumption that a formant estimate derived from a portion of a period would provide sufficient formant information for most uses. Thus, the sampling frequency for the formant measurements is the fundamental frequency. The approach employs a combination of analog and digital techniques. Speech is first processed by wide-band, overlapping filters, which are spaced so that all frequencies above 300 cycles occur in the pass-band of two different filters. Thus the analog part of the process is a form of frequency domain analysis. The formant estimation technique is a pitch synchronous technique in which successive samples of the speech wave are analyzed in the digital computer at the spacing of the fundamental voice frequency. Chapter 5 describes the equipment used in the feasibility study and Chapter 6 describes the general form of the

computer processing used in the formant estimation technique. In Chapter 7 two calibration runs on the system are described: one is to determine the results of varying fundamental laryngeal frequency and the other is to determine the error in actual formant frequency estimation at a given fundamental laryngeal frequency. The results of analyzing various speech inputs is presented in Chapter 8.

It is a conclusion of the study that a complete formant tracker could operate in the following manner. The speech wave would first be analyzed by a set of wide bandwidth overlapping filters. The outputs of the filters would then be multiplexed and quantized. Next, in a digital computer, the outputs of all of the filters which are in the range of the first formant would be processed to determine the approximate period boundaries and thereafter the first formant frequency. On the basis of the period boundaries and the first formant frequency, the higher formant frequencies could be determined. The knowledge of the period boundaries makes it possible to adjust the measurements to the wide variation in the excitation of the higher order formants. It also eliminates the problem created by the fact that there is no boundary below which only one particular formant occurs and above which only higher order formants occur. This present report is not concerned with the implementation of a complete system according to the above description; rather it thoroughly demonstrates the feasibility of determining the period boundaries and of tracking the first formant.

Chapter 2

Representation and Analysis of Speech Signals

The purpose of this chapter is to provide a perspective for the remainder of the report. Many different mathematical models are used to investigate speech, and they are necessarily restricted models. The general nature of some of the models which have been used to study speech are described in this chapter. In addition, the information content is given for the signals associated with some of the models. Finally, a survey of the research relating to formant measurement is presented.

Speech signals may be represented and/or analyzed at many different levels. While a general model of speech will include a specification of the relations among the different levels, a restricted model may be constructed for any particular level. Each such model describes a somewhat different set of constraints on the signal. For instance, one could model speech signals using only crude observations about them, or one could apply the results of the acoustical theory of speech production and consider the classes of functions implied by this theory. Similarly, there exist models of speech which are used in linguistics; utterances can be represented by strings of phones, phone types, or phonemes. These sequences can be further restricted to strings of syllables, morphemes, words, or phrases. Perhaps the day will come

when one can talk about meaningful strings associated with strings of morphemes, thus accounting for the semantic function of speech signals.

1. Information Content of Speech

One of the outstanding formalisms—if not the outstanding formalism—in communication theory is that by Shannon. Shannon's basic theory defines a concept, channel capacity, which can be used to determine the effectiveness of the utilization of a channel.

For instance, one may consider a speech wave as an acoustical signal with a bandwidth of 10,000 cps and a 50 db signal-to-noise ratio [52]. An increase in either the signal-to-noise ratio or the bandwidth fails to increase the intelligibility further. If the inherent redundancy in speech is ignored, then, according to Shannon's formula, the entropy associated with a 10,000 cycle bandwidth and a 50 db signal-to-noise ratio is 166,000 bits per second. Shannon's formula, of course, is based upon a white Gaussian signal and white Gaussian noise. As is well known, however, in the case of speech neither the signal nor the noise is white Gaussian noise.

One can also consider a voiced speech signal as consisting of a number of modulated carriers, where the carrier frequencies are harmonically related to a slowly varying fundamental frequency [27]. Fano obtained values from 3680 to 1560 bits/second as the channel capacity necessary for transmitting such signals.

The fact that the acoustic waves are produced by a particular mechanism provides the basis for another level of encoding speech. The physiological basis of speech has been known for some time, but it has

only been within the last few years that an Acoustic Theory of Speech Production [30] has been in existence. Flanagan [39] found that a channel capacity of 500 bits/second would be sufficient to transmit speech using a coding scheme based on the acoustic theory of speech production.

The phonetic sequences of speech form another level at which encoding may be analyzed; that is, the information contained in the phonetic representation of utterances may be used as a basis for determining the information rate of speech. For most purposes, however, a phonemic symbolization is used in representing speech. There are many types of encodings which could be used to reduce the bandwidth necessary for transmitting phonemic sequences. For instance, one could consider speech to be a k -stage Markov process. A single-state Markov model for phoneme sequences is one in which each phoneme is assumed to occur independently of the preceding phoneme and with a fixed probability. A Markov process is a possible model for this level in speech and one can use published data [112] to estimate the associated bit rate. With the assumption of a rate of 10 phonemes/second, the average rate for such a source is about 50 bits/second. If only a syllabic representation (various length strings of phonemes) is desired, rates of 23 bits/second [83] are possible for speech.

With either of the above representations, phonemes or syllables, one preserves only part of the linguistic information of the message: one loses the prosodies (or suprasegmentals) and part of the social-linguistic or personal information about the speaker [67].

By considering higher level linguistic information, one can construct an encoding which requires that messages correspond to a particular grammar. One experiment which showed that grammar provides a constraint on the perception of speech was performed by George Miller [77]. He used five sets of five words and sequences of words were then formed by selecting a word from each set. These sets of words were arranged in various orders. One way of ordering the sets resulted in "grammatical" sentences when a word was selected from each set in turn. The other orders for the sets of words formed ungrammatical sentences. A grammatical restriction on the speech source is inferred from the fact that the "grammatical" sentences were more intelligible. As another illustration, consider an acoustical signal corresponding to "I do not know." If one were only interested in sequences of words, there would be at least 16 possible strings of words corresponding to the same acoustical wave, for instance, "Eye due knot no." In this case, it should be clear that a grammar does reduce the number of possibilities for interpreting the acoustical signal. Since a complete grammar of English has not yet been constructed, it is not possible to state a bit rate for signals on which the constraints of English grammar have been imposed.

Finally, it may be observed that semantics, too, places a constraint upon the bit rate of signals. The semantic restrictions are exceedingly difficult to specify, and certainly no data on bit rates under this additional constraint are yet known.

Whatever the constraints, there appears to be an upper limit to the bit rate at which the human can interpret the information received. It has been stated that: "For sustained communication beyond a few seconds, the practicable rates for most human activities have been closer to 10 bits/second than to 40 bits/second." [61]. Since speech is generally produced by, and intended for, humans, it would seem that 10 bits/second might well represent an average rate at which information can be interpreted in speech communication.

As indicated above, if the speech signal is considered to be totally unstructured (white Gaussian noise) it involves bit rates of the order of 166,000 bits/second. This is the informational capacity required of a channel to transmit such a signal. The ear performs both an analysis of the acoustical waves and a conversion of the results to neural impulses. The auditory branch of the eighth cranial nerve is undoubtedly capable of a relatively high information rate. It must be, then, that a large part of the information reduction of the signal occurs at higher levels in the central nervous system. At this stage of knowledge about the processing of information by the central nervous system, it is difficult even to speculate on the way in which the decoding and message interpretation takes place. Apparently there is an information reduction in the interpretation of such complex signals as speech. It seems reasonable to assume that the signal is reduced to its information-bearing parameters, and that interpretations are then performed on these parameters. It is presumably in these latter stages of the signal processing that phonemic, grammatical, etc. constraints are imposed. It would appear, however, that if the concept

of message destination can be defined in this process, an information rate of the order of 10 bits per second is a reasonable estimate.

It seems reasonable to conclude that it should be possible to encode speech for transmission or for human interpretation at a relatively low information rate. Clearly, there is a wide range of effective encodings which might be employed for speech. Equipment complexity and time delays introduced by the equipment are of considerable concern in the process of such encoding. It is thus important to consider the relation between these factors and the expected reduction in bit rate. In addition, there is the problem of determining the properties of the signal to be communicated, i.e. of defining the message. In some cases it is desirable to include speaker information, such as identity, mood, etc. In other cases the written equivalent of the message is sufficient. In many speech compression schemes an attempt is made to preserve all of the information which may be relevant. Accordingly, an attempt is made to implement substantial compression without affecting the signal in a way that is noticeable to a listener.

2. Techniques Used in Speech Analysis

The purpose of this section is to present a brief review of some of the acoustic analysis techniques which have been applied to speech formant analysis. An identifying name has not been well established for certain of the forms of analysis. In an attempt to catalog some of the major techniques, however, particular terms have been used in the following pages. These terms do not apply to disjoint techniques:

sometimes several of the terms can be applied to a single technique. It is not practical to discuss each technique in detail in the present report. Also, the examples given of each technique are certainly not meant to be exhaustive; they are intended only to show some of the diversity which has been achieved in the analysis of speech.

a. Spectrographic analysis

The sound spectrograph was the first analyzer developed which provided both a convenient and a meaningful graphic display for research studies on speech. The instrument was designed at the Bell Telephone Laboratories by Koenig, Dunn, and Lacy [63]. The spectrograph analyzes waves and produces permanent visual recordings showing the relations between time, frequency, and energy. The spectrograph is still the most commonly used device for speech analysis [72]. It is especially valuable because of the effectiveness with which it displays the continuity of formant movements. The accuracy of measurements from this device has been investigated. Lindblom found that there is a mean error of 40 cps in formant estimation from spectrograms using normal techniques [73]. One type of spectrographic output, an amplitude section, allows a simple measuring technique which is generally felt to improve the estimation accuracy. This measuring technique uses pre-calculated formant envelopes. Using templates containing such envelopes, Dunn made a large number of measurements of formant bandwidth [25]. His article contains an excellent figure which shows the general tendency to underestimate the position in frequency of the true formant peak when a template is not used.

b. Filter banks

Filter banks provide a very common and convenient means of spectral processing. Flanagan [39] has shown two different techniques for estimating formant frequency from the output of a filter bank. One of these methods employs successive time samples which consist of the short-time spectrum of speech. Flanagan designed circuits which operate on each of these samples to determine the spectral maxima. This technique is called spectrum sampling. The second technique which Flanagan used is called a spectrum-segmentation formant extractor. The essence of this technique is to pre-divide the audio range into a first, second, and third formant range and then to use the previous technique within each of the ranges. In a companion paper [40] it was reported that the spectrum segmentation technique is the better of the two. With the latter technique the results were within 150 cps of the first formant 93% of the time and within 200 cps of the second formant 91% of the time. The measurements were also simultaneously within 150 cps of the first formant and 200 cps of the second formant 85% of the time.

A filter bank is the basic input unit for the work on speech recognition conducted by Forgie and Forgie [44] at Lincoln Laboratories. They use the peaks in the output of a filter bank as an estimator of formant frequency and only use higher resolution when necessary to reduce ambiguity.

Houde, Larkin, and Stewart [55] have made an attempt to use a local spectrum-matching technique in order to achieve an estimate of the formant frequencies without employing a detailed analysis by

synthesis.

Harris and Waite [49] have determined the response of several types of filter banks to vowels. They considered the use of single tuned, double tuned, Butterworth, and Gaussian filters. They investigated the implications for vowel analysis of using both constant bandwidth and constant mel¹-bandwidth filters with and without frequency equalization.

c. Zero-crossing measurements

Fant [30] has observed that

From an engineering point of view it seems rational to measure and to define the formant from the time function. Given a high speed oscillogram of the speech wave, with a prefiltering arranged to isolate the formant of interest, the formant frequency can be measured from the time of a full period of oscillation.

Thus, one zero crossing (synonymously, axis crossing) technique is to isolate the formant of interest and to measure its period. At least one system of this sort has been designed [5, 6] and a description of the equipment employed has been presented. The first of the two preceding references is to a paper presented at a meeting of the Acoustical Society of America. In the system described, there is a filter with cut-off points which can be varied to some extent.

¹The Mel scale is a perceptual scale based on observer responses to frequency relationships. Constant mel-bandwidth filters are of equal width along the mel scale, but differ in width on a linear frequency scale.

One zero crossing scheme has been described on which work was started at the Bell Telephone Laboratories before World War II [79]. With this technique the speech was filtered into three bands: 300 cps to 1100 cps, 900 cps to 3000 cps, and 3000 cps to 8000 cps. The axis crossing rates of the outputs of these filters were then used to specify the acoustical signal.

In 1951, E. Peterson, of the Bell Telephone Laboratories, used an axis crossing counter as a means of obtaining frequency information about speech [86]. His goal was to obtain an automatic formant tracker. He showed the operation of the circuit on speech which had been prefiltered into two bands. Results were also plotted for two sine waves as input and an appendix shows the theoretical operation for two sine waves. Peterson found that using the signal to modulate a carrier and then counting axis crossings improved the measurements during the transfer in a particular filter from one sine wave frequency to the other. Modulation has also been discussed as an aid to formant tracking by Cherry and Phillips [10] in their paper on the possible application of single side-band techniques to formant tracking.

The axis crossing method was also the main spectral estimator used in a digit recognizer completed in 1952 [19] at the Bell Telephone Laboratories. The input speech was divided into two frequency bands, one above and one below 900 cps. The axis crossing rates of these bands were made into two-dimensional patterns and the recognizer picked the digit which had the closest match to the input pattern. The system achieved a recognition rate of approximately 97 to 99% for a speaker

that it "knew," although it had difficulties with new speakers.

Axis crossing information was used in a speech compression system designed at Northeastern University. The investigators designed a ρ_0 meter which determines one-half of the input axis crossing rate [8, 56] and uses this information to position a moveable filter [23]. Then the output of the moveable filter is used as the input to a second ρ_0 meter which is used to determine the frequency of the formant.

Zero crossing rates are also used in a Japanese speech recognizer [96, 97, 98] as a means of obtaining spectral information about speech waves.

d. Diversity approaches

The term "diversity approaches" comes from a paper by David [16], and provides an excellent description of a class of analysis schemes. These schemes are grouped together because they attempt to operate on the gross structure rather than the fine structure of the input signal. The fact that the techniques are concerned with the gross structure of the signal does not imply inaccuracies in the results; it simply specifies another approach to the analysis of signals. Some of the techniques characterized as diversity approaches will next be described.

Sebestyen [102] has presented a particular approach to the general problem of the recognition of membership in classes where the only known description of each class is by several examples of its elements. The theory which he developed was applied to the particular problem of recognizing spoken numerals. Sebestyen [101] has also applied his

theory to the problem of speaker recognition. Such classification may be employed in the use of adaptive procedures for speech recognition. Talbert, Groner, and Koford [110] have simulated an adaptive speech recognition scheme using "Adaline" threshold logic devices. A different adaptive system employing "Adaline" elements for automatic speech recognition has also been investigated by Dammann [14].

Olsen [82, 83, 84] and his co-workers at the RCA Laboratories have developed a series of phonetic typewriters which employ a filter bank for the analysis of the speech wave. They use the syllable as the input unit to their systems. The analysis uses an adaptive time encoding of the spectral patterns. Thus the interpretation is not based upon spectra derived at fixed time intervals. Each input syllable is transformed into an 8×5 binary matrix. There are eight fixed filters which span the frequency range of interest and five time samples are taken as the spectral pattern changes. Each matrix element is "on" or "off" according to a threshold level on the amplitude of the filter output. Recognition is achieved by comparison with ternary matrices: must be on, must be off, and don't care.

Davis [18] has built a recognizer in which reference spectra are correlated with input spectra in order to determine which phonetic element is present.

Wiren and Stubbs [118] built an automatic system for phoneme classification. The system represented an attempt to implement the distinctive features phonemic theory for recognition of spoken American English. The instrumentation involved weighted band-pass filters to complete the classification.

Kersta [16, 62] has done work on several types of recognition problems using the diversity approach. In 1947, he devised a binary matching scheme for digit recognition. He divided the spectrogram into time-frequency sections, 200 cps by 67 milliseconds. He then made each section black or white, according to whether the average darkness of the section was greater than one-half of the average darkness of the darkest section. This study has historical significance as well, since it appears to be the first speech study to have used data processing equipment. Another technique used by Kersta is contour spectrographic analysis which he employs for voice-print identification. The first use of contour analysis was made by Koenig and Ruppel [64].

e. Spectral moment calculation

Moment calculations have been proposed as a means of extracting formant frequencies [80, 109]. One motivation for such a procedure is to achieve a compromise between the accuracy and the time required for measurements with the very complex and the very simple techniques which have been used for formant frequency extraction. In order to obtain greater accuracy, some moment calculation systems have used a filter bank as an input to a computer. The computer was used to process the input to determine spectral regions in which the formants were isolated. Then the computer found the moment of the energy in the isolated region.

f. High energy harmonic trackers

There are at least two techniques that could be used to track the highest energy harmonic. One would be to take a time sequence of harmonic spectra and operate upon each successive element of the time sequence to obtain a trace of the high-energy harmonics. Alternatively, one could build a device which developed a voltage proportional to the frequency of the highest energy harmonic within a specified frequency region.

Hellwarth [52] has described a feedback system in which a modulated signal is moved across a filter to produce a voltage which is proportional to the highest energy harmonic in the signal. The carrier of the modulator is continually adjusted so that the highest energy harmonic is passed through the filter for the single sideband modulator. The voltage obtained from measuring the frequency of the output of the filter not only specifies the harmonic frequency but also is used to control the carrier. One use of this system, other than for formant frequency estimation, is for psychoacoustical studies of band-passed speech in which the frequency band is continuously adjusted according to the location of a particular formant. Filters may be placed at the input to the system so that the device will usually track the highest harmonic in the region of either a first, second, or third formant. The output of a set of such heterodyne filters could also be used to implement a speech compression system. In this case a logical constraint may be imposed on the location of the filters so that at any given time each filter is assigned to the appropriate frequency region for a particular formant.

Hughes [59] used a bank of 35 filters as input to his speech recognition system. He employed a computer to determine which filter had the highest energy as well as to locate the various formants; he used the center frequency of the filter as the estimate of the formant frequency. The bandwidths of the filters were relatively narrow: 50 cps at the lowest frequency, 50 cps at 493 cps, 100 cps at 940 cps, 200 cps at 1800 cps, with increasing bandwidths to 1000 cps at 4500 cps.

g. Phase techniques

A few years ago Huggins [58] developed a phase technique for analyzing various types of simple resonance phenomena. The scheme is dependent upon the phase properties of a resonator as a resonance frequency in the signal is swept past the resonance frequency of the analyzer. Using this phase technique Huggins was able to develop a visible speech display, called a resnogram, which is considerably less sensitive to the nature of either the source or the amplitude of the input wave than is the spectrograph. The display has the property that the representations of the resonances are not as broad as those on a broad band spectrum.

Krokstad [65], at the Royal Institute of Technology, evaluated three different phase techniques as means of tracking formants. In one of these techniques the input and output of a filter were compared to obtain a measure of the phase relation between them. The other two techniques involved a measure of the phase difference between the outputs of two distinct filters; one set was separated by 250 cps

and the other was separated by 100 cps. The general conclusion was that the use of a set of two filters offered the better possibility of development for automatic formant tracking.

h. Analysis by synthesis

Analysis by synthesis is a technique for speech analysis proposed by Stevens [1, 48, 85, 106]. It is a method for reducing the speech wave to certain of its essential parameters. The technique involves the comparison of sampled speech spectra with synthetic spectra which are generated according to rules based on an acoustic theory of speech production. The input to the system is a 36 element filter bank spanning a 7000 cycle range. Samples are taken every 8.3 milliseconds in the current system [85]. A computer has now been programmed to use automatic matching strategies in the implementation of this technique.

i. Pitch synchronous techniques

A few years ago Mathews, Miller, and David [75] performed a pitch synchronous analysis of voiced sounds. They used visual inspection of an oscillogram as a means of determining the period. The actual samples analyzed consisted of the wave from the first axis crossing before a major peak to the first zero before the major peak in the following period. The line spectra of these samples were then computed and compared with known spectra in order to determine the spectral description of the sample.

Pinson [92] has developed a weighted-least-squares fitting procedure on fundamental periods for determining formant frequencies.

He obtains a best fit for a function of the form

$$f(t) = a_0 + \sum_{i=1}^n e^{-\pi B_i t} (a_i \cos 2\pi F_i t + b_i \sin 2\pi F_i t).$$

The fit is evaluated as a function of the variables a_0 , a_1 , B_1 , b_1 , and F_1 . His estimates are based only on those portions of the acoustic waveform which are produced when the glottis is closed and the lower tract is decoupled from the system. In this way Pinson avoids averaging over both open and closed glottis conditions. Since bandwidth is another parameter which Pinson estimated with this technique, it was important that he restrict his measures to the closed vocal tract configuration.

Manley [74] has obtained pitch synchronous representations in terms of fixed orthogonalized exponentially damped sinusoids. This procedure avoids the difficulty of determining the angular frequency associated with each formant at each successive period of the speech wave. The sampled speech wave, as well as period markers, are used by Manley as the input for a computer program which calculates the coefficients in terms of the basis functions.

The possible use of orthogonalized exponentials has been well described by Huggins [57]. Dolansky [22] has also considered this set of functions. Since the pressure wave associated with a voiced non-oral vowel is a square summable function for each period, it is possible to use any of the mathematically complete sets of orthogonal functions to represent a single period. The study reported above

used one such set. There are many other sets of functions which could be used. A general introduction to this type of approximation has been given by Clement [11], who shows several sets of functions which have realizable filters for the basis set. Square summable representations are described in detail in many mathematical texts [12, 60].

Most of the analysis schemes which use an error criterion are based on a mean square error; that is, the mean of the error squared is minimized. One alternative is to minimize the maximum error. This criterion leads to Chebyshev approximations. As the result of a paper, "Synthesis of R-L-C Networks by Discrete Tschebyscheff Approximations in the Time Domain," [95], it is clear that this form of analysis could be applied to speech. Swigert, at the Communication Sciences Laboratory, attempted to implement such a system and found that a great deal of machine time and input accuracy was necessary for the analysis. One major problem was a large matrix inversion. It was estimated that with the particular program used, an isolated pitch period would require at least one minute to analyze on an IBM 7090 computer. In the work by Swigert the approximation was made using exponentials. This type of approximation is described in the mathematical literature [21, 94], and possibly it should be reconsidered as a means of analyzing speech signals.

j. Inverse filtering

The general basis for inverse filtering is that it is possible to perform an operation which is the inverse of that performed by the vocal tract. This operation can be accomplished either by analog

circuits or by a digital computer. Miller [78] was the first to use the technique. His primary interest in the experiment was in the vocal cord wave shape. If the proper settings of inverse filters are obtained, however, one can infer the formant frequencies. Inverse filtering has been studied in many laboratories. Lawrence [68] has worked on this approach, as has Fant. Fant's Quarterly Progress Reports contain a great deal of information about particular inverse filters. Also David [15] has shown how a computer can be used to remove the poles and leave the spectrum associated with the glottal function.

3. Discussion

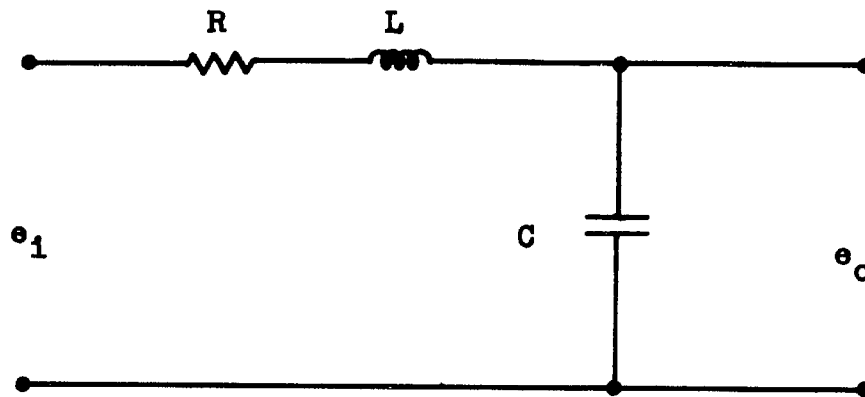
The previous pages illustrate some of the numerous approaches which have been used in the analysis of the speech wave. In general, the chief problem is to derive the fundamental period and the formant frequencies from the quasi-periodic waveforms of speech. As indicated earlier, many of the techniques which have been investigated are in some manner related. The fact that so many different methods of analysis have been tried, however, suggests that there are basic problems in formant frequency analysis which have not yet been resolved. In fact, there continues to be a need for an accurate and dependable formant analyzer. Such an analyzer must operate on a large variety of waveforms, and must function under a considerable range of harmonic and formant amplitudes.

Chapter 3

Formant Theory

Electrical analogs have long been of great value in the analysis of mechanical and acoustical systems. In the case of speech it has been found very convenient to simulate the passive part of the human vocal tract with an electrical ladder network in which each section of the network represents an acoustical section of constant length. It is sometimes overlooked that the dissipation in the acoustical sections is actually a function of frequency, whereas the dissipative elements normally inserted in the electrical networks are not a function of frequency. Nevertheless, rather good approximations have been achieved by electrical analogs. An analysis of the transfer function of the passive part of the human vocal tract shows that it can also be approximated by a series of isolated electrical resonators. Much can be learned about the properties of the individual formants in speech by a detailed consideration of their electrical simulation. Figure 3-1 shows a single electrical resonator circuit for simulating a formant. This is the basic model used by Fant in his book, and it is generally considered to have the characteristics of the formants of speech.

In this chapter the properties of such a circuit will be analyzed. The response of the circuit to various inputs will be derived, and these results will be related to certain properties of speech formants. The



Simple Resonator

Figure 3-1

response of the circuit will be determined when it is excited by pulses whose epochs are equally spaced in time; such sequences of pulses will be referred to as periodically recurring sequences of pulses. Actually, the pulses are not strictly periodic since periodic implies for all time, not for some arbitrary interval of time. By starting the pulses at a particular point in time, however, it will be possible to consider the representation of the output of the circuit and to ascertain its convergence properties. The representation of the output will be used later in this chapter to show some of the formal implications of certain formant frequency measuring schemes.

1. The Resonator Output in the Periodic Case

The purpose of this section is to obtain an expression for the output of the simple resonator corresponding to two different recurring sequences of input pulses.

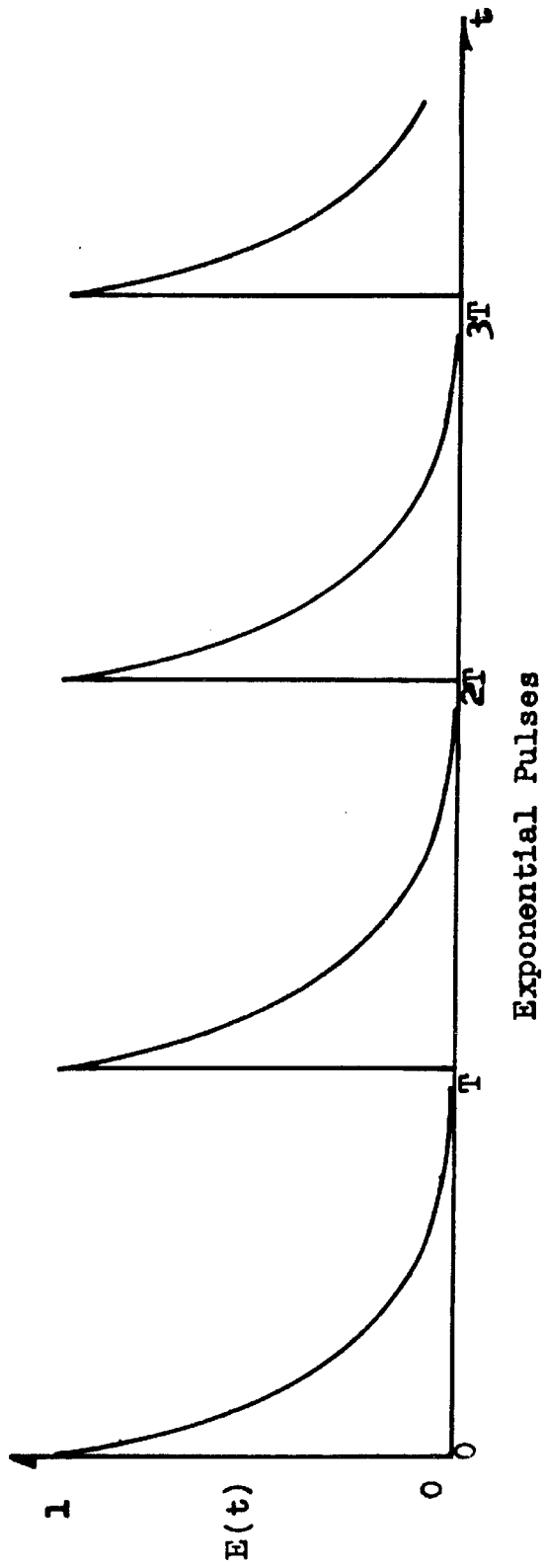
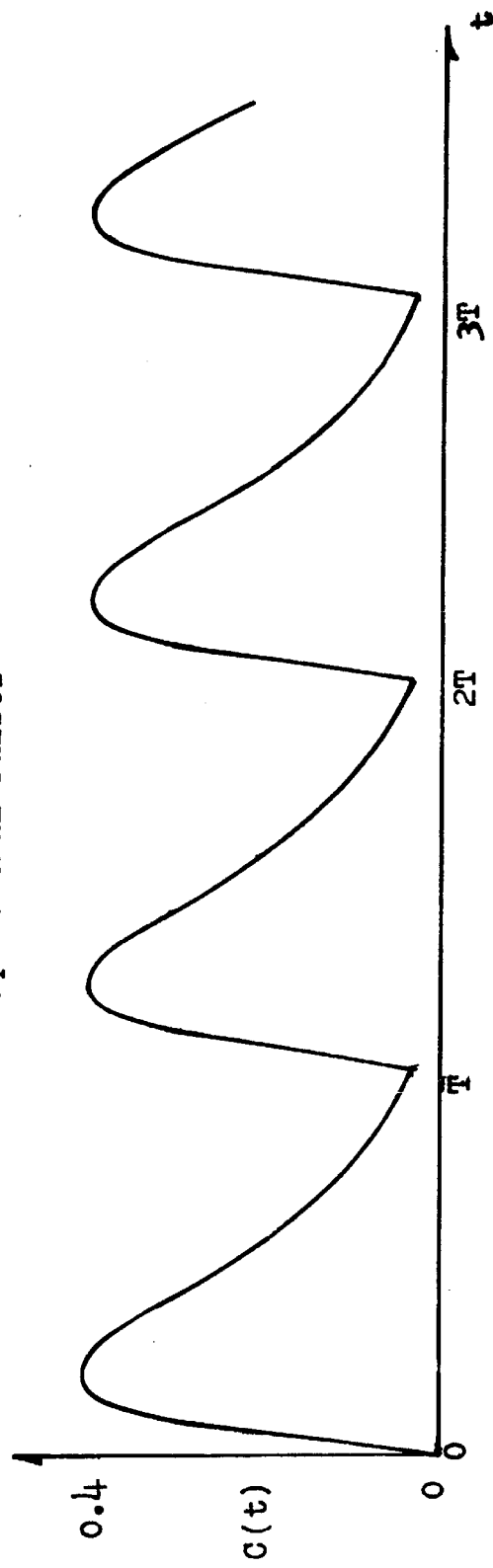
Previous analyses have considered this same problem. Gandhi, Peterson, and Yu [46] determined the representation of the output of the resonator for a periodically recurring sequence of rectangular pulses and a periodically recurring sequence of exponential pulses. The output was described with standard Fourier series techniques, and was represented by a transient term and a periodic term, where the periodic term was a Fourier series. In the above-mentioned study it was noted that the equations for the steady-state portion "did not contain the resonator frequencies" and that "in speech analysis these frequencies can be recovered only by some nonlinear analysis of the wave." It seemed desirable to obtain representations of the output of the resonator which

include the resonator frequency. The representations which will be derived have the advantage of being realizable by one form of nonlinear analysis as well as being simple enough to be heuristically informative about the nature of the process which is taking place.

The mathematical representations which will be developed contain damped sinusoids with the formant frequencies; this allows an exact formulation of some of the operations which were mentioned in the last chapter. The previous report [46] suggested that one very promising technique would be an analysis which is synchronous with the fundamental voice frequency. For this type of analysis, simple representations of the individual periods of the speech wave are particularly valuable.

In this section two different periodic sequences of pulses will be considered. These are shown in Figure 3-2. The first set of pulses was used by Dudley, et al. [24] as early as 1939. The current feeling, however, is that the second set of pulses gives a much better representation of a real glottal source [30].

Since considerable computation is involved in analyzing the response of the resonator to such pulses, most of the details have been relegated to appendices. Appendix A contains a derivation of the output of the resonator corresponding to a recurring input, and Appendix B contains a derivation of the interval representation of the three sums which occur in the output. The remainder of this section is devoted to obtaining closed form interval representations for the output of the resonator when excited by each of the above sequences of pulses.



Critically Damped Pulses
Periodically Recurring Sequences of Pulses

Figure 3-2

a. Exponential input pulses

In Appendix A (Equation A.49) a representation of the output of a simple resonator corresponding to an arbitrary recurrent sequence of exponential pulses is derived. For the special case where the pulses are periodic the equation becomes:

$$e_E(t) = \sum_{n=0}^{\infty} U(t-nT) [E_1 e^{-\lambda(t-nT)} - E_2 e^{-k(t-nT)} \sin[v(t-nT) + \phi_e]], \quad 3.1$$

where

$$E_1 = \frac{x}{[(\lambda - k)^2 + v^2]}, \quad 3.2$$

$$E_2 = \frac{x}{v[v^2 + (\lambda - k)^2]^{1/2}}, \quad 3.3$$

$$\tan \phi_e = \left[\frac{v}{k - \lambda} \right], \quad 3.4$$

$$k = \frac{R}{2L}, \quad 3.5$$

$$x = \sqrt{\frac{1}{LC}}, \quad 3.6$$

$$v = [x - k^2]^{1/2} \quad 3.7$$

T is the period, U(t) is the unit step function, and the input pulse is:

$$e(t) = e^{-\lambda t}$$

The output of a single resonator corresponding to a recurring sequence of exponential pulses applied at the input can be expressed for any given period of the input pulse train. The following expression is based on Equations B.9 and B.41 of Appendix B.

$$e_E(\tau + [m-1]T) = E_1 \xi_m e^{-\lambda \tau} - E_2 \alpha_m e^{-k \tau} \sin[\nu \tau + \phi_e + \beta_1 - \beta_m] \quad 3.8$$

where m is the period for which the representation is valid,

$$t = \tau + [m-1]T, \quad 3.9$$

$$\xi_m = \frac{[1 - e^{-\lambda mT}]}{[1 - e^{-\lambda T}]}, \quad 3.10$$

$$\alpha_m = \frac{[1 - 2e^{-kmT} \cos \nu mT + e^{-2kmT}]^{1/2}}{[1 - 2e^{-kT} \cos \nu T + e^{-2kT}]^{1/2}}, \quad 3.11$$

and

$$\tan \beta_m = \frac{\sin \nu mT}{e^{kmT} - \cos \nu mT}. \quad 3.12$$

The other terms are as previously defined. The variable τ is introduced to make time (i.e. τ) zero at the start of the m^{th} interval. This allows a direct comparison of the representation on various intervals. Equation 3.8 is a representation which shows the effects of the "filter-pulse" constants E_1 , E_2 , and ϕ_e as well as the effects of the recurrence constants, ξ_m , α_m , and β_m . That is, the above equation shows the interaction between the constants determined by the pulse and the filter and the constants determined from the fact that there had been $[m-1]$ pulses previously.

Steady-state behavior would mean that on each interval the representation would be the same as a function of τ . An easy means of determining the steady-state solution is to consider the limit of Equation

3.8 as a function of m . Letting $\bar{P}_E(\tau)$ be the steady-state output on an interval, it follows that

$$\begin{aligned}\bar{P}_E(\tau) &= \lim_{m \rightarrow \infty} e_E(\tau + [m-1]T) \\ &= \xi E_1 e^{-\lambda \tau} - E_2 \alpha e^{-k\tau} \sin[v\tau + \phi_e + \beta_1]\end{aligned}\quad 3.13$$

where ξ and α are the limits of the sequences ξ_m and α_m , respectively: that is:

$$\xi = \frac{1}{1 - e^{-\lambda T}}, \quad 3.14$$

$$\alpha = \frac{1}{[1 - 2e^{-kT} \cos vT + e^{-2kT}]^{1/2}}, \quad 3.15$$

and the other terms are as previously defined. The above simple representation is an exact representation of the output in the steady-state, just as Equation 3.8 is a simple representation of the output which is exact during the m^{th} period. It allows one to appreciate easily the nature of the waveform during a period, whereas the nature of the waveform is not easily grasped when the representation contains an infinite sum.

Clearly, the input to this resonator is not periodic. However, the output can be represented as the sum of a periodic and a transient term. On the first period the periodic term is given by:

$$0 \leq t < T \quad \bar{P}_E(t) = E_1 \xi e^{-\lambda t} - E_2 \alpha e^{-kt} \sin(vt + \phi_e + \beta_1). \quad 3.16$$

From Equations B.53 and B.62 it follows that the transient term, $\bar{T}_E(t)$ is given by

$$\bar{T}_{e_E}(t) = -e^{-\lambda T} \xi E_1 e^{-\lambda t} - \frac{E_2 e^{-kt}}{\mu_1} \sin [vt + \phi_e - \gamma_1] \quad 3.17$$

where

$$\mu_1 = [1 - 2e^{kT} \cos vT + e^{2kT}]^{1/2} \quad 3.18$$

and

$$\tan \gamma_1 = \frac{\sin vT}{e^{-kT} - \cos vT}. \quad 3.19$$

A knowledge of the transient form with respect to the periodic form is quite useful in determining how soon the transient contribution can be ignored as a function of time. This condition can also be deduced from the interval representation of Equation 3.8.

Using Equations 3.16 and 3.17, the output of a resonator corresponding to periodically recurring exponential input pulses can be written:

$$0 \leq t < \infty \quad e_E(t) = \bar{T}_E(t) + \bar{P}_E(t). \quad 3.20$$

Three exact closed form representations have been found for the output of the simple resonator when the input is a periodically recurring sequence of exponential pulses: a representation valid for any period, a steady-state representation on each period, and a form which shows the output as a periodic and a transient term.

b. Critically damped input pulses

The next case to be considered is the representation of the output when the input is a periodically recurring sequence of critically damped pulses. The output of the resonator corresponding to a recurrent sequence of critically damped pulses is given by Equation A-50. For the special case of recurring critically damped pulses which are also periodic, the equation can be written as:

$$E_C(t) = \sum_{n=0}^{\infty} U(t-nT) [C_1 e^{-\lambda(t-nT)} + C_2 (t-nT) e^{-\lambda(t-nT)} + C_3 e^{-k(t-nT)} \sin[v(t-nT) + \phi_c]], \quad 3.21$$

where

$$C_1 = \frac{2x[\lambda - k]}{[(\lambda - k)^2 + v^2]^2}, \quad 3.22$$

$$C_2 = \frac{x}{[(\lambda - k)^2 + v^2]}, \quad 3.23$$

$$C_3 = \frac{x}{v[(\lambda - k)^2 + v^2]^2}, \quad 3.24$$

and

$$\tan \phi_c = \frac{2v(k - \lambda)}{[(\lambda - k)^2 - v^2]}. \quad 3.25$$

Using the representations for such sums as are derived in Appendix B (Equations B.9, B.14, and B.41), the above expression can be written for the m^{th} period in the following manner:

$$\begin{aligned}
E_C(\tau+[m-1]T) &= C_1 \xi_m e^{-\lambda\tau} + C_2 [\xi_m \tau e^{-\lambda\tau} + \rho_m e^{-\lambda\tau}] \\
&\quad + C_3 \alpha_m e^{-k\tau} \sin[v\tau + \phi_c + \beta_1 - \beta_m] \\
&= [C_1 \xi_m + C_2 \rho_m] e^{-\lambda\tau} + C_2 \xi_m \tau e^{-\lambda\tau} \\
&\quad + C_3 \alpha_m e^{-k\tau} \sin[v\tau + \phi_c + \beta_1 - \beta_m]
\end{aligned} \tag{3.26}$$

where

$$\rho_m = T e^{-\lambda T} \frac{[1 - m e^{-\lambda[m-1]T} + [m-1] e^{-\lambda m T}]}{[1 - e^{-\lambda T}]^2} \tag{3.27}$$

and where the other terms have been previously defined.

The above expression has the same possible interpretation as did the interval representation found for the case of periodically recurring exponential pulses; namely, it shows the interaction of the effects of the pulse, the filter, and the pulse recurrence. That is, Equation 3.26 shows the output of the resonator for the m^{th} period in such a way that the effects of the basic pulse-filter coefficients, C_1, C_2, C_3 , and ϕ_c are separated from the recurrence components, ξ_m, ρ_m , and β_m .

As indicated previously, steady-state behavior means that there is a limiting behavior on each period. An easy means of determining the steady-state behavior is to consider the limit of Equation 3.26 as a function of m . Letting $\bar{P}_C(\tau)$ be the steady-state output on such an interval, it follows that

$$\begin{aligned}
\bar{P}_C(\tau) &= \lim_{m \rightarrow \infty} e_C(\tau + [m-1]T) \\
&= [C_1\xi + C_2\rho]e^{-\lambda\tau} + C_2\xi \tau e^{-\lambda\tau} \\
&\quad + C_3\alpha e^{-k\tau} \sin[v\tau + \phi_c + \beta_1]
\end{aligned} \tag{3.28}$$

where

$$\begin{aligned}
\rho &= \lim_{m \rightarrow \infty} \rho_m \\
&= \frac{T e^{-\lambda T}}{[1 - e^{-\lambda T}]^2}
\end{aligned} \tag{3.29}$$

and all other terms are as previously defined.

The above equation shows that during each period the steady-state output of a resonator, resulting from a recurrent sequence of critically damped input pulses, consists of an exponential pulse, a critically damped pulse and a damped sinusoid. It also exhibits the amplitude coefficients of each as the product of a term associated with the basic pulse-filter configuration and a term associated with the nature of the recurrence. The phase angle is expressed by a sum consisting of a term which reflects the basic pulse-filter configuration and a term which reflects the nature of the recurrence.

An input of periodically recurring critically damped pulses is not a periodic input. However, the output can be written as the sum of a periodic and a transient term. On the first period the periodic term, $\bar{P}_C(t)$, is given by:

$$\begin{aligned}
0 \leq t < T \quad \bar{P}_C(t) = & [C_1 \xi + C_2 \rho] e^{-\lambda t} + C_2 \xi t e^{-\lambda t} \\
& + C_3 \alpha e^{-kt} \sin[\nu t + \phi_c + \beta_1]
\end{aligned} \tag{3.30}$$

Using Equations B.53, B.58, and B.62, it follows that the transient term, $\bar{T}_C(t)$ is given by

$$\begin{aligned}
\bar{T}_{e_C}(t) = & - [C_1 e^{-\lambda T} \xi + C_2 \rho] e^{-\lambda t} - C_2 e^{-\lambda T} \xi t e^{-\lambda t} \\
& - \frac{C_3 e^{-kt}}{\mu_1} \sin [\nu t + \phi_c - \gamma_1]
\end{aligned} \tag{3.31}$$

where all terms have been previously defined.

It is not surprising that the transient term also consists of the sum of an exponential pulse, a critically damped pulse, and an exponentially amplitude modulated sinusoid. In the above, it is easily seen that the terms can be separated into those which result from the pulse-filter configuration and those which result from the pulse recurrence. Using Equations 3.30 and 3.31, the output of a resonator corresponding to periodically recurring critically damped pulses is given by

$$0 \leq t < \infty \quad e_C(t) = \bar{T}_C(t) + \bar{P}_C(t). \tag{3.32}$$

The same three types of representations have been derived for the periodically recurring critically damped pulses as were obtained for the exponential pulses.

2. The Simple Resonator and Speech

The theory of recurrently impulsed resonators has now been developed and simple representations have been determined for the output corresponding to two different periodically recurring sequences of input pulses. In all cases the representations were exact and there was no attempt to make approximations. The remainder of this section is devoted to applying some of the previous results to speech.

In order to consider speech we shall first make use of some of the measurements which have been reported by Dunn [25]. He has recently determined the formant frequency and bandwidth for 400 of the vowels that were used in the Peterson-Barney study [91] at the Bell Telephone Laboratories. In Dunn's article the data were presented in several forms, including the average bandwidth (for each of the first three formants) as a function of average formant frequency for each of the vowels. Since the current interest is in the signal structure of the formants, an alternate form has been adopted for presentation. Figure 3-3 shows one-half the bandwidth, $B/2$, plotted as a function of the frequency of the spectral amplitude density maximum, $\omega_{\max}$, for each formant measured by Dunn. That is, the graph in Figure 3-3 shows all those points $(B/2, \omega_{\max})$ which were observed at least once in Dunn's work. As was mentioned in Appendix A, the damping coefficient, k , is frequently (and closely) approximated by $B/2$ when the bandwidth is expressed in radians per second.

For the simple resonator discussed in Appendix A the following conventions were used: k is the damping coefficient, $x^{1/2}$ is the natural

frequency, v is the resonance frequency, and $\omega_{\max}$ is the spectral maximum. The basic relations among these variables are shown graphically in Figure A-2. The basic equations relating the variables were shown to be:

$$v^2 + k^2 = x \quad 3.33$$

and

$$\omega_{\max}^2 + 2k^2 = x. \quad 3.34$$

There is a line on the graph of Figure 3-3 labeled $\omega_{\max} = 10k$ and for any point which is on or below this line, one may use the approximation

$$\omega_{\max} \approx v \quad 3.35$$

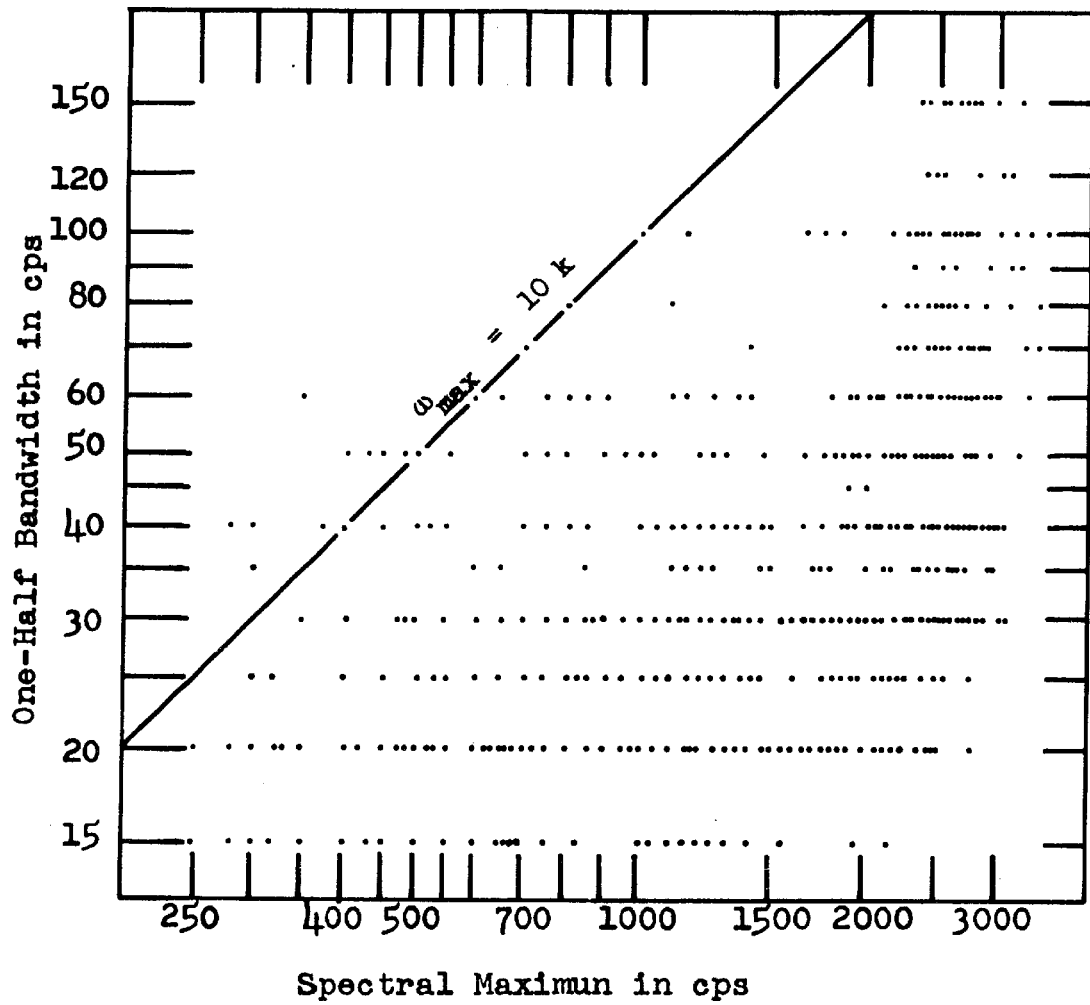
which results in a maximum error of less than one percent. The approximation

$$v \approx x^{1/2} \quad 3.36$$

holds with the same error limitation for those points which are below the line. If one were to make the approximation

$$\omega_{\max} \approx x^{1/2} \quad 3.37$$

for those points which lie below the line, the maximum error is only two percent. The above remarks, although obvious in a mathematical sense, are of considerable significance in work on speech. This is especially true since most of the points in Figure 3-3 lie below the line $\omega_{\max} = 10k$. It should be noted, however, that the remarks are conditional; namely, if



Vowel Formant
Frequency-Bandwidth Data
(After Dunn)

Figure 3-3

$$\omega_{\max} \geq 10k$$

3.38

then an error of less than the magnitude stated above would be made by the approximations. In Dunn's data it was the case that for all second and third formants the relation of Inequality 3.38 was satisfied.

In the case of the first formant the situation is a little more complex. Out of 400 first formant points which were measured only 11 were not on or under the line $\omega_{\max} = 10k$; that is, over 97 percent of the first formant points also satisfied Inequality 3.38. It is interesting to note that of the 11 points which were above the line, 4 were for the same speaker, and 6 for the same vowel. At present there is no way to explain the fact that one speaker frequently exceeded the line and that one vowel frequently exceeded the line; it would be interesting to consider physiologically and psychoacoustically the vowels for these points. In Dunn's data the first formant is the only formant for which the Inequality 3.38 is not satisfied. Even in the case of the first formant, however, the error which would have been involved by assuming either Equation 3.35 or 3.36 would have been less than two percent in 99.5 percent of the cases; and in the worst case, the error would have been only four percent.

Using Dunn's data it has been confirmed that there are two very good approximations which can be made in the case of speech: first, the resonance frequency is approximately the frequency of the spectral maximum, and second, the natural frequency is approximately the frequency of resonance. Since Flanagan [42] has shown that the difference

limen for formant frequencies is about three percent, it would seem that there would be no loss of perceptual information in making the above approximations. Lindblom [73] found in his experiments that the mean error in the standard spectrographic determination of the spectral peak is about 40 cps. Therefore anyone using standard spectrographic techniques could use the spectral amplitude density maximum as an estimator of the formant frequency. This does not mean that the spectral amplitude density maximum is the formant frequency, but only that it is a reliable estimator of the formant frequency when accuracies of the above-mentioned magnitudes are satisfactory. The only meaningful definition of formant frequency is the frequency of a pole in the transfer function of the vocal tract. This point is adequately discussed by Fant [30], Lindblom [73], Flanagan [38], and Stevens and House [108] to name but a few. Lindblom's reply [73] to a statement that some consider the formant frequency to be the spectral maximum is especially worthy of serious consideration.

In Appendix A it is shown that the maximum point on the amplitude density spectrum for the transfer function of a simple resonator is given by

$$|H(j\omega_{\max})| = \frac{x}{2kv} . \quad 3.39$$

Since it has been shown that the approximation $v^2 \approx x$ is a good approximation for the natural vowel resonances, the following approximation to the amplitude of the spectral maximum can be used for the case of natural vowel resonances:

$$|H(j\omega_{\max})| \approx \frac{v}{2k}, \quad 3.40$$

which shows that the maximum formant amplitude is approximately proportional to the formant frequency and inversely proportional to the formant bandwidth.

It is believed by many [45], [69] that intensity is one physical correlate of linguistic stress. Generally, one thinks of the vowel spectrum as fixed and of variations in intensity as achieved by modification of the vocal pulse. When the glottal input spectrum is held constant, there is a relation between the various formant frequencies and the speech output spectrum which has already been described by Fant. The above relations, however, show that another possible means of obtaining an increase in output level would be to decrease the bandwidth. The equations show that if the bandwidth is halved, the spectral maximum shifts upward less than 1/2 percent in frequency, but the amplitude of the formant peak is doubled. This would seem to open a possible area of further investigation; namely, the interaction between the vocal tract and the glottal pulse used to vary the output intensity.

Another interesting observation which can be made from Equation 3.40 is that if the bandwidth is fixed, then the maximum amplitude of the spectral envelope increases with frequency at a rate of 6 db/octave.

Critically damped pulses are often used in synthesis to obtain appropriate spectral characteristics for the glottal source in speech. From time domain considerations, however, they do not really form a

natural type of pulse. One type of phonation is that in which the vocal folds remain closed during an appreciable interval of each period. This type of pulse has frequently been shown in the literature. Holmes [54] has observed this type of pulse as has David [15], Fant [30], Flanagan [35], Mathews, et al. [75], Timke, von Leden and Moore [111], and Miller [78]. It has been reported by Miller and later by Holmes [54] that the first formant is mainly excited at the closing of the glottis. Holmes observed that the second formant seems to be excited by both the opening and the closing of the glottis; that is, the second formant is excited twice per pitch period. He noted that the third formant is excited at both the opening and closing of the glottis, but much less at the opening than at the closing. Therefore, the following equation can be used as a time domain model of a first formant excited at a constant fundamental frequency:

$$S_{\varphi}(t) = \sum_{n=0}^{\infty} U(t-nT) e^{-k(t-nT)} \sin[v(t-nT)+\varphi] \quad 3.41$$

where φ is a parameter. In our applications it will be φ_e , φ_c , φ_r or 0. It is known that actual vocal pulses do not just begin and remain the same. None the less, the above is a reasonable time domain model for consideration. The above equation can be exactly represented on any interval of the form

$$[m-1]T \leq t < mT \quad 3.42$$

and with a change of variable

$$\tau + [m-1]T = t \quad 3.43$$

the equation becomes

$$S_{\phi}(\tau + [m-1]T) = \alpha_m e^{-k\tau} \sin(v\tau + \phi + \beta_1 - \beta_m) \quad 3.44$$

where α_m and β_m have been defined previously. It is easily verified that:

$$\alpha(1 - e^{-kmT}) \leq \alpha_m = \frac{[1 - 2e^{-kmT} \cos v mT + e^{-2kmT}]^{1/2}}{[1 - 2e^{-kT} \cos vT + e^{-2kT}]^{1/2}} \leq \alpha[1 + e^{-kmT}] \quad 3.45$$

where

$$\alpha = \frac{1}{[1 - 2e^{-kT} \cos vT + e^{-2kT}]^{1/2}} \quad 3.46$$

and, as is shown by Equation B.42,

$$\frac{-1}{e^{\frac{kT}{2}} - 1} < \tan \beta_m < \frac{1}{e^{\frac{kT}{2}} - 1}. \quad 3.47$$

Since, for small z , $\tan z \approx z$, as is evident from

$$\tan z = z + \frac{z^3}{3} + \frac{z^5}{15} + \dots, \quad 3.48$$

it follows that the rate of convergence of the amplitude and the rate of convergence of the phase angle are bounded by convergent series which are independent of the resonant frequency v . This would also follow heuristically from the fact that the sum can be represented as a transient term and a periodic term. The envelope of the transient term is of the form Ae^{-kt} which shows that there is a bound on the rate of convergence which is not a function of frequency (only the value of A is dependent upon frequency). Thus, the two parameters which are needed

for determining the rates of convergence are the damping coefficient, k , and the period, T . In general, the convergence to steady-state is within one percent in the first 3 to 5 periods; of course, the wider the bandwidth, the faster the convergence.

Table 3-1 gives some values for the product $2\pi kT$. One can refer to Table 3-1, and the abscissa of Figure 3-4 to determine the percentage error in the convergence of $\alpha_m \rightarrow \alpha$. The table and figure together show the speed of convergence for particular values of k , T , and m .

3. Zero Crossing Frequency as a Measure of Formant Frequency

In 1951, E. Peterson [86] of the Bell Telephone Laboratories experimentally evaluated a zero crossing technique as a means of obtaining spectral information about signals. In the following year an automatic digit recognizer was built at the Bell Telephone Laboratories [19] which used zero crossing information to obtain spectral information. The inputs to this device consisted of the ten digits; the device had to be retrained for new speakers. One modern speech recognizer, that of Sakai [97], uses the zero crossing technique as an estimate of the spectral information. It must be noted that in this recognizer the formant frequency is quantized into intervals of 100 cycles. Thus, the fact that the zero crossing technique is used for recognition is not to be construed as proof that it measures the formant frequency.

For several years the main problem with the zero crossing technique as an estimator of formant frequency was considered to be the disturbances introduced into the measure by other formants and by the vocal pulse.

Bandwidth cps $k/2\pi$	Fundamental Frequency (cps)			
	50	100	200	400
15	1.88	.94	.47	.24
20	2.51	1.25	.63	.31
25	3.14	1.57	.79	.39
30	3.77	1.88	.94	.47
35	4.40	2.20	1.10	.55
40	5.02	2.51	1.26	.63
45	5.65	2.83	1.41	.71
50	6.28	3.14	1.57	.79
60	6.54	3.77	1.88	.94
70	8.79	4.40	2.20	1.10
80	10.05	5.02	2.51	1.26
90	11.30	5.65	2.83	1.41
100	12.56	6.28	3.14	1.57
120	13.07	6.54	3.77	1.88
150	18.84	9.42	4.71	2.35

Value of $2\pi kT$ as a Function of f_0 and $k/2\pi$.

Table 3-1

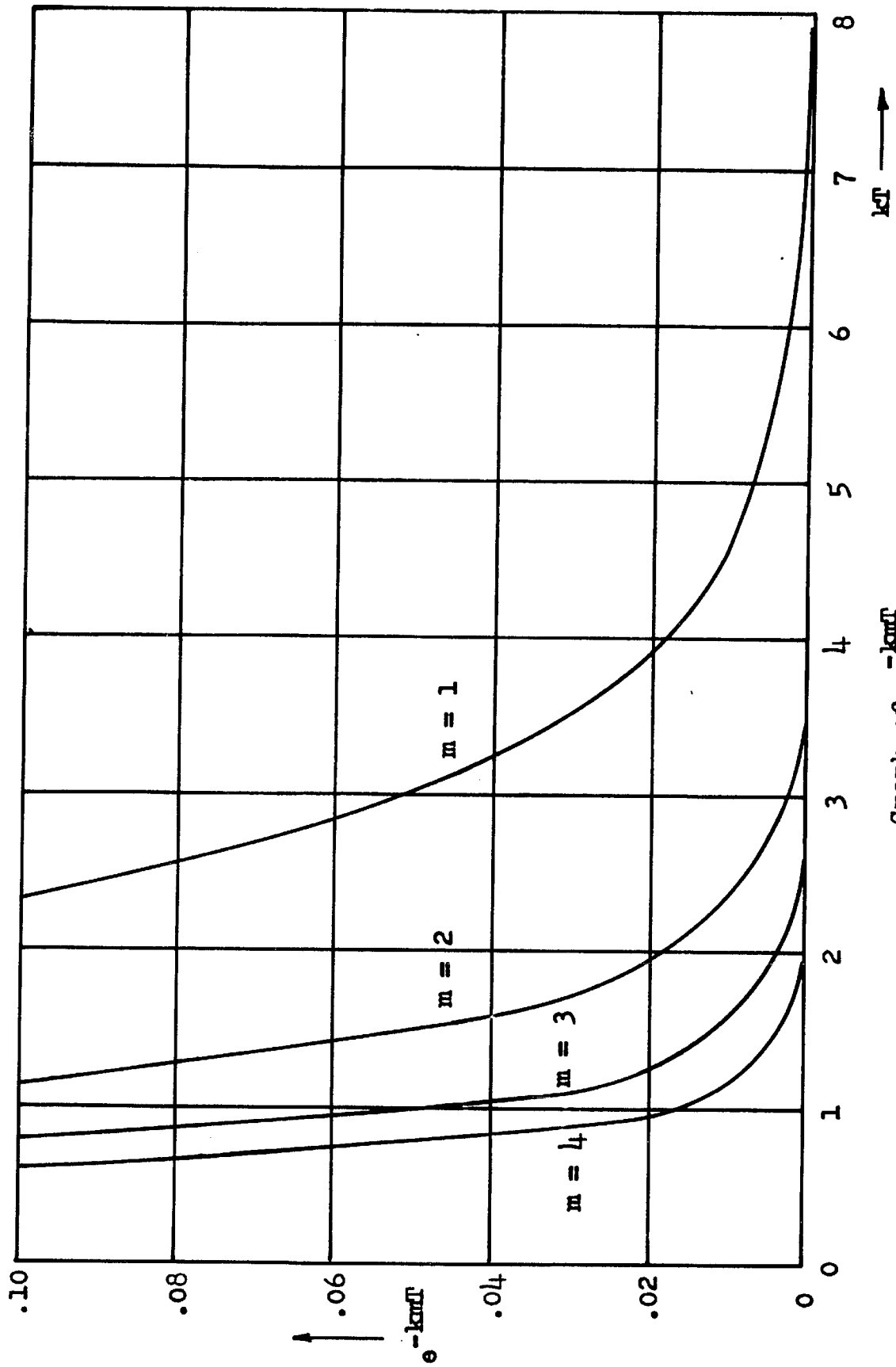
Graph of e^{-kmT}

Figure 3-4

Some of the zero crossing meters which have been built have had dynamically tuned filters in order better to isolate each formant frequency from the other formants and to filter out as much of the vocal pulse as possible, c.f. [9], [56].

Recently, [29], Fant wrote that if zero frequency counting is to be used as an estimate of formant frequency, then formant oscillations must be isolated from each other by some kind of adjustable filter. Fant states that even then the estimated value of the formant frequency will follow the frequency of the strongest harmonic.

It is a rather simple matter to show that a harmonic frequency is obtained when a zero crossing frequency is derived from the output of a single periodically excited resonator if the output has reached the steady-state condition. Let $N(t)$ denote the number of zero crossings for a formant and let $A(Z)$ be the time average of the number of crossings. Then

$$A(Z) = \lim_{T \rightarrow \infty} \frac{N(T)}{T} \quad 3.49$$

and if the formant is in the steady-state condition (periodic with period T_0), it follows that

$$A(Z) = \frac{N(T_0)}{T_0} . \quad 3.50$$

Since $N(T_0)$ is the number of zero crossings in any period, it is an integer. Thus

$$A(Z) = N(T_0)f_0 \quad 3.51$$

is a harmonic. This demonstrates Fant's statement that a zero crossing frequency is a harmonic.

In order to investigate the zero crossings of a single formant in greater detail one may consider the zero crossings of

$$S(t) = \sum_{n=0}^{\infty} U(t-nT) e^{-k(t-nT)} \sin[v(t-nT)]. \quad 3.52$$

A consideration of the zero crossings of the above expression implies that all of the other formants have been filtered out as well as the effects of the vocal pulse. All that is left then, is the time domain representation of a single formant. It is necessary to consider only the steady-state part of the equation; accordingly, the zero crossing frequency for the periodic portion of $S(t)$, will be determined. Thus, the average number of zero crossings per second will be determined for periodic functions of the form

$$0 \leq t < T \quad P_S(t) = \alpha e^{-kt} \sin[vt + \beta_1] \quad 3.53$$

where

$$\alpha = [1 - 2e^{-kT} \cos vT + e^{-2kT}]^{-1/2} \quad 3.54$$

and

$$\tan \beta_1 = \frac{\sin vT}{e^{kT} - \cos vT}. \quad 3.55$$

In order both to simplify the analysis and to maintain a continuous output, the additional phase angle ϕ has been omitted from Equation 3.53.

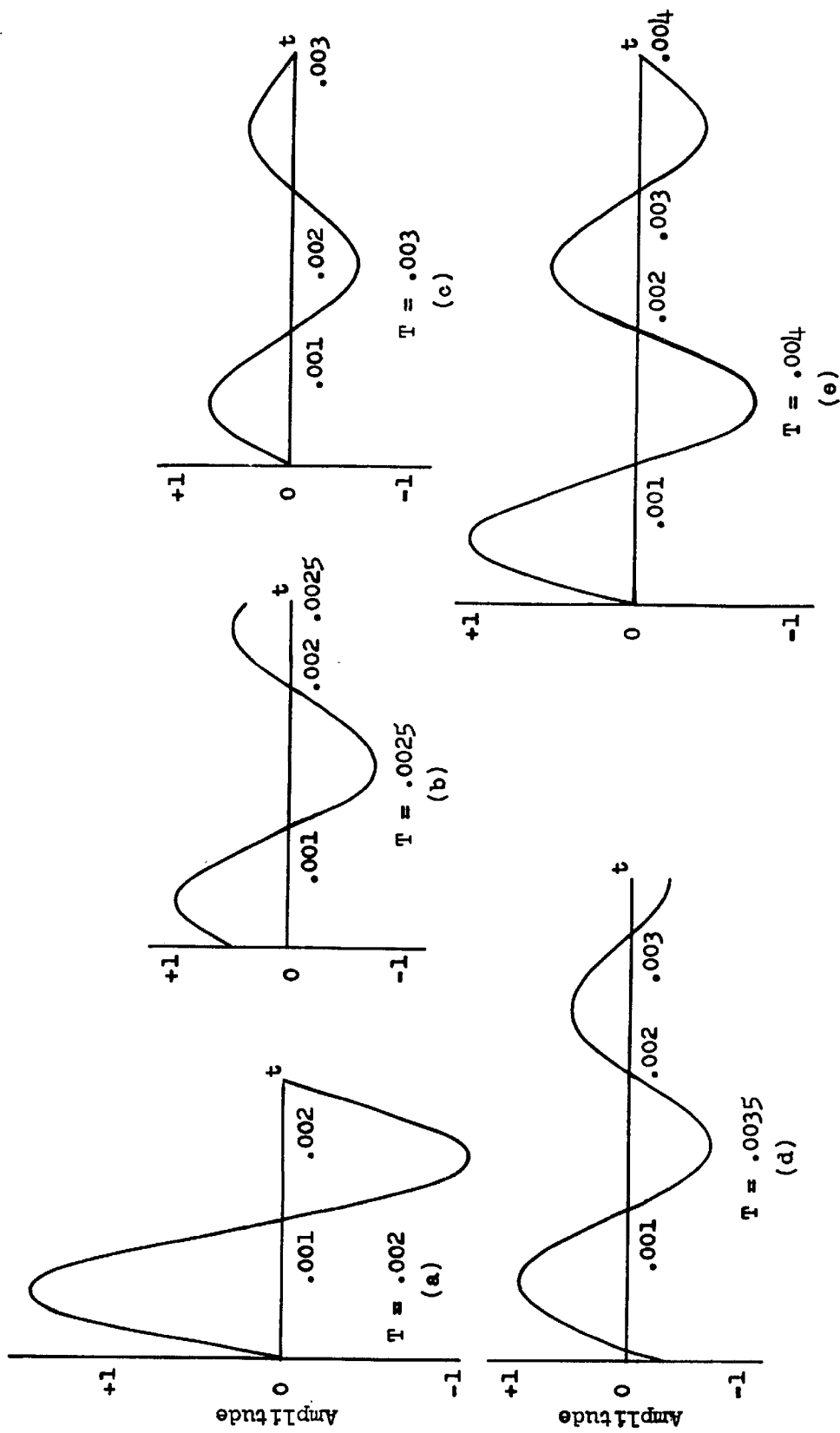
Since all of the terms in the sum are continuous, we know that $S(t)$ is continuous for all finite values of t . The periodic solution is also continuous: that is

$$P_S(0) = P_S(T) \quad 3.56$$

which can be shown by evaluation.

Figure 3-5 shows $P_S(t)$, where v and k are held constant, but where T takes on the values .002, .0025, .003, .0035, .004. In the curves the resonant frequency is 500 cps and the damping factor is 50 cps. In figures (a) and (b) there are two zero crossings per period; in figure (c) there is a possibility of various interpretations (2, 3, or 4 crossings), and in (d) and (e) there are four zero crossings per period. One major observation is clear: the number of zero crossings is constant for limited sets of values of T . $P_S(t)$ is a continuous function. Therefore, except for those cases where the initial value is zero, there will be an even number of crossings. This follows from the fact that at the start and the end of every period the value will be the same (Equation 3.56).

The frequency which is estimated by the number of zero crossings per second is obviously twice the resonant frequency. If one prefers to count only the positive going crossings or to divide by two, it is possible to have a zero crossing counter estimate the formant frequency rather than twice the formant frequency. For the present analysis it is just as easy to work with all of the zero crossings and estimate twice the formant frequency.



Steady State Waveform as a Function of T

for $v = 500$ cps and $k = 50$ cps

Figure 3-5

Clearly, for $0 \leq t < T$, the function

$$P_S(t) = \alpha e^{-kt} \sin[vt + \beta_1] \quad 3.57$$

will have the same number of zero crossings as the function

$$f(t) = \sin[vt + \beta_1]. \quad 3.58$$

The zeros of $f(t)$ are spaced π/v seconds apart. Since there are an even number of zero crossings, it should be clear that for

$$0 \leq T < \pi/v \quad 3.59$$

there will be no zero crossings (since there could not be two); for

$$\pi/v \leq T < 3\pi/v \quad 3.60$$

there will be two zero crossings, and for

$$[2n-1]\pi/v \leq T < [2n+1]\pi/v \quad 3.61$$

there will be $2n$ zero crossings. The number of zero crossings/second will be denoted by $N(t)$, thus for $n \geq 1$ and for

$$[2n-1]\pi/v \leq T < [2n+1]\pi/v$$

it follows that

$$N(t) = 2n/T. \quad 3.62$$

An easy way to visualize $N(t)$ is to consider it as a series of sections on the family of rectangular hyperbolas which are defined by

$$y = 2n/z. \quad 3.63$$

It has been previously noted that there is an element of ambiguity

about the number of zero crossings for the special cases where

$$T_n = \pi[2n+1]/v \quad 3.64$$

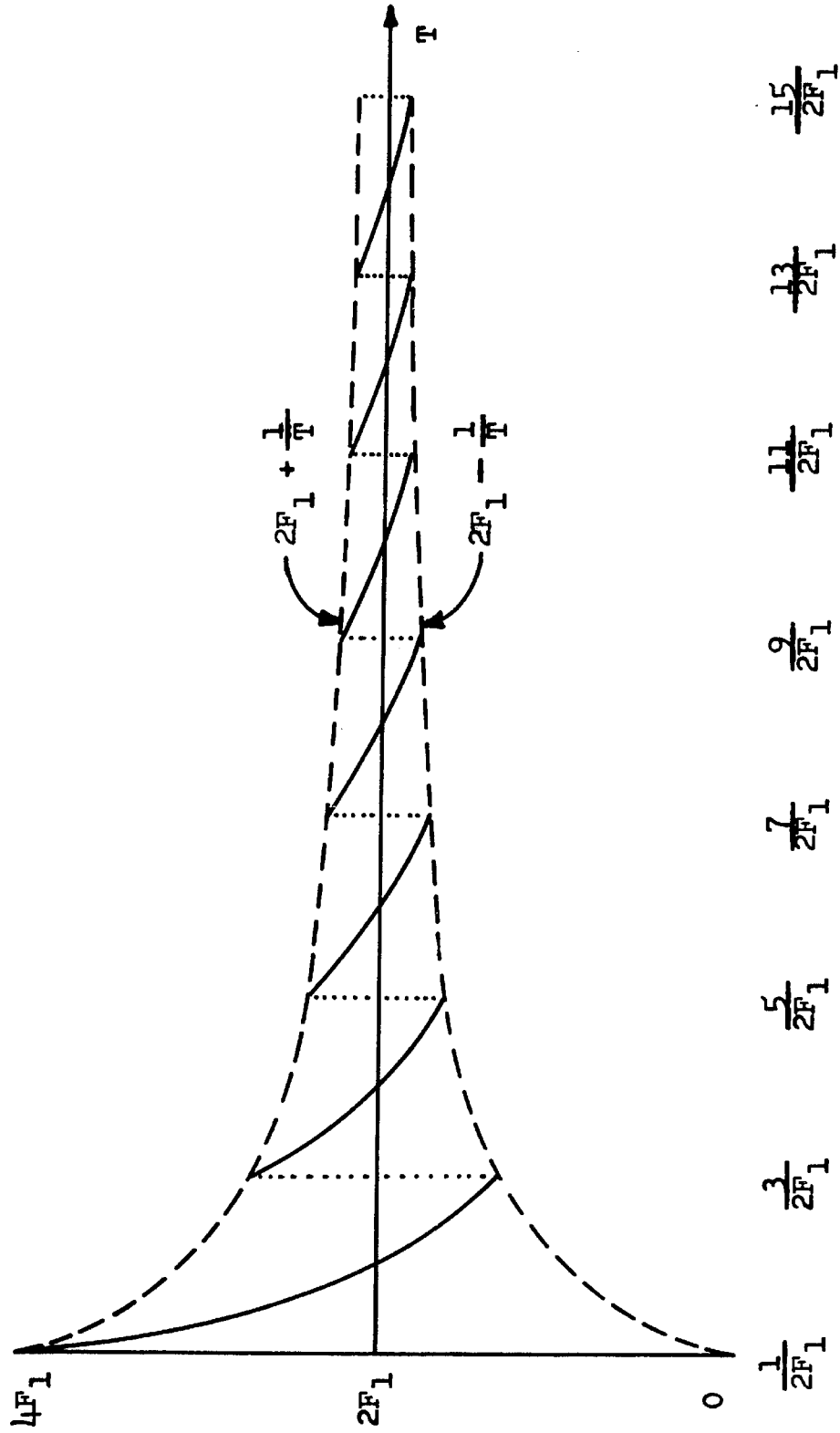
At these points there is a discontinuity in $N(T_n)$ of magnitude $\frac{1}{2T_n}$.

The values of $N(t)$ have been plotted as a function of t in Figure 3-6. The discontinuities at the points $[2n+1]\pi/v$ are shown by the dotted lines. The envelopes passing through the points of maximum error are shown on the figure by the dashed lines. The envelopes are given by

$$E(T) = \frac{v}{\pi} \pm \frac{1}{T} . \quad 3.65$$

These envelopes show that there is always an uncertainty in this technique which is equal to the fundamental voicing frequency, or in the case where the formant frequency is estimated, one half the fundamental frequency. The magnitude of this error shows why this technique has been sufficient for speech recognition schemes based upon either limited vocabularies or crudely quantized spectral information. It also shows why it has been difficult to measure formant frequencies exactly with the technique, even when moveable filters are employed.

In [29] it was stated that zero crossing equipment would follow the strongest harmonic. Since the amplitude density spectrum of a simple filter is not symmetric about the resonant frequency, there is a distinction between the highest energy harmonic and the harmonic nearest the formant frequency, although they are frequently the same. Consider the points nf_0 and $[n+1]f_0$ where the n^{th} harmonic and the $(n+1)^{\text{st}}$ harmonic of a fundamental frequency f_0 are equally distant from



Steady State Zero Crossing Frequency of a
Single Formant as a Function of Period

Figure 3-6

the formant F_1 . These points are related by

$$F_1 - nf_0 = [n+1]f_0 - F_1 \quad 3.66$$

which means that

$$f_0 = \frac{2F_1}{[2n+1]} \quad 3.67$$

Accordingly, nf_0 and $[n-1]f_0$ are points of discontinuity as illustrated in the graph drawn in Figure 3-6. Thus, it has been shown that for the simple model which was investigated, the zero crossing frequency follows the nearest, not necessarily the strongest, harmonic. This fact does appear to be known by some [5].

It has been shown here that for the common model of a formant the zero crossing frequency will be the same as the zero crossing frequency of the nearest, not the strongest, harmonic. Since the zero crossing measure has the property that it is always a harmonic, it has an inherent uncertainty with respect to the formant frequency of one-half of the fundamental frequency. According to Flanagan's [42] three percent difference limen, this measure is generally unacceptable for a first formant. It is marginal for a second formant of a male speaker, and probably sufficient for many third formants in vowels produced by male speakers.

4. Highest Energy Harmonic as a Formant Frequency Estimator

It is possible to determine the formant frequency that would be estimated by a highest energy harmonic tracker as a function of the ratio between the input formant frequency and the fundamental frequency.

For the case of the highest energy harmonic, the result is a function of the formant bandwidth. In some cases this result is different from that for a zero crossing counter which has been shown to estimate the nearest harmonic. We shall use the assumption that

$$v > 10 k \quad 3.68$$

as a justification for the approximation

$$v \approx [v^2 - k^2]^{1/2}. \quad 3.69$$

In this way the results are independent of bandwidth, and yet, still applicable to speech.

The spectrum of a single steady-state formant is nonsymmetric: it falls off faster above the spectral maximum than it does below it. In order to determine the point at which the highest energy harmonic switches from the n^{th} to the $(n+1)^{\text{th}}$ harmonic, we need only equate the amplitude density spectrum for the harmonics: namely,

$$|H(jn\omega_0)| = |H(j(n+1)\omega_0)|. \quad 3.70$$

Substituting into Equation A.9 we find:

$$\frac{x}{[(x - [n\omega_0])^2]^2 + 4k^2[n\omega_0]^2]^{1/2}} = \frac{x}{[(x - [(n+1)\omega_0])^2]^2 + 4k^2[(n+1)\omega_0]^2]^{1/2}} \quad 3.71$$

which is equivalent to:

$$[(n+1)^2 - n^2]\omega_0 \{ [(n+1)^2 + n^2]\omega_0 - 2[x - 2k^2] \} = 0. \quad 3.72$$

The root of interest is:

$$\begin{aligned}
 \omega_o &= \sqrt{\frac{2[x-2k^2]}{(n+1)^2 + n^2}} \\
 &= \sqrt{\frac{2[v^2 - k^2]}{(n+1)^2 + n^2}}.
 \end{aligned}
 \tag{3.73}$$

Equation 3.73 can be simplified by using the approximation

$$\begin{aligned}
 v &\approx \omega_{\max} \\
 &\approx [v^2 - k^2]^{1/2}.
 \end{aligned}
 \tag{3.74}$$

Thus, we have

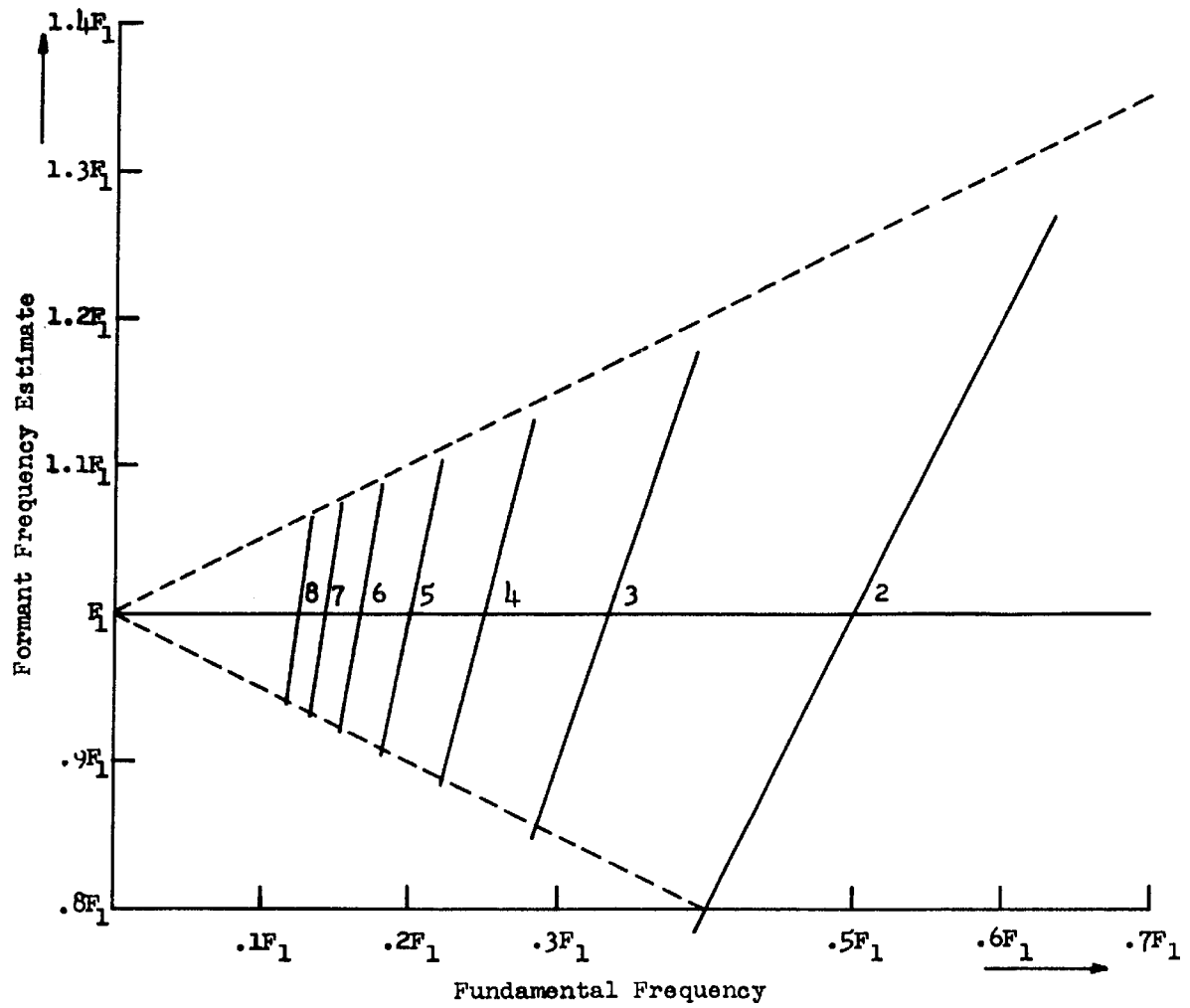
$$\omega_o \approx v \sqrt{\frac{2}{(n+1)^2 + n^2}}
 \tag{3.75}$$

as the relation between the formant frequency and the fundamental frequency when the n^{th} and $(n+1)^{\text{th}}$ harmonics are equal in intensity. The n^{th} harmonic is therefore

$$n\omega_o \approx v \sqrt{\frac{2n^2}{(n+1)^2 + n^2}}
 \tag{3.76}$$

at the point where the highest energy harmonic shifts from the n^{th} to the $(n+1)^{\text{th}}$ harmonic.

The formant frequency estimated by the highest energy harmonic has been plotted as a relative function of fundamental frequency in Figure 3-7. The estimates are always harmonics, these appear as the solid



Formant Frequency Estimated by a Highest Energy Harmonic Tracker as a
Function of Fundamental Frequency

Figure 3-7

straight lines. The dashed lines correspond to where the straight lines would end if we had made the plot for the nearest harmonic rather than the highest energy harmonic. It can be seen from the graph that most of the time a zero crossing counter and a highest energy harmonic tracker would give the same result. The numbers by the harmonics show which harmonics have been represented. The first harmonic was omitted since there is no cut-off when the fundamental is higher than the input formant.

5. Comparison of Errors in the Two Harmonic Trackers

We shall determine two measures of the errors inherent in the two harmonic trackers previously discussed as a function of the harmonic number: the maximum magnitude of the error and the maximum percentage error in estimating the input formant frequency. The interesting result is that the highest energy harmonic tracker makes the maximum magnitude error and the zero crossing counter (nearest harmonic tracker) makes the largest percent error. The difference is due to the manner in which the errors are made.

a. Magnitude errors

The maximum magnitude of error associated with any harmonic in a zero crossing (nearest harmonic tracker) is $f_0/2$ for all harmonics.

The maximum magnitude error in a highest energy harmonic tracker occurs when the formant changes from the n^{th} to the $n+1^{\text{th}}$ harmonic. At this point the n^{th} harmonic is less than the formant frequency and the maximum magnitude of the error for the n^{th} harmonic, $M_H(n)$, is

$$\begin{aligned}
M_H(n) &= F_1 - nf_o \\
&= f_o \sqrt{\frac{(n+1)^2 + n^2}{2}} - nf_o \\
&= f_o \left[\sqrt{\frac{(n+1)^2 + n^2}{2}} - n \right].
\end{aligned} \tag{3.77}$$

It can be shown that

$$\frac{1}{2} = \lim_{n \rightarrow \infty} \left[\sqrt{\frac{(n+1)^2 + n^2}{2}} - n \right] \tag{3.78}$$

and thus, for large n , the magnitude of the maximum error for a highest energy harmonic tracker approaches the maximum magnitude of error for a zero crossing (closest harmonic) tracker.

The two methods of formant frequency estimation are compared in Table 3-2. The magnitude of the maximum error as a function of the fundamental frequency is given for several of the harmonics. No entry is made for the first harmonic, since the tracker can follow it until an error of f_o is made.

b. Percent error of the input formant frequency

One would intuitively expect that the system which made the smallest maximum magnitude error would also make the smallest maximum percent of input frequency error. In this case, however, one's intuition is wrong.

Harmonic Number	Maximum Error Zero Crossing	Maximum Error Highest Energy Harmonic
2	$0.5 f_o$	$0.550 f_o$
3	$0.5 f_o$	$0.536 f_o$
4	$0.5 f_o$	$0.528 f_o$
5	$0.5 f_o$	$0.523 f_o$
6	$0.5 f_o$	$0.519 f_o$
7	$0.5 f_o$	$0.517 f_o$
8	$0.5 f_o$	$0.515 f_o$

Maximum error in cps for each particular harmonic for the zero crossing and the highest energy harmonic estimates of the input formant frequency.

Table 3-2

Before proceeding let us consider a simple example. If we assume a fundamental frequency of 100 cps, then the highest energy harmonic tracker switches from the first to the second harmonic for a formant frequency of 157 cps, and it switches from the second to the third harmonic at a formant frequency of 255 cps. The nearest harmonic tracker would switch from the first to the second harmonic at a formant frequency of 150 cps and would switch from the second to the third harmonic at a formant frequency of 250 cps. Thus, for the second harmonic, the maximum error for the nearest harmonic tracker is 50 cps whereas the maximum error for the highest energy harmonic is 55 cps. At this point our intuition is still content since the nearest harmonic tracker turned out to be nearest the formant. The percent errors are different however. The maximum percent error for each of the systems occurs when the formant frequency is the lowest estimated by the harmonic. In the case of the nearest harmonic tracker we find a maximum error of $50/150 = 33\%$, whereas for the highest energy harmonic we find a maximum error of $43/157 = 27\%$. This is the crux of the matter; the highest energy harmonic tracker makes a bigger error when the formant is above the harmonic and a smaller error when the formant is below the harmonic: the percent errors are maximum for the case where the formant frequency is below the harmonic, and thus the highest energy harmonic is a better percent estimator than the nearest harmonic.

The percent error is maximum for the n^{th} harmonic in the nearest harmonic tracker when the input formant is the lowest tracked by the

harmonic. In this case the formant is related to the harmonic by the equation

$$F_1 = nf_o - [f_o/2]. \quad 3.81$$

The maximum percent error for the n^{th} harmonic in a zero crossing counter, $\%_Z(n)$, is given by:

$$\begin{aligned} \%_Z(n) &= \frac{\frac{f_o}{2}}{F_1} \\ &= \frac{\frac{f_o}{2}}{nf_o - \frac{f_o}{2}} \\ &= \frac{1}{2n - 1}. \end{aligned} \quad 3.82$$

The maximum percent error for the highest energy harmonic tracker occurs at the point where the input formant frequency is the lowest tracked as the n^{th} harmonic. This is the case where the tracker switches from the $(n-1)^{\text{st}}$ to the n^{th} harmonic and the relationship between the harmonic and the formant is given by:

$$nf_o = F_1 \sqrt{\frac{2n^2}{(n-1)^2 + n^2}}. \quad 3.83$$

The percent error of the input formant frequency for a highest energy harmonic tracker for the n^{th} formant, $\%_H(n)$, is given by:

$$\%_H(n) = \frac{nF_0 - F_1}{F_1}$$

$$= \sqrt{\frac{2n^2}{(n-1)^2 + n^2}} - 1. \quad 3.84$$

The values for these functions have been tabulated in Table 3-3. Again there is no entry for the first harmonic since the error becomes essentially infinite in this case.

Harmonic Number	Maximum % Error Zero Crossing	Maximum % Error Highest Energy Harmonic
2	33.3%	26.5%
3	20.0%	17.7%
4	14.3%	13.1%
5	11.1%	10.4%
6	9.1%	8.6%
7	7.7%	7.4%
8	6.7%	6.4%

Maximum percent error in estimating the formant frequency as a function of harmonic number for the zero crossing counter and the highest energy harmonic tracker.

Table 3-3

Chapter 4

A Formant Estimation Technique

The remainder of the present study is an inquiry into the possibility of the relatively accurate derivation of the formant frequencies at a rate approaching real time. In order to achieve a relatively fast rate, analog equipment was used to obtain the required linear transformations of the input signal. In order easily to achieve the nonlinear portions of the analysis, a digital computer was used to process the individual filter outputs.

The formants appear in the time domain as damped sinusoids. The formant frequency is physically realized as one-half of the reciprocal of the time between zero crossings within a laryngeal period. If the effect of the laryngeal frequency is not eliminated, that is, if the average zero crossing rate is used to determine the formant frequency, then the estimate approaches a harmonic frequency. Thus, it is necessary to perform a selective type of zero crossing measure.

Several formants are characteristically present simultaneously in the quasi-periodic waveforms of vowels and consonants. The measurement of these formants is a problem of considerable magnitude. There are practical as well as theoretical considerations which are relevant to the development of an effective system for formant frequency analysis. The excitation of the vocal cavities is not always simple

and uniform, and the excitation varies in rate over a considerable range. The higher frequencies from the laryngeal source may be low in amplitude and noise may be associated with the excitation as in the case of laryngealized (irregular) or breathy laryngeal action. Further, the frequencies of the formants may be relatively close together, so that the measure of either is disturbed by the presence of the other.

In the system to be investigated the formants are first isolated by frequency domain techniques, then time domain techniques are used to measure the actual formant frequencies. An analog system seems particularly suited to the analysis in the frequency domain. First, it is relatively easy to adjust filter characteristics. Second, a scanning filter or a filter bank can be used to process speech in real time for such analysis. Equivalent processing in currently available digital computers is exceedingly time consuming, and, at least at present, is not applicable to the real time systems of interest in speech analysis. Under the present circumstances it appears that a combination of digital and analog techniques is by far the best approach to the problem of speech formant analysis.

The system investigated first isolates the formant by means of a set of fixed filters. Then relatively crude boundaries for the fundamental laryngeal period are identified. Finally, the formant frequency is determined from the zero crossing rate of the time function within a single period.

1. Isolation Technique

Most systems employ filter banks for the frequency domain analysis of speech. It is also possible to isolate the frequency region within which a formant lies by using a movable filter or possibly a set of movable filters. Equivalently, the speech spectrum may be shifted across a fixed filter or a set of fixed filters to isolate a particular formant. In the sound spectrograph of Koenig, Dunn, and Lacy [63] the speech spectrum is slowly shifted across a single analyzing filter to produce the visual display. The formant tracker developed by Hellwarth [52] shifts the speech spectrum in real time to follow a particular formant. Howard's [8, 56] ρ_0 meter positions a filter of the type designed by Dolansky [22] in the frequency region indicated by the axis crossing rate. For the purpose of laboratory observation it is possible to employ an adjustable filter and an oscilloscope (especially a memory scope) to observe a formant waveform. The formant frequency can then be estimated from the display on the oscilloscope.

It is also possible, as in fact was done in the present study, to employ fixed filters for the analysis and then to determine automatically which of the filters presents the most suitable output for measuring a particular formant. If the actual formant frequency is of interest, then that part of the time domain function which corresponds to a single laryngeal period must first be identified. Thereafter the formant frequency may be estimated from the zero crossing rate within a single period. With this type of isolation of both the formant and the period, one can obtain very good measurements of the formant

frequency.

The above procedures cannot be carried out with what is customarily used as a speech processing filter bank. The bandwidths of the filters in most speech processing filter banks are chosen to be small enough that the filter center frequency provides a sufficiently good estimate of the frequency of the wave at the filter output. For instance, Hughes [59] employed bandwidths of 50 cycles for the low frequencies in his filter bank. Such filter banks normally have relatively small overlap in the frequency domain. Specifically, in such filter banks the upper frequency which is three decibels below the maximum for one filter normally corresponds to the lower frequency which is three decibels below the maximum for the filter with the next highest center frequency.

The sound spectrograph was one of the first instruments to make use of extensive filter overlap in speech analysis. The continuity of the formants is typically well displayed in a broad-band spectrogram. A continuous analyzer with the same properties as the sound spectrograph originally developed at the Bell Telephone Laboratories would require a bank of 200 filters. Each filter would have a bandwidth of 300 cps. If the first filter had a center frequency located at 150 cps; then the second filter would have the center frequency located at 167.5 cps, and the n^{th} filter would have a center frequency located at $150 + 17.5(n-1)$ cps. Any single frequency in the central portion of the filter bank would be within the pass-band of approximately 17 filters simultaneously. The effectiveness of the display achieved with the sound spectrograph strongly suggests that filters which have very considerable

overlap are desirable in the analysis of the speech wave.

The bandwidth of a filter is associated with its time constant. Small bandwidth filters have longer time constants than wide bandwidth filters. The speech input to these filters which is of interest to the present investigation is a quasi-periodic waveform. The normal manner of evaluating the operation of a filter upon a waveform of this type is to consider the product of the spectrum associated with the input and the spectrum associated with the filter. Thus, since the input is a line spectrum the filter output would be considered as a line spectrum. There is another manner for evaluating the filtering operation which allows a better insight into the result for the case in which we are interested. Consider the impulse response reversed in time to be a time window. If the time constant, as shown by the impulse response, is long enough to include several periods of the input, then the resultant spectrum approximates a line spectrum. If the time constant is short with respect to the input period, then the output approximates a continuous spectrum. In both cases, the result is very much dependent upon the relation between the fundamental frequency and the time constant. Therefore, in order to obtain a damped sinusoid as the output of a filter, it is necessary that the filter have a wide bandwidth relative to the harmonic spacing of the input fundamental frequency. Koenig, Dunn, and Lacy [63] have discussed the fact that various types of time-frequency patterns result from different filters. They point out that regardless of which of the two filter widths one employed "the spectrograph does not give a false picture; it is simply a matter of choice which interpretation of

the signal is more convenient." One further observation should be made about the filter used in the sound spectrograph. The filter has approximately a 25 db drop in 265 cycles, which is independent of the speech frequency being analyzed. This filter thus has the property of a very sharp cut-off which, in db/octave, gets sharper as the speech frequencies increase.

In our problem, that of estimating the formant frequency, the convenient signal representation is the damped sinusoid; therefore, we shall use a wide bandwidth. In fact, we shall employ a 300 cps bandwidth just as in the spectrograph. In this way a signal representation is obtained which is approximately as good as that from a spectrograph for determining formant frequencies. The system has the same problems as the spectrograph, however, when the fundamental frequency is increased beyond the low range of female speakers. Overlapping filters were employed in the present system, but the extent of the overlap is much less than in the sound spectrograph. By the use of overlapping filters it is possible to examine the output of adjoining filters to determine which filter best isolates any particular formant. In brief, we shall use a small set of overlapping filters to isolate the first formant. The filters employed provide enough isolation, including most cases where the first two formants appear near each other in back vowels. It appears that it would be difficult, however, to use a system of passive filters to isolate the second and third formants in certain cases due to their relative closeness in octaves. The increasing sharpness of the filter skirts as a function of the frequency being analyzed is the reason this is so well

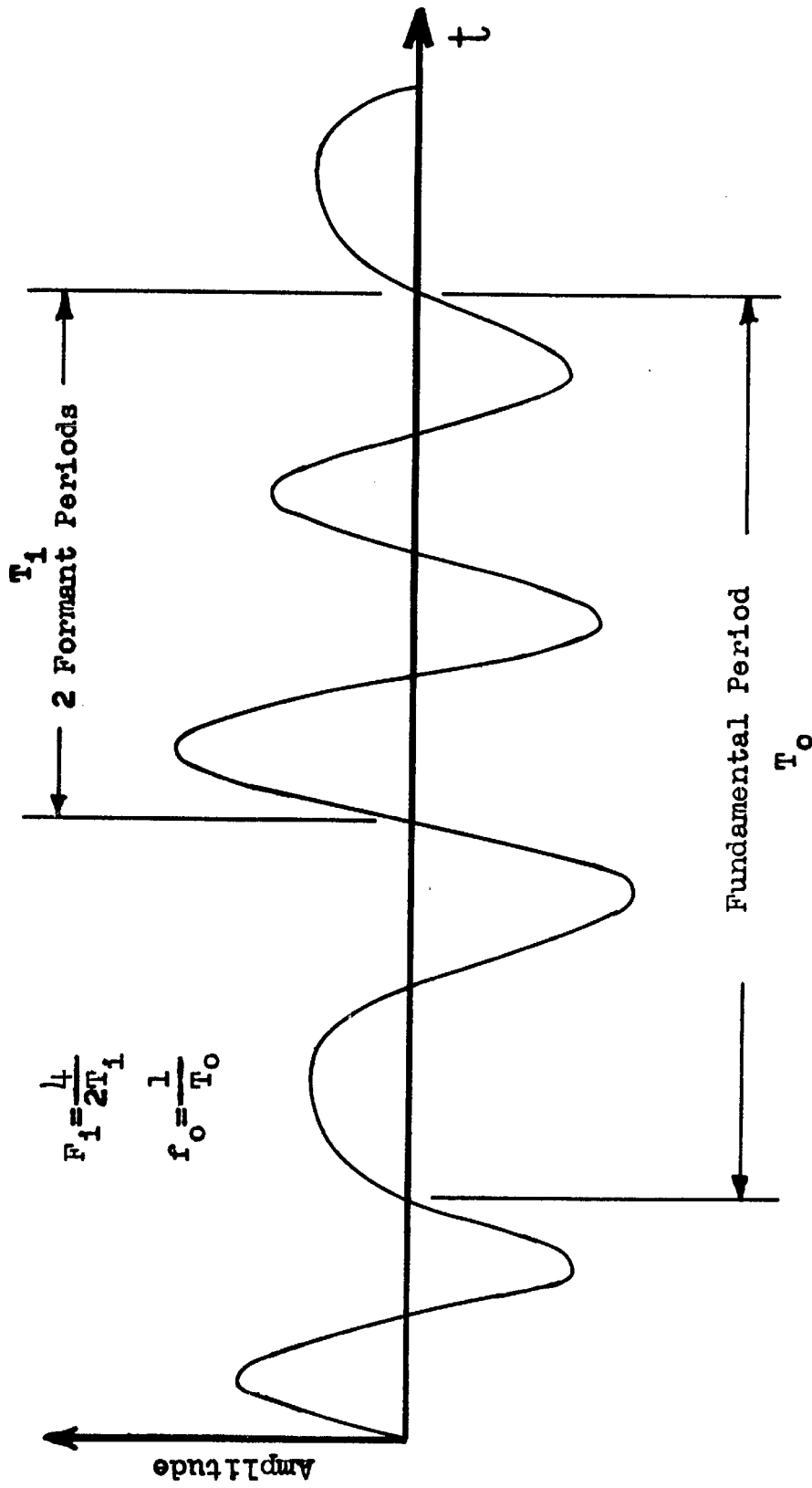
achieved by the sound spectrograph.

2. Processing Technique

The standard technique for processing the output of a filter bank is to rectify and smooth the output of each individual filter. It is quite proper, for example, that a spectrograph should process the output of the filter in this manner to give an amplitude density spectrum. However, we shall not have the equivalent of 200 filters, nor are we interested in spectral maxima. Rather, we shall attempt to determine the actual formant frequencies as a function of the zero crossings of the filter outputs.

Figure 4-1 shows an output of a filter corresponding to one human manifestation of a phone type which would be classified as the phoneme /r/. The output of this filter will be processed in the following manner. First, an automatic "period detector" will be used in order to determine the beginning and ending of each period. The method of measuring the period is somewhat unusual, yet is simple and has no more (or less) error associated with it than most period measuring techniques. The theoretical aspects of the accuracy of this method of measuring the period of the fundamental frequency are discussed later. The main object of this procedure, however, is not to achieve accuracy in estimating fundamental period: rather it is to determine the fundamental period sufficiently well that the formant frequency can be estimated by the zero crossings within the fundamental period.

The processing scheme can be described easily in terms of the graph shown in Figure 4-1. First the period boundaries are determined.



A Single Formant from Actual Speech Isolated by a Band-pass Filter

Figure 4-1

In the figure these are marked by the lines for the "Fundamental Period". The period is always assumed to begin and end at the value zero, and the reciprocal of the length of the interval can be used as an estimator of the fundamental frequency. The formant frequency is obtained by determining the number of half formant periods (not counting the first whole formant period) in the fundamental period, and then by dividing by twice the time interval of the fundamental period minus the first whole formant period. For example, in Figure 4-1 there are four half formant periods which occur in time T_1 ; thus the formant frequency is given by

$$F_1 = \frac{4}{2T_1} = \frac{2}{T_1} . \quad 4.1$$

If the outputs of the other filters can be ignored automatically, then only the filters near the formant frequency need have accurate answers. Therefore, it is only important that the estimation technique for fundamental period and formant frequency be effective in the case of filters near a formant in frequency.

3. Summary

The estimation technique to be used has been described as follows. An initial form of spectral analysis and isolation is performed by the use of a filter bank, consisting of overlapping wide-bandwidth filters. The output of each filter is then processed in a computer to attempt to determine the period and the formant frequency. By methods to be described subsequently, a selection is next made of the filter output

which has the most accurate representation of the formant frequency.

Chapter 5

Equipment Used to Implement the Formant Tracking Method

The purpose of this chapter is to describe the equipment used to study the approach to formant analysis suggested in the preceding chapter. While the individual components of the system are all well-known devices, their combination into an analysis system differs in certain major aspects from previous approaches described in the literature. The details of the instrumentation are described in the present chapter in order to show the nature of the system more specifically.

1. Introduction

The system input consists of two channels from a 4-channel Ampex 300 tape recorder. One of the channels provides the recorded audio speech signal; the other channel provides a pulse source which is used logically by various parts of the system, as will be described later. The audio input is processed by an adjustable filter bank which consists of five Krohn-Hite Variable Band Pass Filters. The output of these filters is fed into a Texas Instrument Multiplexer which converts the parallel filter outputs into a serial form. Generally the total speech wave and a constant voltage are also used as inputs to the multiplexer. The serial output of the multiplexer is used as the input to a Texas

Instrument Analog-to-Digital Converter. The output of the converter is controlled by a tape control unit and is used as the input to an Ampex digital tape recorder. The digital tapes are hand carried to the Computing Center where they are processed. The Computing Center has an IBM 1410 which is used for I/O (input-output) and an IBM 7090 which is used as the main computer. The digital tape of the filter outputs is processed by programs on the IBM 7090. A system diagram of the formant tracker is shown in Figure 5-1.

2. Audio Tape Recorder

The audio tape recorder used as the input to the Formant Tracker is a 4-channel Ampex 300 with associated Ampex electronics. The recorder is operated at 15 i.p.s.; at this speed its rated frequency response is ± 2 db from 50 to 15,000 c.p.s. Only two channels of the tape recorder are currently being used. One is used for the speech signal; the other is used for logical pulses which start and stop the digital tape recorder. The pulses also switch on and off the digital pulse input to the digital tape unit. A third channel would be used if the speech signals came from a microphone instead of another audio tape. In this case, a 7500 c.p.s. sine wave would be recorded on the channel concurrently with the speech. The sine wave would then be multiplied by a factor of 6 and used as the clock for the multiplexer and analog-to-digital converter. The advantage of using the sine wave is two-fold. It eliminates errors due to instability in either the multiplexer clock or the audio tape speed. Although the multiplexer clock is relatively stable, it is limited by the fact that it is capacitor controlled. The

capacitor control makes it possible, of course, easily to change the internal clock rate on the multiplexer.

3. Filter Bank

The filter bank consisted of Krohn-Hite Variable Band-Pass Filters: 3 Model 310AB filters, a model 310ABR filter, and a Model 330M filter. The above filters have a voltage transfer function [66] which is approximately:

$$H(\omega) = \frac{\omega^4 T_1^4}{[(1+2jA\omega T_1 - \omega^2 T_1^2)^2][(1+2jA\omega T_2 - \omega^2 T_2^2)^2]} \quad 5.1$$

where the lower cut-off frequency is

$$f_L = \frac{1}{2\pi T_1}, \quad 5.2$$

the upper cut-off frequency is

$$f_U = \frac{1}{2\pi T_2}, \quad 5.3$$

and the value of the "peaking factor," A , is approximately 0.6. The value reduces the attenuation at the corner frequencies by approximately 9 db from that which would be achieved by a filter composed of independent resistive-capacitive elements (for such a filter $A=1$). The filters have a slope of approximately 12 db/octave at the cut-off frequencies, and by the 12 db down points have essentially reached their asymptotic values of 24 db/octave. The difference between an independent 4-section filter and the Krohn-Hite filters can be seen from the

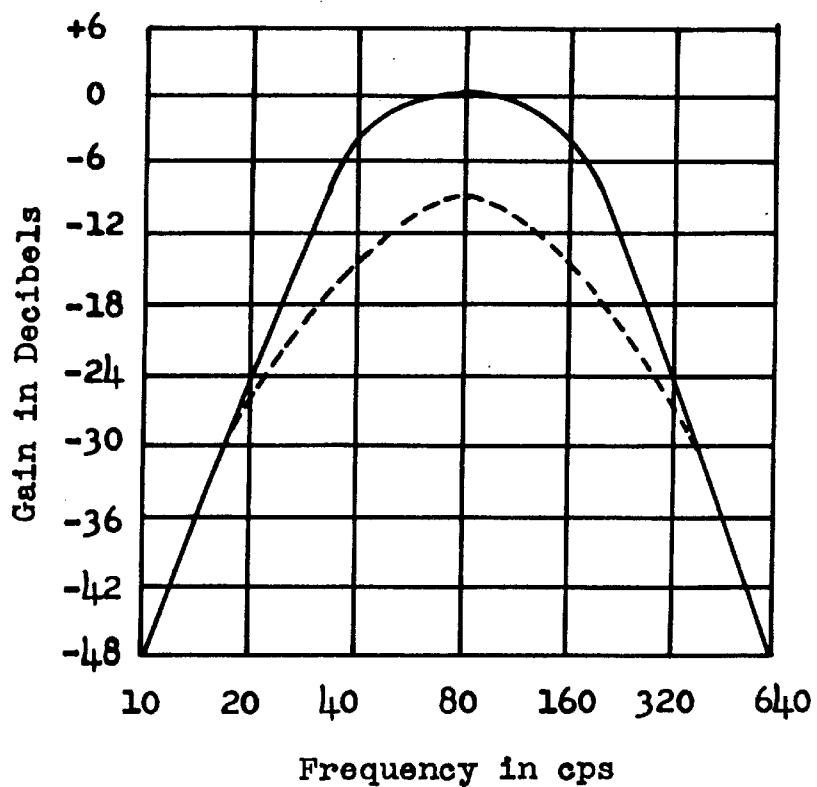
frequency response curve shown in Figure 5-2.

The filters have a pattern like that shown in Figure 5-2 when the bandwidth is greater than one octave. In the formant tracker the filters are generally set to a bandwidth of less than an octave. The frequency response of the filter settings used for the calibration runs are shown in Figures 5-3 to 5-6.

4. Multiplexer

The multiplexer used in the formant tracker is a Texas Instrument Series 845B time-division unit. The multiplexer has a possible sampling rate of 50,000 samples per second; it is being operated at about 44,000 samples per second. The computer program which measures the formants can be run in either of two ways: with a constant sampling rate or with the actual sampling rate at the time the speech record is converted to digital tape. In the second case, the sampling rate is measured by a Hewlett-Packard Model 522B Electronic Counter, recorded by hand near the time the record is taken, manually put on punched cards, and read into the computer as data.

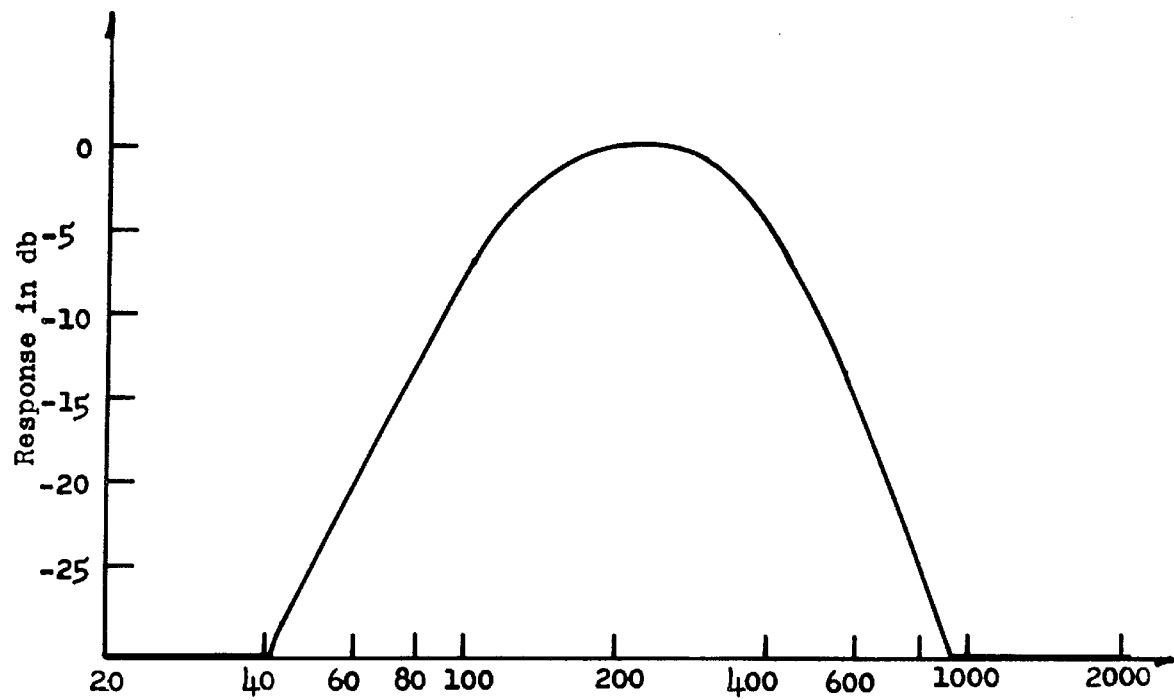
The switch impedance of the multiplexer is 50 ± 20 ohms when on and 2000 Megohms when off. The multiplexer provides both an analog (signal) line and a digital (convert enable) line to the analog-to-digital converter. The multiplexer has a capability of 160 channels, but at present, there are only digital cards available for 32 channels. Only six channels were used since the sampling rate is a limitation in the present work.



Solid curve shows the Band-Pass response of the Krohn-Hite filter. Dashed curve is the response of a similar filter with independent RC sections.

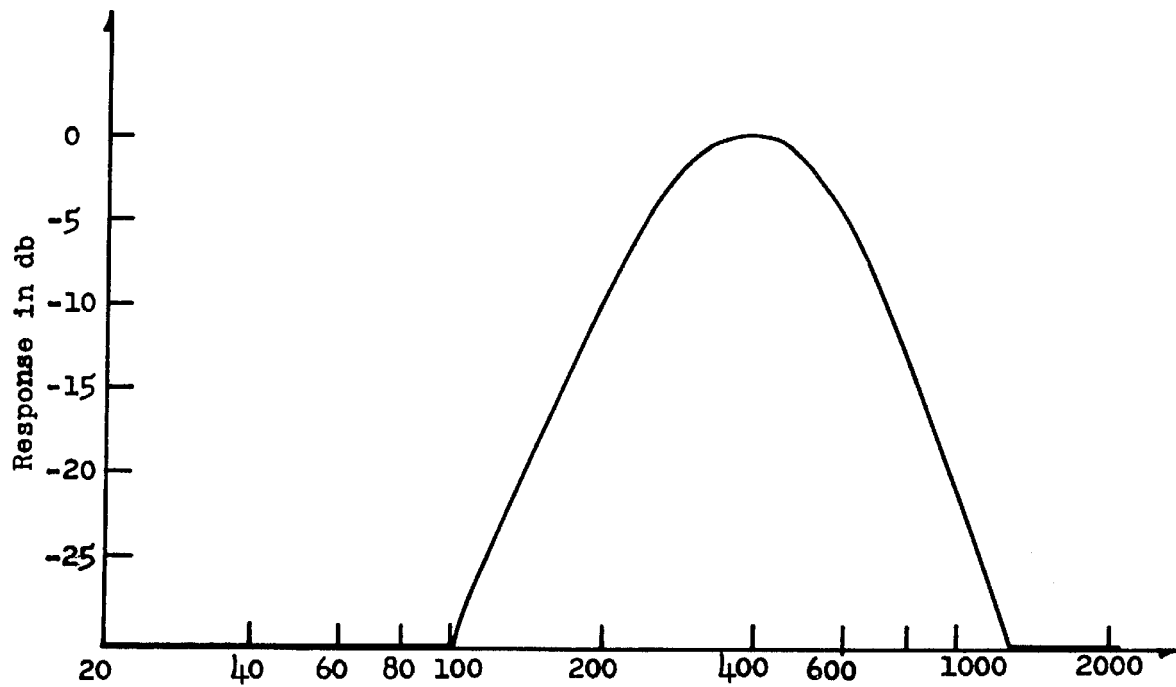
Amplitude Spectrum of a Krohn-Hite Filter

Figure 5-2



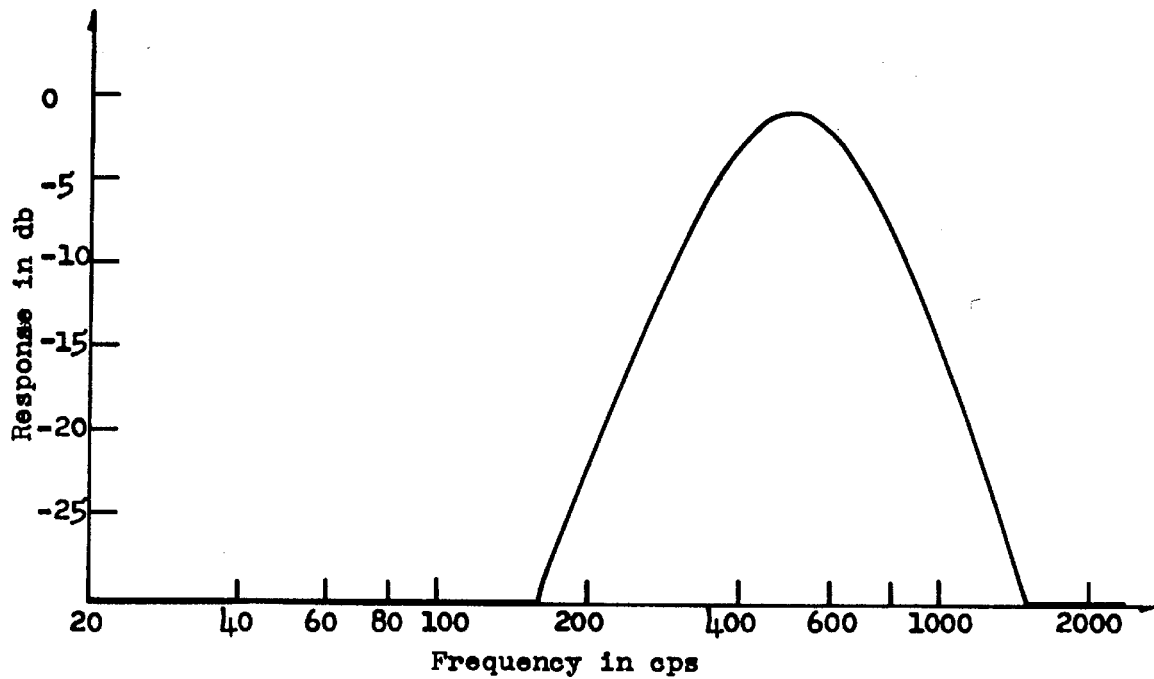
Amplitude Response of Filter 2

Figure 5-3

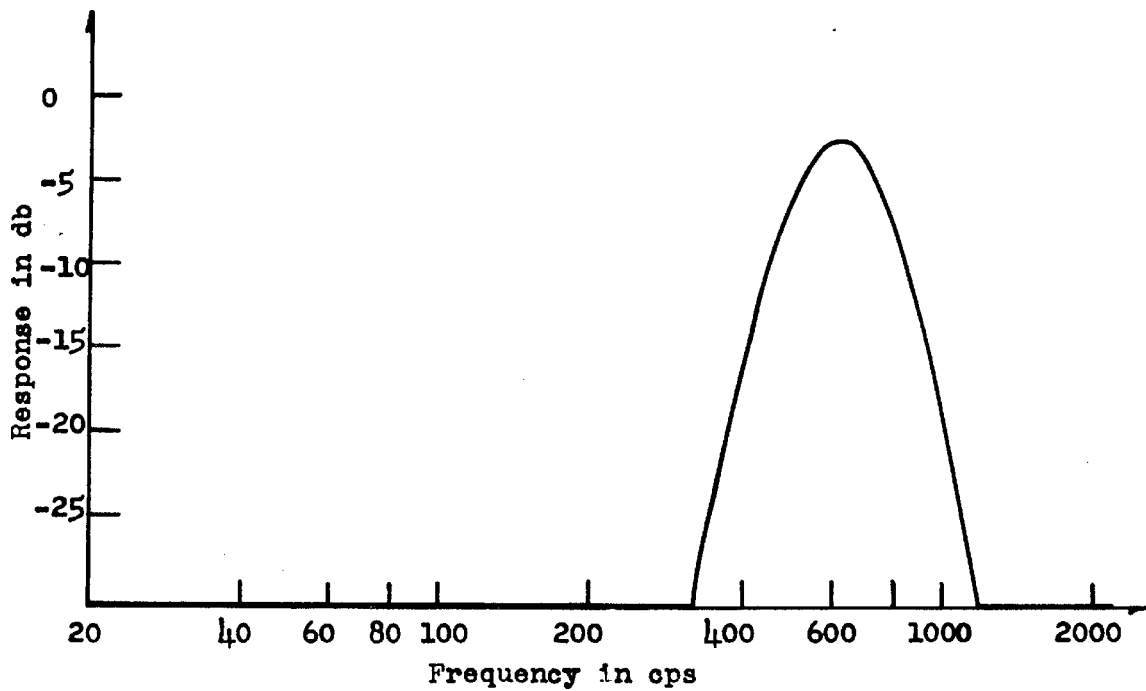


Amplitude Response of Filter 3

Figure 5-4



Amplitude Response of Filter 4
Figure 5-5



Amplitude Response of Filter 5
Figure 5-6

5. Analog-to-Digital Converter

The formant tracker also uses a Texas Instrument Series 834 Analog-to-Digital Converter. The input impedance is 100 Kilohms. There are three possible input voltage scales, 2.5, 5 and 10 volts maximum; the instrument is protected for inputs greater than 12 volts on any scale. We use the 2.5 volt scale to have maximum overload capability, since the peak clipping which occurs for small overloads does not effect the estimation technique. There are, however, levels of clipping which would affect the technique.

The conversion accuracy is $\pm 0.05\%$ of full scale plus and minus one-half the least significant bit. Although the equipment has a 12 bit conversion, the formant tracker only uses the sign and 5 most significant bits. The limitation is caused by tape speed and sampling requirements.

6. Tape Control

The tape control unit was built in the Communication Sciences Laboratory by Norris McKinney. It has two main functions. It alternately starts and stops the digital tape unit on pulse command from the audio tape unit. It also gates the digital input to the tape unit. The digital input to the digital tape unit starts approximately 10 milliseconds after the tape unit starts; the digital tape runs at full speed for 4 milliseconds; the control unit turns off the digital input when the digital tape is turned off. The tape control also has the capability of converting a sine wave at 7500 cps into a series of pulses at 45,000 cps which can be used as an external clock for the multiplexer

and analog-to-digital converter.

7. Digital Tape Unit

The formant tracker is currently using an Ampex FR400 8-track digital tape transport. Only seven channels are being employed, since the tape units of the computer only use seven channels. The Ampex FR400 in the laboratory has speeds of 37-1/2 and 75 i.p.s. It is currently being operated at 75 i.p.s.

8. Digital Computer

The computing facilities at The University of Michigan's Computing Center were used for processing the digital tapes. The Computing Center has an IBM 1410 which is used for general I/O work. Programs, in the form of cards, are converted to tape and then read into an IBM 7090 for processing. The output of the 7090 is put on tape and converted to print or punched cards by the IBM 1410. The tapes at the Computing Center are normally "read" at 125 i.p.s., 556 bits/inch. This tape reading mode is used by the program for the formant tracker.

The 7090 used in the formant tracker has two active I/O channels (potentially there are eight data channels on such a machine, but only two are available on the machine which was used). In theory one could be computing, loading data, and writing output tapes, all in parallel. All output tapes for printing or punching are under system control and do not allow parallel operations. Therefore, the highest level of parallel operation possible was reading an input data tape and either computing or writing one of the system output tapes. This level of parallel

operation was used for all acoustic inputs to the formant tracker.

The digital computer had two main tasks to perform in formant tracking: to process the outputs of the filters individually and to estimate the formant frequency from the processed filter outputs. The details of these two operations are discussed at some length in Chapter 6.

9. Summary

It should be recalled that one purpose of this study was to investigate the possibility of developing a formant tracking system based upon the principles outlined in Chapter 4. However, such a feasibility study must consider both the appropriateness of the mathematical model formulated to solve the problem and the possibilities for implementing that model using currently available hardware. The only way to prove implementation of the model was, so to speak, by an "existence proof." The current chapter has outlined the hardware which was used for implementation of the essential process of a complete system based on the model.

Chapter 6

Computer Program for First Formant Frequency Estimation

The purpose of this chapter is to describe the basic computer program used for estimating the formant frequencies from the digital tape input to the computer. That is, this chapter is a description of how the multiplexed digitized filter outputs are analyzed to determine the first formant frequency of the speech input as a function of time.

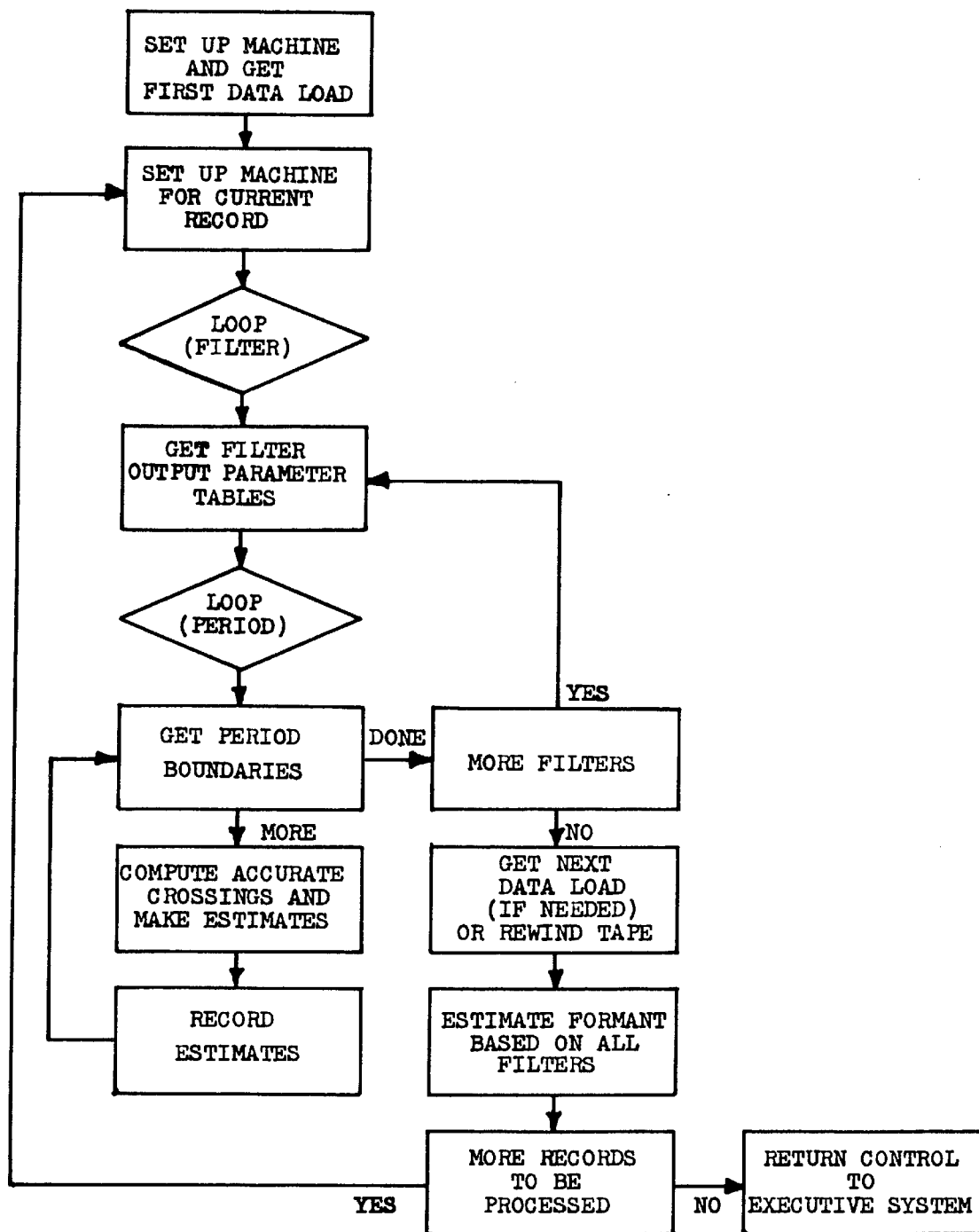
1. Basic Machine Operation

The basic machine flow chart is given in Figure 6-1. The program begins by initializing the machine and bringing in the first record to be processed. The machine is then set to process the first record. Then the loop used to process each filter is started. The main filter output parameters are determined next; these parameters include the following:

- a) time (in sample points) between successive axis crossings.

This time defines the interval referred to below.

- b) time of axis crossing or time immediately preceding the axis crossing if there is not an exact time of crossing (i.e. if the sample point does not coincide with the axis crossing).
- c) the formant properties on the interval (between successive axis crossings); that is, whether the formant was positive or



Basic Program Flow Chart

Figure 6-1

- negative and whether the last value on the interval was zero.
- d) the maximum amplitude for each interval, and
- e) a string over the alphabet {A,B,C,D} which encodes the relations between the amplitude of each set of four consecutive intervals.

The above is only a brief statement of the filter output parameters which are computed. A more detailed description of these parameters, their computation and their use occurs in section 3 of this chapter.

Within the loop indexed by the filters there is a loop which operates upon the previously mentioned parameters as well as on the input data table. This interval loop determines the period boundaries and estimates the formant frequencies associated with the output of each filter. The period boundaries are determined solely from the sign of the filter outputs on the intervals and from the string over the alphabet {A,B,C,D}. As mentioned above, this string encodes a relation between formant amplitudes on consecutive intervals. The exact relation used to define the string will be discussed in detail later, as will the technique used to determine the period boundaries.

Based upon the knowledge of the period boundaries, more exact estimates (using linear interpolation between sample points) are made of the fundamental period and the interval from which the formant frequency is to be estimated. Then the estimates of fundamental and first formant frequency are made and the pertinent information is stored. This process is repeated for each period until the output for the filter has been entirely processed. When the output of the filter has been processed the program repeats the analysis for the next filter, or if

there are no more filters, proceeds to the determination of the frequency of the formant based upon all of the filter outputs.

When all filter outputs have been processed the computer starts a parallel data loading operation as long as another record remains to be analyzed. The tape is rewound after the last record is read. The formant frequency is then estimated making use of the redundant form of the input. The criteria used for the final estimation are:

- a) any estimate must lie within the center $1/2$ of the pass-band of the filter,
- b) if there are several such estimates, then each must agree in time ($\pm \frac{10}{7433}$ sec) and frequency (± 50 cps) with either the filter immediately above or below the estimating filter, and
- c) the filter lowest in frequency must either be chosen or must confirm (as in b) the filter chosen.

The above conditions are described in detail in section 7 of this chapter. After the formants have been determined, then the machine either proceeds to the next record or, in the case of no more records to be processed, returns control to the executive system.

2. Initialization of Machine for Current Record

There are many miscellaneous operations which must be performed with each new input record. Some of these are of major importance, some are not essential but avoid confusion.

A data card is read with each new record. This is an identification card punched in such a way as to describe the data being processed. This card is used as a label on all outputs associated with the particular

record from the digital input tape. The sole purpose of this label is to avoid confusion in the output records.

The program then transfers to subroutine PUNC, which is mnemonic for PUNCTuation. The subroutine is so-named because its main purpose is to find the constant input channel which is used to punctuate (separate) the other five data channels. When this channel is found a check is made to determine whether any characters were dropped from the input tape. There is a real need for this operation since the tape is written without check sums and is read with the redundancy traps off. In this case, the constant input channel is used as a check for synchronization. If synchronization is present for the entire length of the record, then the program proceeds. Otherwise, the length of the input record (duration of the speech sample being processed) is reduced to that initial segment on which synchronization is present. It should be pointed out that this synchronization failure occurs very infrequently; much less than five percent of the records have the problem, but the failure must be detected because of the result when the condition is not detected. There are procedures which can be used in the recording process to detect the condition when it occurs in the tape writing process.

Once the punctuation marks have been found it is possible to determine the relative positions of the input channels within the computer words. Since there are 6 channels with 6 bits/channel, and since there are 36 bits/word in the computer, the position of the input characters are assigned a fixed position in the computer words for the input table. A table is then computed in which the location of the character for the

n^{th} filter is coded in the n^{th} entry.

The program then transfers to a subroutine which will print a graph corresponding to the input data. The letters A,B,C,D,E, and F are used for the 6 channels. If the same point on the page occurs for two different channels, only the higher occurring channel symbol occurs. A page of this type of output is shown in Figure 6-2. The label along the left margin is from the identification card for the record. The last two numbers are added by the machine. The next to the last number on the identification line corresponds to the number of the record on the input tape and the last number corresponds to the number of sample points in the record being processed. The six lines of periods are the respective time axes for the graphs, and the sample point number for each point is given across the top of the page. The current program plots the same number of graphs for each record, but it is a trivial change to read an additional card which would specify the number of graphs to be printed. This card could also be used to specify the sampling rate for each record. Currently the sampling rate is one of the inputs with the first record and is assumed constant for each run. After plotting the requested number of graphs the computer proceeds to start the analysis of the input, filter by filter.

3. Compute Filter Output Parameter Tables

The filter output parameter tables are computed by two subroutines: subroutine ZCC and subroutine ESUB, neither of which are especially mnemonic. Subroutine ZCC computes:

- a) the length of the interval between successive zero crossings,

Figure 6-2

- b) the time at which each zero crossing occurs,
- c) the nature (direction) of the crossing, and
- d) the maximum amplitude and sign of the function on the interval.

The results of a), b), and c) are stored in a table called CF and the results of d) are stored in a table called CFO. The tables have one entry per zero crossing. The results of table CFO are then used by subroutine ESUB to compute a table called ETAB which encodes each set of four consecutive crossings as either A,B,C, or D according to the scheme which is described in detail below.

The CF table is encoded in the following manner. The length of the interval is stored in the address of the word, the nature of the zero crossing is stored in the tag of the word, and the time of the zero crossing is stored in the decrement of the word.

The length of a zero crossing interval is defined as the number of non-zero points of the interval if the interval does not end with zero, and is defined as the number of non-zero points plus one when the interval ends with zero. The interval is always assumed to begin with the first non-zero point. This definition has the property that successive lengths may be added to obtain the total length when there are not consecutive zeros.

The nature of the crossing is used to encode two aspects of the zero crossing: whether the value of the filter output on the interval was positive or negative and whether the interval did or did not end exactly on zero. The four entries made in the tag of the word are:

2_8 - positive, did not end with zero

3_8 - positive, ended with zero

4_8 - negative, did not end with zero

5_8 - negative, ended with zero.

The advantage of the code is that the tag always includes a 1 if the interval ended with zero and has different bits corresponding to plus and minus. (Clearly, the above information could have been encoded with 2 bits, but there were 3 bits available and it was easier to perform the required logical checks used elsewhere in the program using the three bit code described above.)

The time, stored in the decrement, is stored relative to the end of the record. Whenever times are printed, however, they are given relative to the start of the record.

The CFO table contains an estimate of the maximum amplitude on each interval. It is a simple task to compute the maximum value of the sample points on a given interval exactly. However, such a detailed computation did not seem necessary and an approximation was used. In particular, the amplitude of the sample point which was the integer part of one-half the length of the interval from the zero crossing end of the interval was used as an estimator of the maximum amplitude. The nature of the type of error introduced by this estimation will not be belabored here. It should be pointed out that no difficulties seem to arise from the approximation, although trouble would be expected with frequencies in the second formant range at the current sampling rate.

The ETAB table computed by subroutine ESUB also has one entry per zero crossing. Its values are used in determining the fundamental

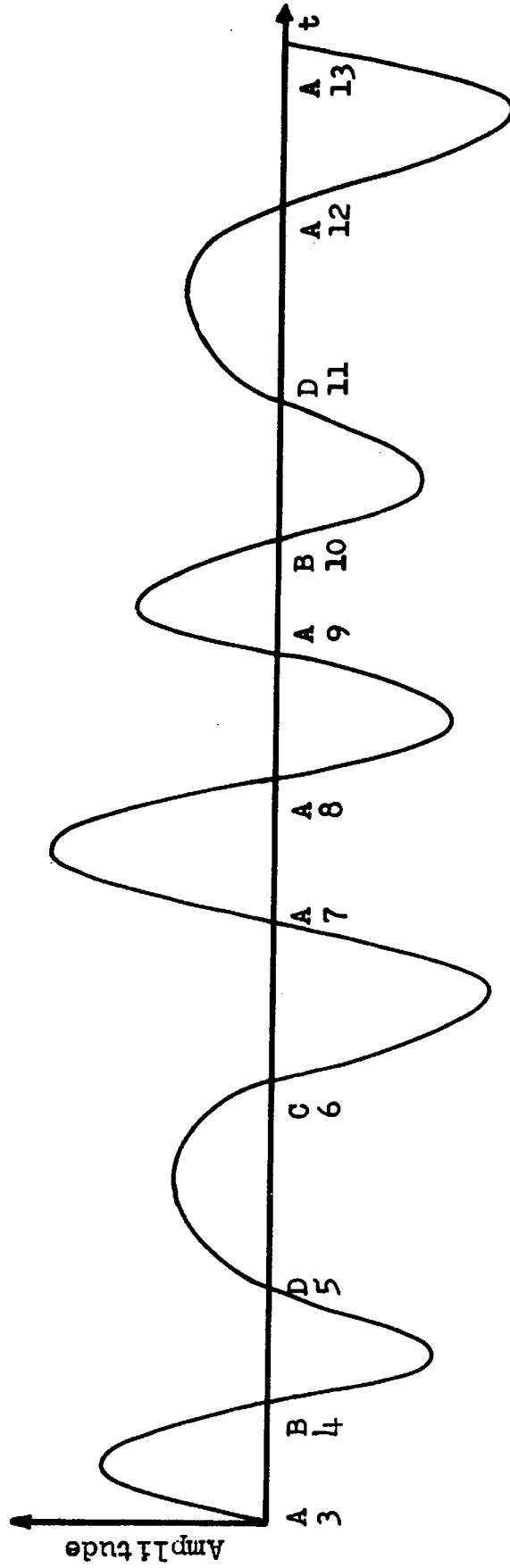
period, within which the formant is to be measured. There are four possible values for entries in the ETAB table. Let us denote the maximum value on the interval preceding the i^{th} crossing zero by CFO_i . Then the i^{th} value of the E table is defined by the following inequalities.

<u>ETAB Table</u>	<u>Conditions on CFO Table</u>
$\text{ETAB}_1 = A$ iff	$\text{CFO}_1 \geq \text{CFO}_{1+2}$ and $\text{CFO}_{1+1} \geq \text{CFO}_{1-3}$
$\text{ETAB}_1 = B$ iff	$\text{CFO}_1 \geq \text{CFO}_{1+2}$ and $\text{CFO}_{1+1} < \text{CFO}_{1-3}$
$\text{ETAB}_1 = C$ iff	$\text{CFO}_1 < \text{CFO}_{1+2}$ and $\text{CFO}_{1+1} \geq \text{CFO}_{1-3}$
$\text{ETAB}_1 = D$ iff	$\text{CFO}_1 < \text{CFO}_{1+2}$ and $\text{CFO}_{1+1} < \text{CFO}_{1-3}$

Clearly, the conditions of the CFO entries are mutually exclusive and exhaustive, that is, all conditions are covered by one and only one of the conditions. The ETAB table was motivated by the consideration of a formant in the time domain. Figure 6-3 shows a human speech formant in which the zero crossings are labelled. The letters corresponding to the ETAB table values are included under the axis crossing number. (There is not sufficient information on the graph to determine either the first E table entry or the last two entries. There was, of course, more information available in the computer.)

In case anyone should desire to replicate this section of the program it should be pointed out that if one computes the alternate values of ETAB, then the other values can be uniquely determined without reference to the CFO table. In particular, Table 6-1 could be used for the computation of alternate values.

In the time domain a formant, over the interval when the glottis



Speech Formant (Human) Showing ETAB Table Values at Axis Crossings

Figure 6-3

		ETAB _{N+1}				
		ETAB _N	A	B	C	D
ETAB _{N-1}	A	A	A	B	B	
	B	C	C	D	D	
	C	A	A	B	B	
	D	C	C	D	D	

ETAB values as a function of the
preceeding and following values

TABLE 6-1

is closed, can be exactly represented by a function of the form

$$f(t) = Ae^{-kt} \sin(vt+\phi). \quad 6.1$$

In the case of many first formants the above equation is a reasonable approximation throughout the entire period. For instance, it would be a fairly good representation of the "isolated" formant in Figure 6-3 if the effect of the filter transient could be removed. Thus, during each period there is a waveform which has the property that its envelope goes down, then up fairly quickly at the start of the next period, after which it again goes down. This is the property which was abstracted for the ETAB table.

The ETAB table is isomorphic to a string over the alphabet $\{A,B,C,D\}$. Letting the individual entries be denoted e_i and denoting concatenation by $\overline{\prod}$, the entire table can be denoted

$$\text{ETAB} = \overline{\prod_{i=1}^{\text{ZCCT}} \text{ETAB}_i} \quad 6.2$$

where ZCCT is the number of zero crossings in the output of the filter being processed. There is never a valid means of determining ETAB_1 , $\text{ETAB}_{\text{ZCCT}-1}$, or $\text{ETAB}_{\text{ZCCT}}$. Therefore, the string that is processed to determine the period boundaries is:

$$\text{ETAB} = \overline{\prod_{i=2}^{\text{ZCCT}-2} \text{ETAB}_i} \quad 6.3$$

4. Determination of Period Boundaries

The purpose of this section of the computer program is to select automatically the axis crossings that will be considered as the period boundaries. As has been previously mentioned, this decision is based solely upon the entries in the ETAB table and the signs of the function on the intervals. In fact, it is the case that the current algorithm used to determine period boundaries is based upon treating A,B, and C as one symbol and D as another symbol. That is, the algorithm is a statement about occurrences of D's and not D's in the ETAB table. The initialization of the subroutine sets NFD (mnemonic for "next first D") to the value of the entry which is the first D in the ETAB table. Before describing the algorithm formally, it is necessary to consider a few very basic properties of recurring damped sinusoids.

The formant shown in Figure 6-3 has a period (as determined by the machine) beginning at axis crossing 5 and ending at axis crossing 11. The point to be noticed is that the function has the same sign on intervals ending with the fifth and eleventh zero crossings. The property that the formant waveform has the same sign on two intervals will be called parity; non-parity will be used when the function has different values on two intervals. Thus, parity is a relation defined on pairs of intervals. It should be clear that adjacent intervals have non-parity for a damped sinusoid. Of course, if a zero is also a "tangent" at a period boundary, then the adjacent periods have parity.

The algorithm used for computing the axis crossing based upon the ETAB and CFO table is as follows. Set FD (mnemonically "first D of the

period") to the value in NFD. Anytime that there are three or more consecutive D's beginning with FD, the value of FD is increased by two until there is either one or two D's beginning with the value FD. Once the FD has been computed in the above manner, NFD is set to the axis crossing number corresponding to the first D after the first non-D after the D at FD. (In the case of Figure C-1, if $FD = 5$, then $NFD = 11$). In the case where NFD is over 20 greater than FD, the value of NFD is set equal to $FD+20$ and a formant estimation is made based upon those period boundaries. (This usually occurs in a filter not being used to measure the formant.) In this case, the technique used is called L. In the case that NFD is greater than the total number of zero crossings the program determines whether to process the next filter or to proceed to load the next record.

There are four special cases to be considered, depending upon whether there are one or two initial and one or two final D's starting at FD and NFD, respectively. These cases will be described briefly.

- a) If there is one initial D and one final D, then the technique is labelled A.
- b) If there are one initial D and two final D's, and if there is parity for the FD^{th} and NFD^{th} intervals, then the case is labelled C; if not and if there is parity for the FD^{th} and $NFD+1^{th}$ intervals, the case is labelled E; otherwise it is labelled J.
- c) If there are two initial D's and one final D, and if there is parity for the FD^{th} and NFD^{th} intervals, then the case is labelled G; if not and if there is parity for the $FD+1^{th}$ and

NFD^{th} intervals, the case is labelled D; otherwise it is labelled K.

- d) If there are two initial D's and two final D's, and if there is parity for the FD^{th} and NFD^{th} intervals, then the case is labelled B; if not and there is parity for the $FD+1^{th}$ and $NFD+1^{th}$ intervals, the case is labelled F; if not and if there is a parity for the $FD+1^{th}$ and NFD^{th} interval, the case is labelled H; otherwise it is labelled I.

The conditions associated with each of the above-mentioned techniques for the period boundaries are summarized in Table 6-2. Each of the techniques was given a label (which is carried along throughout the remainder of the program). This procedure permits later study to determine whether any of the techniques should be changed. The output graphs at the end of the program use the letter, corresponding to the technique used for estimating the formant frequency, as the symbol for the point which is plotted.

In each case where parity is obtained the parity values are put into location CTIR2 and CTIR22. In the case of techniques A and J the values of FD and NFD are used. Thus, periods are assumed to begin with the $CTIR2^{th}$ axis crossing and end with the $CTIR22^{th}$ zero crossing.

Sometimes the interval determined by the preceding algorithm is modified before computing the formant frequency. After the above algorithm has been carried out the program checks the filter output to determine whether any two successive intervals within the period have parity. If so, FD is set to the axis crossing number in NFD and the

TECH	PARITY	ETAB(FD)	ETAB(FD-1)	ETAB(NFD)	ETAB(NFD-1)
A		D	~ D	D	~ D
B	FD,NFD	D	D	D	D
C	FD,NFD	D	~ D	D	D
D	FD+1,NFD	D	D	D	~ D
E	FD,NFD+1	D	~ D	D	D
F	FD+1,NFD+1	D	D	D	D
G	FD,NFD	D	D	D	~ D
H	FD+1,NFD	D	D	D	D
I	~ FD,NFD ~ FD+1,NFD+1	D	D	D	D
J	~ FD,NFD ~ FD,NFD+1	D	~ D	D	D
K	~ FD,NFD ~ FD+1,NFD	D	D	D	~ D
L	FD + 20 < NFD → FD + 20 = NFD				

Conditions associated with each technique
for estimating period boundaries

Table 6-2

algorithm is repeated. If there is a series of two or more consecutive zero sample points within the period as determined above, then a switch labelled GO is set to MAYBE. Otherwise, GO is set to YES. The current program only accepts YES values of GO when considering the final formant frequencies. The MAYBE values seem to be too marginal for inclusion, sometimes they are good, sometimes they are bad. The program then proceeds to the next section which is described below.

5. Compute Accurate Zero Crossings and Make Estimates

The computation described in this section is done by subroutine PSUB (which is not mnemonic). The purpose of this subroutine is to compute the fundamental frequency and the formant frequency based upon the period boundaries (as given by CTIR2 and CTIR22), the CF table, and the basic input data.

The non-zero axis crossings are computed to five place (binary) accuracy, that is, to 32^{nds} of a sample time unit. This is computed by linear interpolation between the preceding and following value. The computation is carried out for the beginning and the ending of the interval from which the formant is to be estimated and for the beginning and ending of the interval from which the fundamental is to be estimated. The four parameters are called NSUBFO, NSUBF1, NADD and NADDF, where the first of each pair is for the fundamental and the second is for the formant.

The fundamental period in 32^{nds} of a sample period (NFO) is computed as the increment to be added (NADD) at the final zero crossing, minus the increment to be subtracted at the first zero crossing (NSUBFO), plus 32

times the number of sample points between the first and last zero crossings. The fundamental frequency is obtained by dividing 32 times the sampling frequency (SFLS5, mnemonically the sampling frequency left-shifted 5) by the period (NFO).

The interval used to estimate the formant frequency is determined in the following manner. The number of zero crossings in the period (NZC, mnemonically the number of zero crossings) is computed as $NZC = CTIR22 - CTIR2$. Then, based upon the number of zero crossings within the period, a prescribed number of zero crossings are removed from the beginning and end of the period. The new zero crossing numbers are kept in CST (mnemonically start) and CFN (mnemonically finish). The period from which the formant is to be estimated (NF1) in 32^{nds} of a sample point is computed as the increment to be added (NADDF), minus the increment to be subtracted, plus 32 times the interval from the CSTth zero crossing to the CFNth zero crossing.

The number of 1/2 periods (recomputed and stored in NZC) is determined as $CFN - CST$. The number of 1/2 periods (NZC) times 16 times the sampling frequency (SFLS4, mnemonically, the sampling frequency left shifted 4) is divided by the interval (NF1) in order to obtain the formant frequency.

If the whole period had GO set to MAYBE, then the interval from which the formant is to be estimated is checked to see whether the function is zero at any consecutive sample points. If it is not, the program proceeds to the record estimates section described below. Otherwise, GO is set to YES and an X is inserted in front of the technique to show a last-minute reversal. Then the program proceeds to the record estimates

section described below.

6. Record Estimates

The computation described in this section is all done in subroutine QSUB (not mnemonic). The subroutine serves two main functions: it stores the data needed for later computation and it can, by external request, print the key information computed in the section described above.

The following data are always stored for all periods and filters where GO is YES:

- 1) the fundamental frequency,
- 2) the formant frequency,
- 3) the time, and
- 4) the technique.

An example of the printed output from this subroutine is shown in Figure 6-4. The top line on the page is the label for the record. The second line shows the current ETAB table parameters, CET is the current filter and ZCCT is the length of the table. The next several lines are the current ETAB table concatenated into a string.

The results printed by subroutine QSUB begin with the header line. The items on the header will be briefly redescribed.

PER - the period number

TIME - the number of sample points from the beginning of the record to the start of the current period

TECH - the technique used to estimate the period

FO - the fundamental frequency (cps)

- NFO - the total length of the fundamental period (in 32^{nds} of a sample period)
- NADD - the increment to be added to estimate better the end of the period (in 32^{nds} of a sample period)
- NSUBFO - the increment to be subtracted to estimate better the start of the period (in 32^{nds} of a sample period)
- F1 - the formant frequency (cps)
- NF1 - the interval from which the formant frequency will be determined (in 32^{nds} of a sample period)
- NZC - number of $1/2$ periods in the formant interval of a sample period
- NADDF - the increment to be added to estimate better the end of the formant interval (in 32^{nds} of a sample period)
- NSUB1 - the increment to be subtracted to estimate better the start of the formant interval (in 32^{nds} of a sample period)
- CTIR2 - zero crossing for the start of the fundamental period
- CST - zero crossing for the start of the interval used for estimating the formant
- CFN - zero crossing for the end of the interval used for estimating the formant
- CTIR22 - zero crossing for the end of the fundamental period
- GO - used to show whether or not there were consecutive zeros in the formant interval
- F1L - the input channel (on the multiplexer) to which this output corresponds

Figures 6-2 and 6-4 correspond to the same set of data; thus, it is possible to check the previously described section in great detail with the graph from which it was determined.

7. Formant Estimate Based on All Filters

The purpose of this section of the program is to determine the final estimate of the formant frequency based upon all of the estimates from the individual filters. The particular inputs which are used in this section are printed by subroutine WPR, one output of WPR is shown in Figure 6-5. This is from the same computer run as that displayed by Figures 6-2 and 6-4.

The reason that there are no entries under the heading "filter 1" in Figure 6-5 is that the channel for the filter consisted of the raw speech wave for this run, and therefore, was not analyzed. In the procedures described in Chapter 7, the channel for filter 1 is the output of the speech synthesizer and is used as a standard for the calibration.

The following information is presented in Figure 6-5 as a function of filter number:

- 1) fundamental frequency,
- 2) first formant frequency,
- 3) time at the start of the period, and
- 4) technique used to determine the period.

The end of each of the outputs from subroutine WPR also contains a table with the following columns:

FIL - the filter number,

B/OV/	GEP	CSL TAPE 196	LIST 1 77	SP	150,450	300,600	450,750	2(6,9)	2	3264		
***** FILTER 1 ***** FILTER 2 ***** FILTER 3 ***** FILTER 4 ***** FILTER 5 *****												
F0	F1	TIME	TECH	F0	F1	TIME	TECH	F0	F1	TIME	TECH	PER
128	574	812	XA	137	537	13	X8	125	563	17	XH	1
128	564	1216	X8	128	535	301	B	130	569	73	B	2
128	571	1274	B	128	536	359	B	125	569	130	B	3
130	567	1332	X8	125	537	417	B	125	572	189	B	4
131	484	1900	A	129	541	476	B	123	567	248	B	5
179	428	1956	XC	131	539	533	B	128	567	311	H	6
132	384	2060	XH	130	540	590	B	128	571	366	B	7
132	371	2172	XB	130	540	647	B	128	571	366	B	8
				128	545	704	B	128	573	424	B	9
				130	585	1116	B	132	575	540	B	10
				128	587	1173	B	128	577	596	B	11
				128	594	1231	B	130	576	657	H	12
				130	602	1289	B	129	580	711	B	13
				129	595	1346	B	131	587	768	B	14
				129	585	1403	B	128	589	825	B	15
				158	571	1461	B	128	595	883	B	16
				129	559	1508	XB	130	594	941	B	17
				128	540	1565	XB	128	595	998	B	18
				129	528	1623	B	130	601	1056	B	19
				129	516	1681	B	129	610	1118	H	20
				130	507	1738	B	129	611	1175	H	21
				131	497	1795	B	127	620	1233	H	22
				132	495	1858	D	130	626	1291	H	23
				132	483	1914	C	130	623	1348	H	24
				132	458	2072	XC	119	621	1400	B	25
				132	455	2128	A	130	604	1462	B	26
				130	445	2184	A	131	593	1519	B	27
				132	439	2241	A	151	568	1576	B	28
				132	445	2297	XA	128	557	1625	B	29
				135	464	2353	XA	128	537	1683	B	30
								131	525	1741	B	31
								131	515	1797	B	32
								132	501	1854	G	33
								131	495	1910	A	34
								120	489	1966	A	35
								135	479	2028	XA	36
								132	464	2078	XE	37
												38
												39
***** FILTER 1 ***** FILTER 2 ***** FILTER 3 ***** FILTER 4 ***** FILTER 5 *****												
F0	F1	TIME	TECH	F0	F1	TIME	TECH	F0	F1	TIME	TECH	PER
128	574	812	XA	137	537	13	X8	125	563	17	XH	1
128	564	1216	X8	128	535	301	B	130	569	73	B	2
128	571	1274	B	128	536	359	B	125	569	130	B	3
130	567	1332	X8	125	537	417	B	125	572	189	B	4
131	484	1900	A	129	541	476	B	123	567	248	B	5
179	428	1956	XC	131	539	533	B	128	567	311	H	6
132	384	2060	XH	130	540	590	B	128	571	366	B	7
132	371	2172	XB	130	540	647	B	128	571	366	B	8
				128	545	704	B	128	573	424	B	9
				130	585	1116	B	132	575	540	B	10
				128	587	1173	B	128	577	596	B	11
				128	594	1231	B	130	576	657	H	12
				130	602	1289	B	129	580	711	B	13
				129	595	1346	B	131	587	768	B	14
				129	585	1403	B	128	589	825	B	15
				158	571	1461	B	128	595	883	B	16
				129	559	1508	XB	130	594	941	B	17
				128	540	1565	XB	128	595	998	B	18
				129	528	1623	B	130	601	1056	B	19
				129	516	1681	B	129	610	1118	H	20
				130	507	1738	B	129	611	1175	H	21
				131	497	1795	B	127	620	1233	H	22
				132	495	1858	D	130	626	1291	H	23
				132	483	1914	C	130	623	1348	H	24
				132	458	2072	XC	119	621	1400	B	25
				132	455	2128	A	130	604	1462	B	26
				130	445	2184	A	131	593	1519	B	27
				132	439	2241	A	151	568	1576	B	28
				132	445	2297	XA	128	557	1625	B	29
				135	464	2353	XA	128	537	1683	B	30
								131	525	1741	B	31
								131	515	1797	B	32
								132	501	1854	G	33
								131	495	1910	A	34
								120	489	1966	A	35
								135	479	2028	XA	36
								132	464	2078	XE	37
												38
												39

Summary Printout of Analysis of Each Individual Filter

Figure 6-5

LOW - the lowest estimate of the formant frequency

CTR - the center value of the estimates of formant frequency

UP - the highest estimate of the formant frequency

E(F1) - the mean of the estimates of the formant frequency

V(F1) - a measure of the variance of the estimates of the
formant frequency

N - the number of periods measured

E(F0) - the mean of the fundamental frequency, and

V(F0) - a measure of the variance of the estimates of the
fundamental frequency.

The above items are computed according to the following formulas:

$$\text{CTR} = \frac{\text{UP} + \text{LOW}}{2} \quad 6.4$$

$$\text{E}(\text{F1}) = \frac{1}{N} \sum_{i=1}^N F_1, \quad 6.5$$

where F_1 is the estimate of the first formant frequency on the i^{th} period,

$$\text{V}(\text{F1}) = \frac{1}{N} \sum_{i=1}^N (F_1 - \text{E}(\text{F1}))^2, \quad 6.6$$

$$\text{E}(\text{F0}) = \frac{1}{N} \sum_{i=1}^N f_1, \quad 6.7$$

where f_1 is the estimate of the fundamental frequency on the i^{th} period,
and

$$V(FO) = \frac{1}{N} \sum_{i=1}^N (f_i - E(FO))^2. \quad 6.8$$

It should be clear that the entries in the above table are not of great interest when the input consists of normal speech. For steady-state inputs, however, the above items demonstrate the steadiness of the system as an estimation device. Some of the above parameters of the output are used in the calibration results.

Only the formant frequency and the time at the start of the period are needed as inputs to the section of the program which makes the first formant estimation. Actually, only the formant frequency and time at the start of the period are used to determine the final estimate; the other two items are carried along as superfluous but interesting information.

The following definitions allow a simple statement of the algorithm used to make the final formant estimation. A formant frequency estimate from a single filter will be called acceptable if it lies in the center half of that filter's pass-band. A given formant frequency will be called confirmed if it is acceptable, if there exists another filter which has a formant frequency estimation within 50 cps of the given formant frequency estimation, and if the other filter has an epoch within 10 sample points of the given epoch. The algorithm used for final selection is:

1. If any acceptable estimate does not lie within 30 sample points of another acceptable estimate, then use the acceptable estimate as the final estimate of the formant frequency.

2. If an acceptable estimate lies within 30 sample points of another acceptable estimate, then:
 - a. if there are no confirmed estimates, disregard all of the above values;
 - b. if there is only one confirmed estimate, use it;
 - c. if there are two or more confirmed estimates, take the lowest in frequency unless it confirms the output of the filter used for the last estimation; in which case use the output of the filter employed in the last estimation.
3. Proceed until all acceptable estimates have been processed according to the above rules.

An example of the printed output of the above subroutine is shown in Figure 6-6. It corresponds to the same run as all of the other figures in this chapter. The columns are relatively straightforward:

- FO - the fundamental frequency
- F1 - the first formant frequency
- TIME - the beginning of the period (in terms of sample points)
- TECH - the technique used for estimating the period plus codes corresponding to the following:
 - a. a slash under the "E" in "TECH" signifies the value was acceptable,
 - b. an X under the "T" in "TECH" signifies that the value was confirmed, and
 - c. an X in the column preceding the "T" in "TECH" signifies the existence of two confirmed values

B/DV/	GEP	CSL TAPE 196	LIST 1 77	SP	150,450	300,600	450,750	216,9)	2	3264
FO	F1	TIME	TECH	FIL	PER					
125	563	17	X/KH	4	1					
130	569	73	/ B	4	2					
123	567	248	/ B	4	3					
128	567	311	/ H	4	4					
128	571	366	X/ B	4	5					
128	575	424	X/ B	4	6					
128	573	482	X/ B	4	7					
132	575	540	X/ B	4	8					
128	577	596	X/ B	4	9					
129	580	711	X/ B	4	10					
131	587	768	X/ B	4	11					
128	589	825	/ B	4	12					
128	595	883	/ B	4	13					
130	594	941	X/ B	4	14					
130	601	1056	X/ B	4	15					
129	610	1118	X/ H	4	16					
129	611	1175	X/ H	4	17					
127	620	1233	X/ H	4	18					
130	623	1291	X/ H	4	19					
119	621	1400	X/ H	4	20					
130	604	1462	X/ B	4	21					
131	593	1519	X/ B	4	22					
151	568	1576	X/ B	4	23					
128	557	1625	X/ B	4	24					
131	525	1683	X/ B	4	25					
131	497	1741	X/ B	4	26					
132	495	1795	X/ B	4	27					
132	495	1858	X/ D	3	28					
132	483	1914	X/ C	3	29					
132	458	2072	X/ XG	3	30					
132	455	2128	/ A	3	31					
132	439	2241	/ A	3	32					
132	445	2297	/ XA	3	33					
135	464	2353	/ XA	3	34					
					35					
UPL E(F1)	V(F1)	LOW1	UP0 E(F0)	V(F0)	LOW0	NUM				
626	557	3007	439	151	129	21	119	35		

List of First Formant Frequency Estimates
Figure 6-6

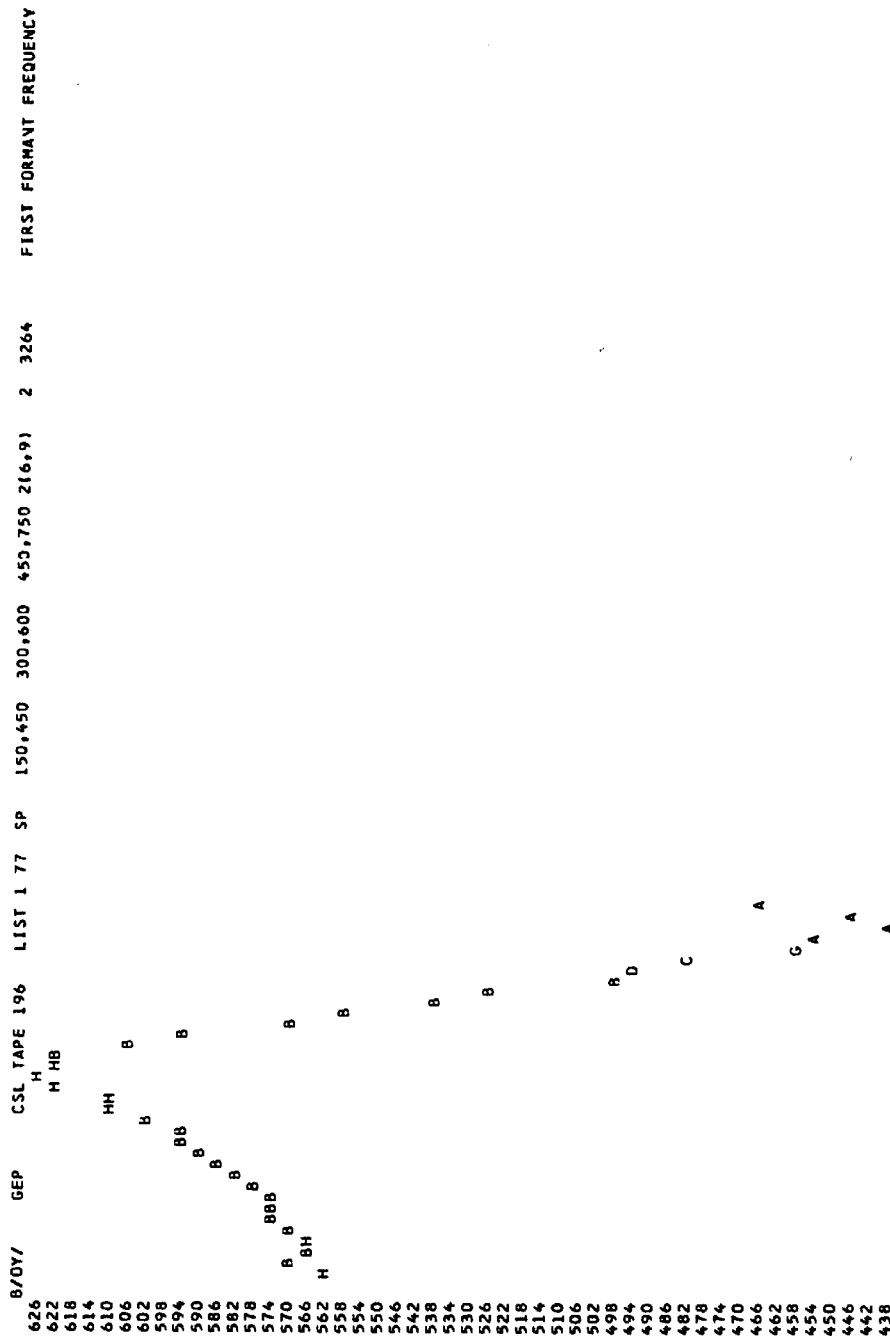
near the epoch of this period.

FIL - the filter from which the estimation was taken

PER - the consecutive period number assigned to this estimate.

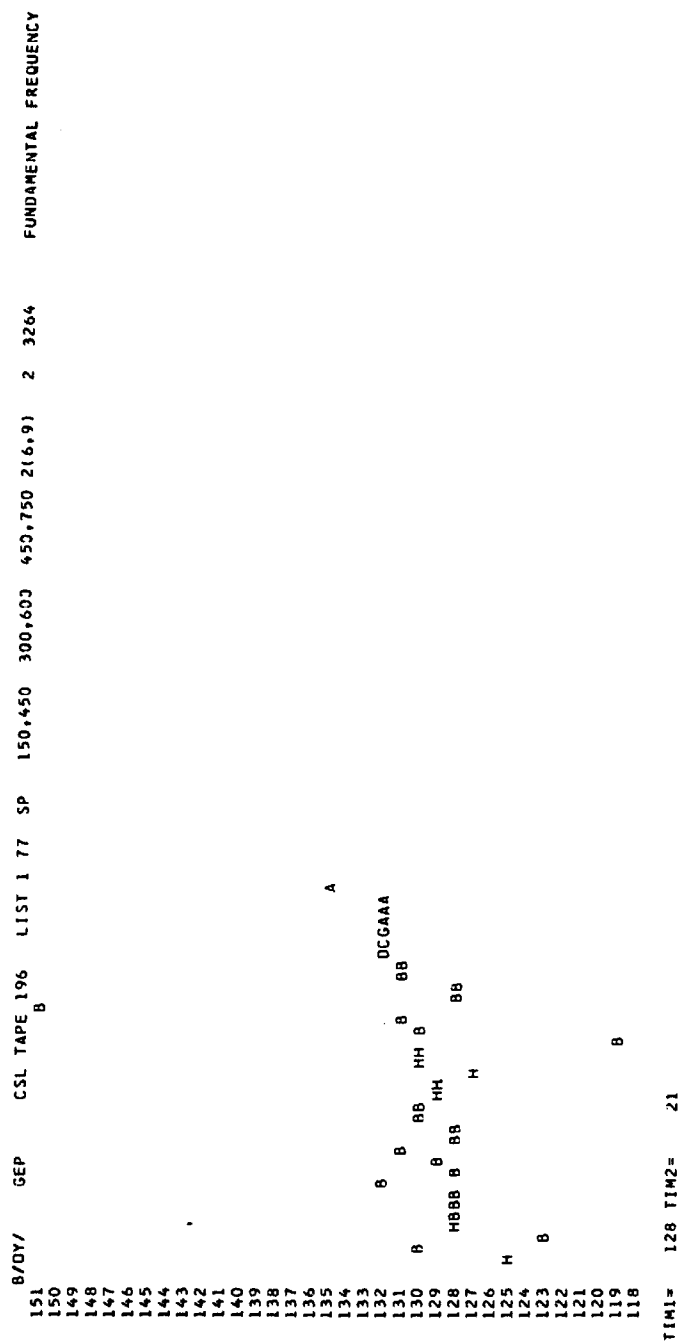
In addition to the table output, the output estimates of the formant frequency are graphed as a function of the period showing the technique as the symbol used to plot the point. The fundamental frequency is also plotted as a function of the period showing the technique used by the symbol employed to plot the point. The graphs for the formant frequency and fundamental frequency for the data previously presented in this chapter are shown in Figure 6-7 and 6-8.

After the final graph is printed, control is returned to the executive system if all input data have been processed; otherwise, the program processes the next record.



Graph of the First Formant Frequency of [ɔɪ]

Figure 6-7



Graph of the Fundamental Frequency of [ɔɪ]

Figure 6-8

Chapter 7

Experimental Study of System Accuracy

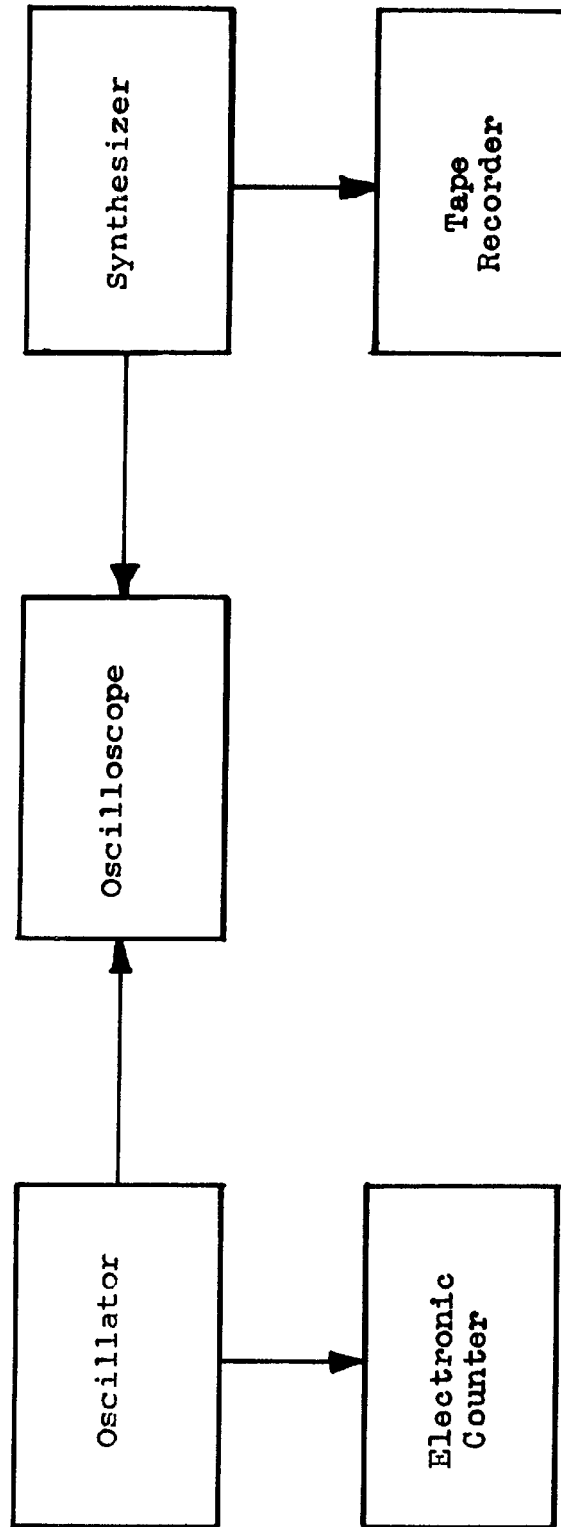
1. Introduction

The study of system accuracy was divided into two parts. One study was designed to determine the effect of the fundamental frequency upon the formant measurement; the other was designed to determine the exactness with which the filters measured the formant frequency.

2. Variation as a Function of Fundamental Frequency

In order to determine the variation as a function of fundamental frequency, a single formant was analyzed for eleven different fundamental frequencies. A single formant was obtained from the speech synthesizer being developed at the Communication Sciences Laboratory by Giles Peterson. Using the formant tracking system (without a filter) the formant frequency was found to be 460 cps. The synthesizer has two alternate control systems, manual (by setting knobs) or automatic (where the control signals are demultiplexed from a digital tape). The manual control was used.

The system diagram used for determining the fundamental frequency is shown in Figure 7-1. The output of the synthesizer was fed to the vertical plates of an oscilloscope and to a recorder. The output of an oscillator was fed to the horizontal plates of the oscilloscope and to



Hewlett Packard Model 522B Electronic Counter
Hewlett Packard Model 650A Test Oscillator
Tektronic Type 555 Dual Beam Oscilloscope
Communication Sciences Laboratory Synthesizer
Ampex 300 tape recorder (4-channel)

System Diagram of Method Used to Measure Fundamental Frequency

Figure 7-1

an electronic counter. The oscillator frequency was shifted to the desired value, then the synthesizer was shifted until the pattern on the oscilloscope showed that the synthesizer was at the desired fundamental frequency. The formant was then recorded while a stable pattern on the oscilloscope was noted. Eleven values of fundamental frequency were used in the study ranging from 70 cps to 170 cps. Approximately 36 periods were analyzed for each fundamental frequency. The exact values of fundamental frequency are given in Table 7-1.

The following statistics were obtained for each fundamental frequency:

- 1) the highest estimate of formant frequency,
- 2) the lowest estimate of formant frequency,
- 3) the average estimate of formant frequency, and
- 4) the variance of the formant frequency.

The variance was computed by

$$V(F_1) = \frac{1}{n} \sum_{i=1}^n (F_1 - E(F_1))^2 \quad 7.1$$

where

$$E(F_1) = \frac{1}{n} \sum_{i=1}^n F_1. \quad 7.2$$

and F_1 was the estimate of the formant frequency on the i^{th} period.

The mean and variance are included in Table 7-1 as a function of the fundamental frequency.

Fundamental Frequency	LOW	E(F1)	UP	V(F1)	N
70	458	459	462	1	18
80	456	458	462	2	23
90	458	460	462	1	33
100	450	451	456	5	5
110	450	456	462	4	39
120	450	454	457	2	38
130	439	449	458	37	40
140	445	450	456	6	51
150	457	463	471	13	52
160	455	460	464	3	59
170	452	455	461	4	54

Formant frequency estimates as a function of
fundamental frequency for a fixed formant frequency of 460 cps

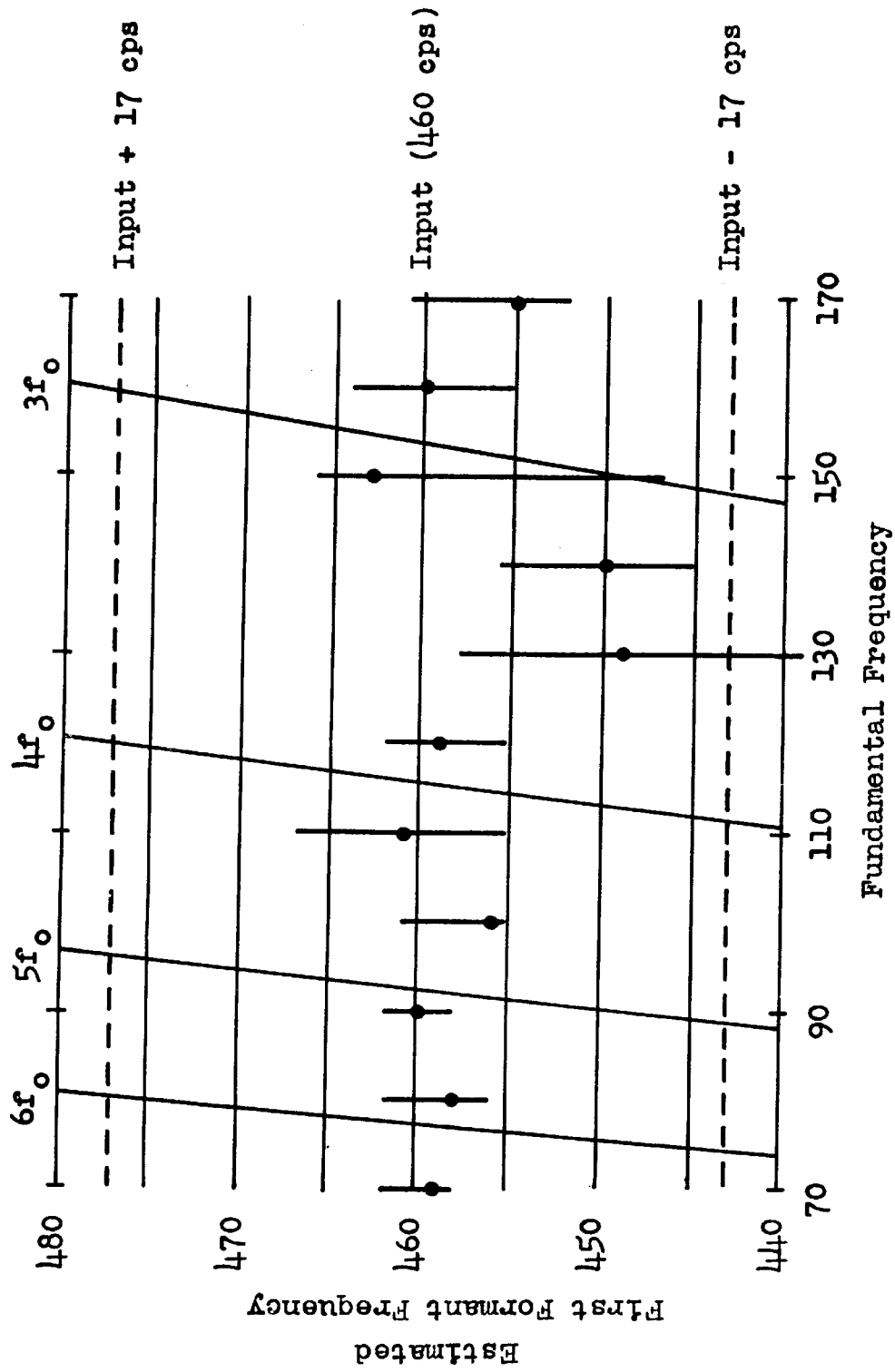
Table 7-1

Figure 7-2 is a graph showing the range and mean of the estimates obtained for each value of the fundamental frequency. The solid lines passing through the graph are the harmonic frequencies. The estimate of the formant frequency appears to be slightly influenced by the harmonics. It can be seen that the average values tend toward the harmonics. It should be clear, however, that the amount of variation is not so large as one might first infer without consideration for the size of the scale used in the graph. For instance, in the Bell Telephone Laboratories Spectrograph the spacing between lines is 17 cps. If one line of a spectrograph were at the input formant frequency, 460 cps, then the first line above the formant frequency would be at 477 cps and the first line below the formant frequency would be at 443 cps. These are the dashed lines drawn on the figure. Only two of the 390 points measured lay on or outside these lines.

Therefore, if a maximum individual error of about 4.5% (the worst error in the 390 periods measured) can be tolerated, then this system can be assumed free of influence of the fundamental frequency for fundamental frequencies between 70 cps and 170 cps. If one excludes the worst single estimate, then the maximum error is only 3.8%, which is close to Flanagan's difference limen for formant frequency.

3. Accuracy of Formant Frequency Measurements

The synthesizer was also used as the input to determine the accuracy of the response of each of the filters to various formant frequencies. For determining the accuracy of the formant frequency the fundamental was held constant at 77 cps.



Estimated First Formant Frequency as a Function of Fundamental Frequency

Figure 7-2

The program was again used to check itself. The formant input (unfiltered) was analyzed by the computer, and the average value of this result was used as a measure of the input formant frequency. The output of each filter was analyzed for each input and the means and variances of the results are presented in Table 7-2. In some cases, the variance is zero. This is generally the result of only using one period for estimation. Only the results which would be used for final estimates are presented. Thus, there are not generally as many periods analyzed for a filter which is not near the formant.

The average measurements for each of the filters is presented in tabular form in Table 7-2 and graphically in Figures 7-3 to 7-6.

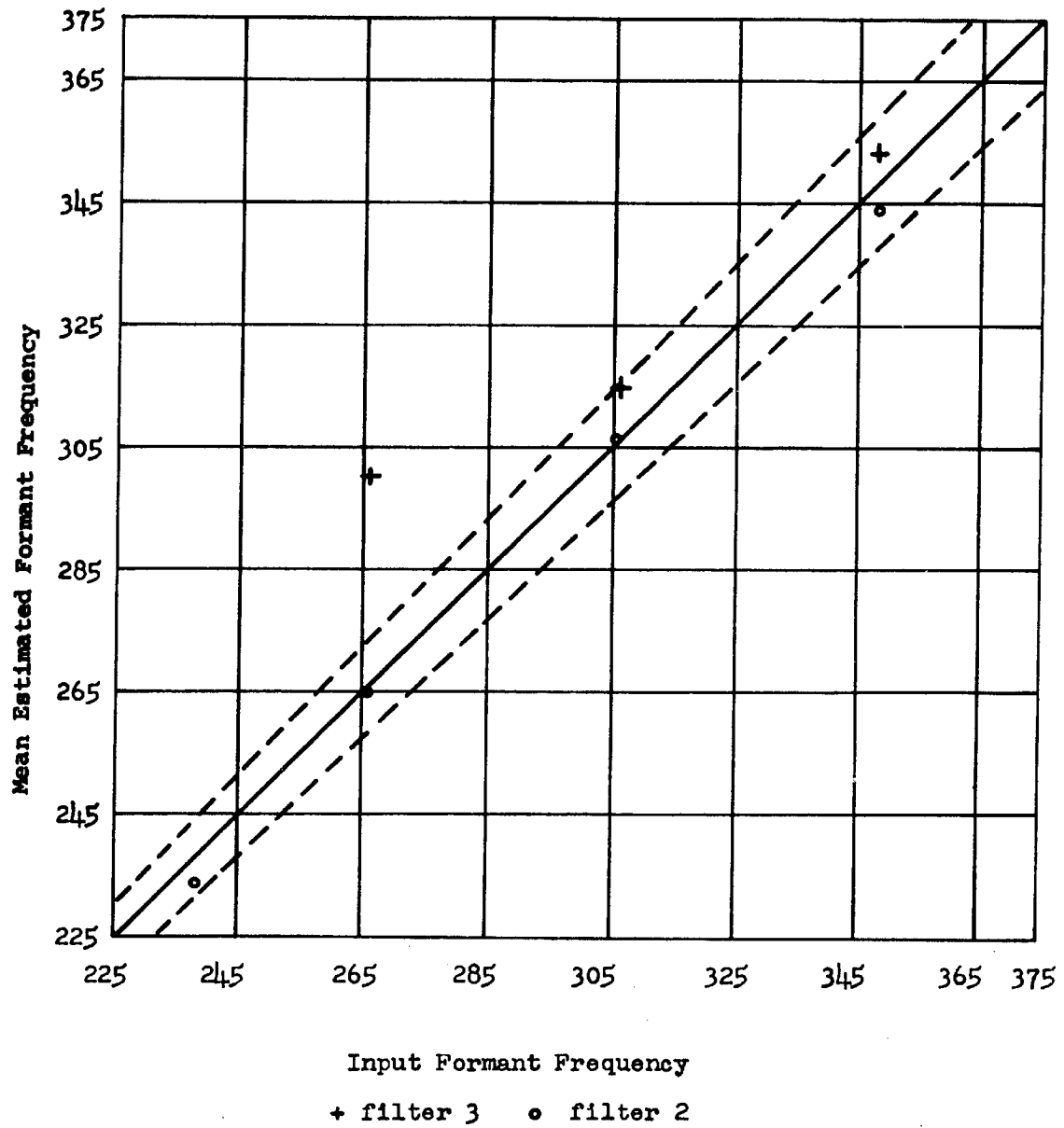
The frequency calibration is shown in four graphs, one for the acceptable region of each of the four filters used in this study. Figure 7-3 shows the calibration in the acceptable region for filter two, Figure 7-4 for filter three, Figure 7-5 for filter four, and Figure 7-6 for filter five. The circles on the graphs correspond to the acceptable filter for the given region.

The dashed lines correspond to $\pm 3\%$ from the center line, the difference limen for formants determined by Flanagan. It can easily be seen that the average output from the acceptable filter always falls within the $\pm 3\%$ lines. It can also easily be seen that in all cases where the input formant was 306 cycles or above there were at least two filters which had average values within the $\pm 3\%$ lines; in fact, many points have three filters within the $\pm 3\%$ lines. It is this feature of "confirming" (discussed in Chapter 6) which is used as a form of redundancy for determining the final estimate of the formant

Input $E(F_1) V(F_1)$	Filter 1 $E(F_1) V(F_1)$	Filter 2 $E(F_1) V(F_1)$	Filter 3 $E(F_1) V(F_1)$	Filter 4 $E(F_1) V(F_1)$
238	3	2	-	-
266	19	3	-	-
306	2	1	-	-
348	1	3	-	-
380	1	1	1	-
400	5	0	0	-
436	1	3	26	-
473	1	2	6	-
486	2	4	8	5
530	1	5	6	50
563	10	1	1	104
616	3	20	4	56
637	3	21	4	15
698	4	39	6	17
722	10	4	3	8
769	4	-	4	25

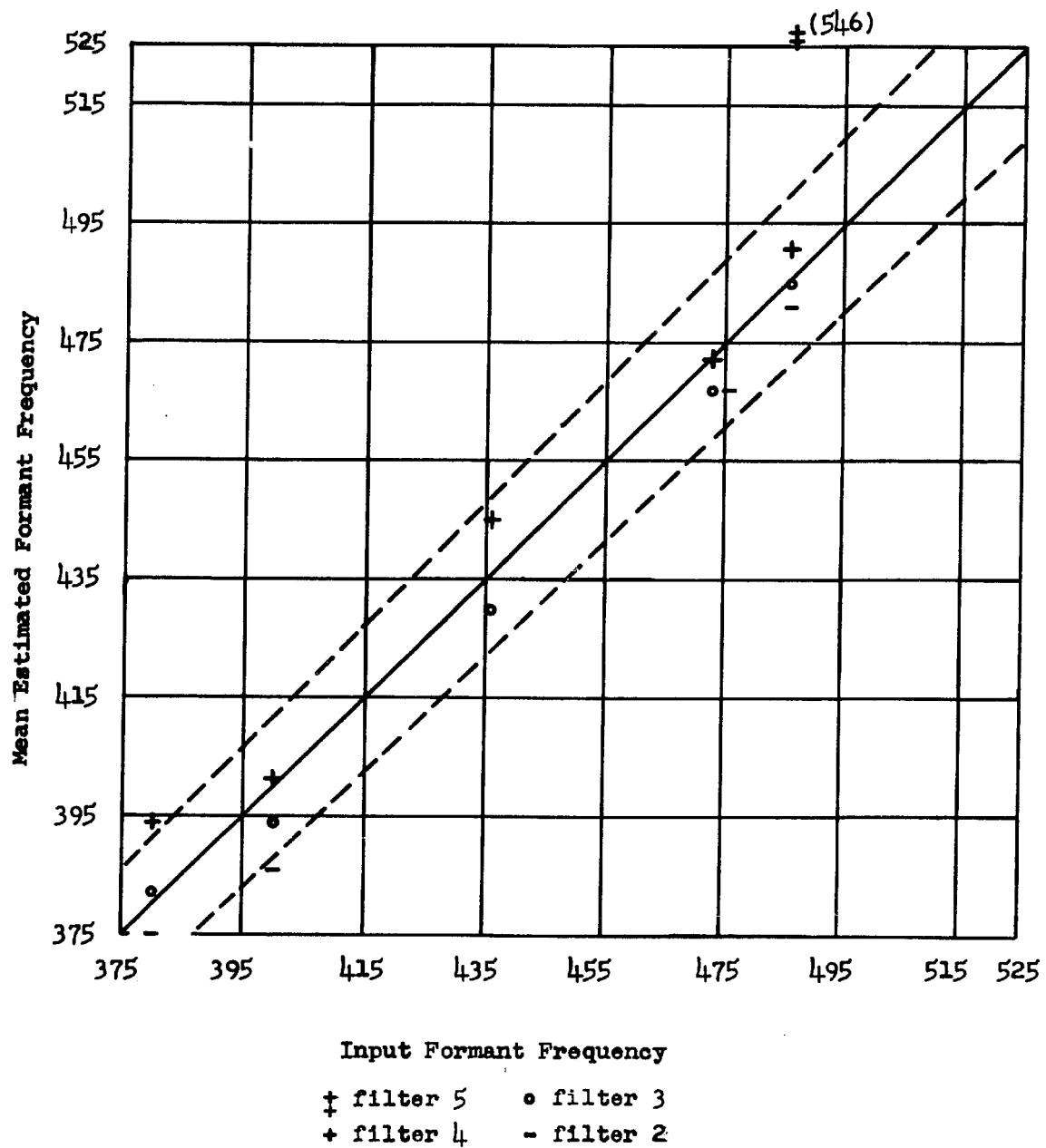
Formant frequency calibration

Table 7-2



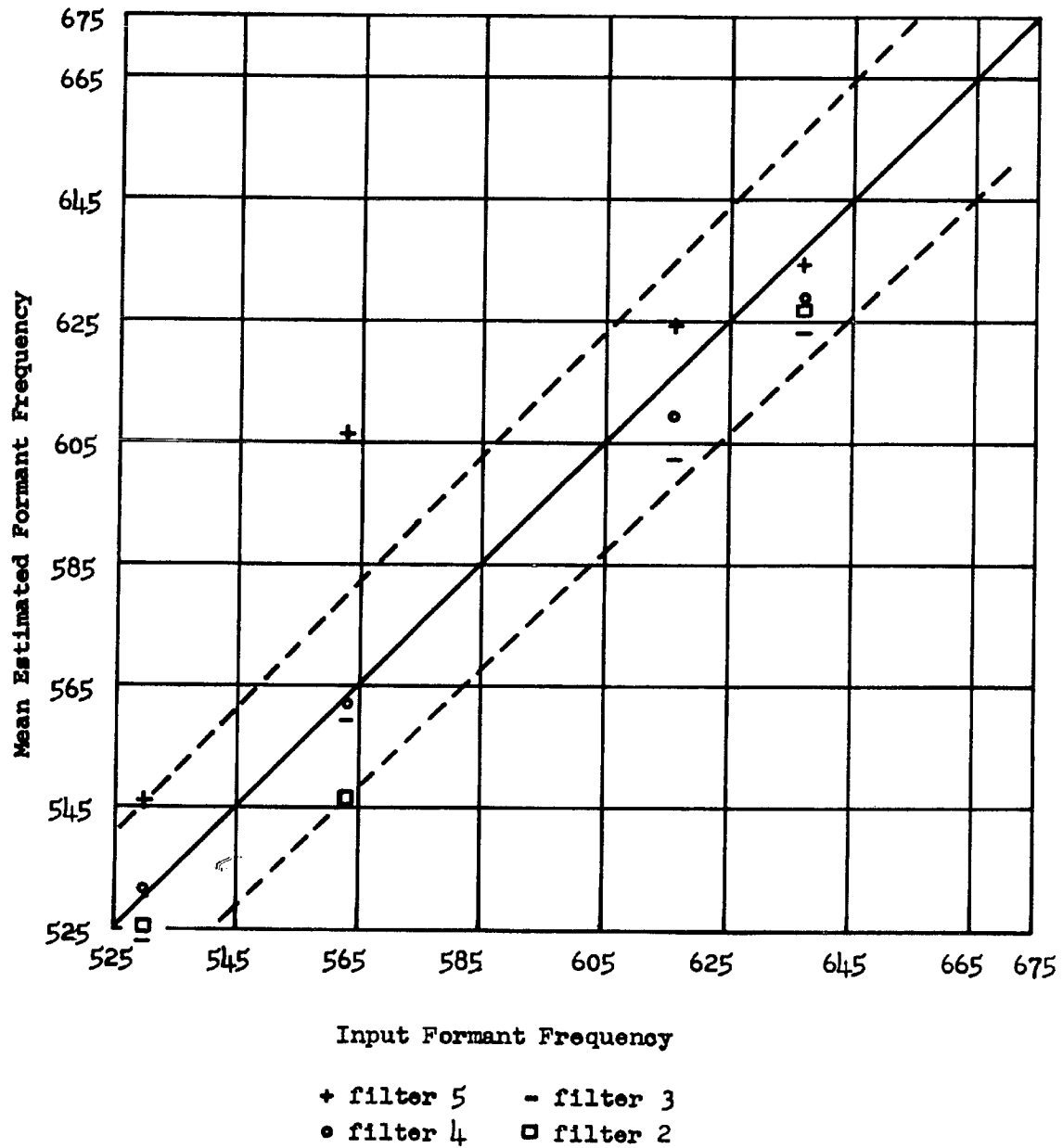
Formant Frequency Calibration: 225-375 cps

Figure 7-3



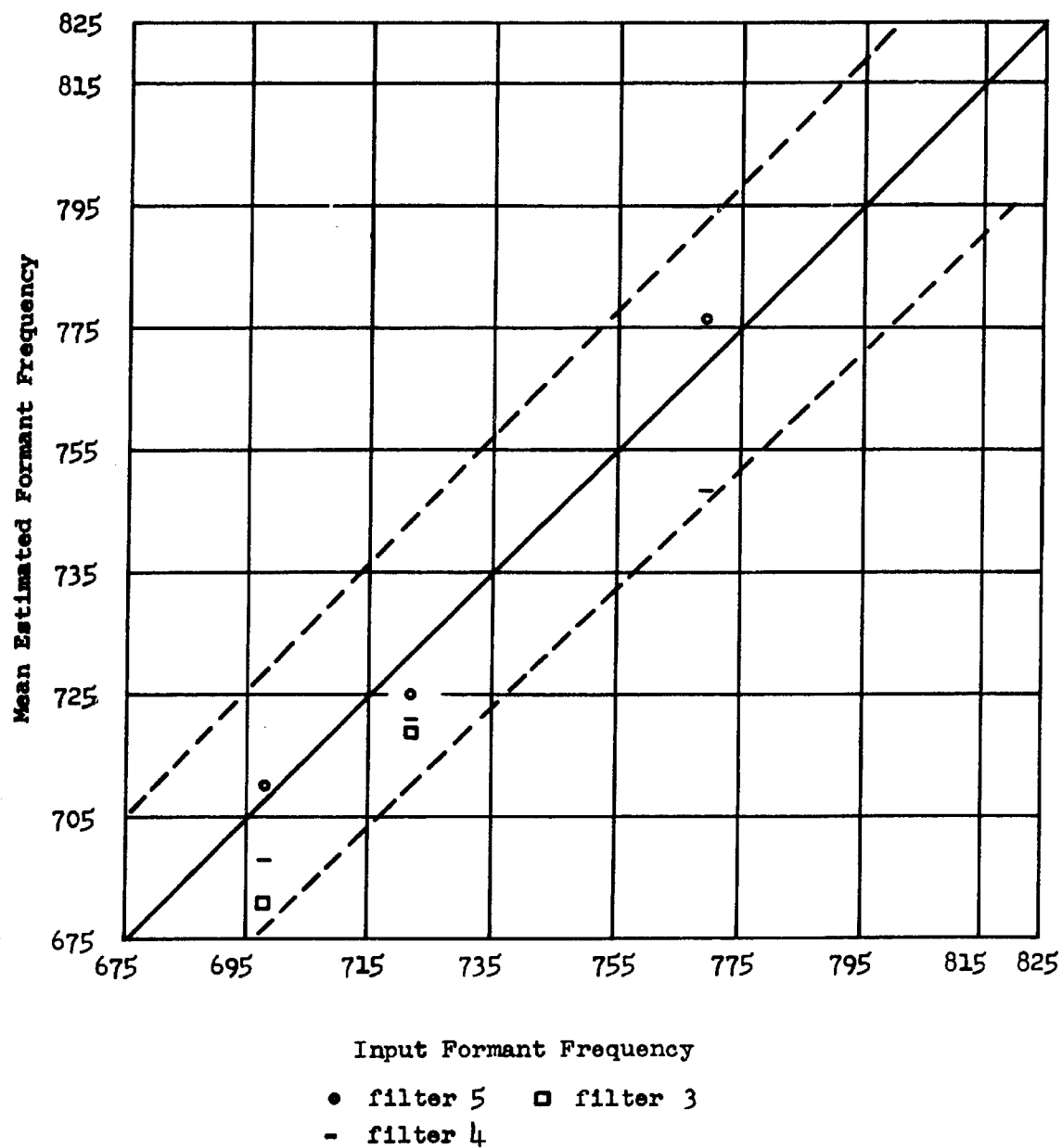
Formant Frequency Calibration: 375-525 cps

Figure 7-4



Formant Frequency Calibration: 525-675 cps

Figure 7-5



Formant Frequency Calibration: 675-825 cps

Figure 7-6

frequency.

4. Summary

The above results show that the fixed filter scheme of tracking formants is reasonably accurate and reasonably independent of fundamental frequency over a limited range. The range of application could be increased either by changing the type of filter used or by increasing the bandwidths of the filters which are used. It is felt that a scheme of the sort being investigated can achieve accuracies within the difference limen for vowel formants by the judicious selection of a filter bank.

Chapter 8

Results of Speech Formant Tracking

The purpose of a formant tracker is, after all, to track formants in human speech. A considerable number of speech samples have been processed by the system described in the preceding pages in order to study its performance. The purpose of this chapter is to present some of the results of tracking first formant frequencies in actual speech. Three different types of speech inputs were used for the illustrations presented.

- a) isolated vowels and diphthongs,
- b) isolated CNC¹ words, and
- c) isolated words recorded in normal room noise.

The isolated vowels and diphthongs were taken from the tapes used by Lehiste and Peterson [69] which were recorded by G. E. Peterson. The isolated CNC words were from a commercially available² tape of CNC words. The final set of isolated words was recorded in a laboratory with blower noise from various instruments in the background. The recordings were made to determine whether the system would be as

¹Words with a syllabic structure which consists of a consonant phoneme, followed by one or more syllable nucleus phonemes, followed by a consonant phoneme.

²CNC word lists one through five, Recording Hi-Fi Studio, Ann Arbor, Michigan.

insensitive to such a noise background as one would heuristically expect it to be.

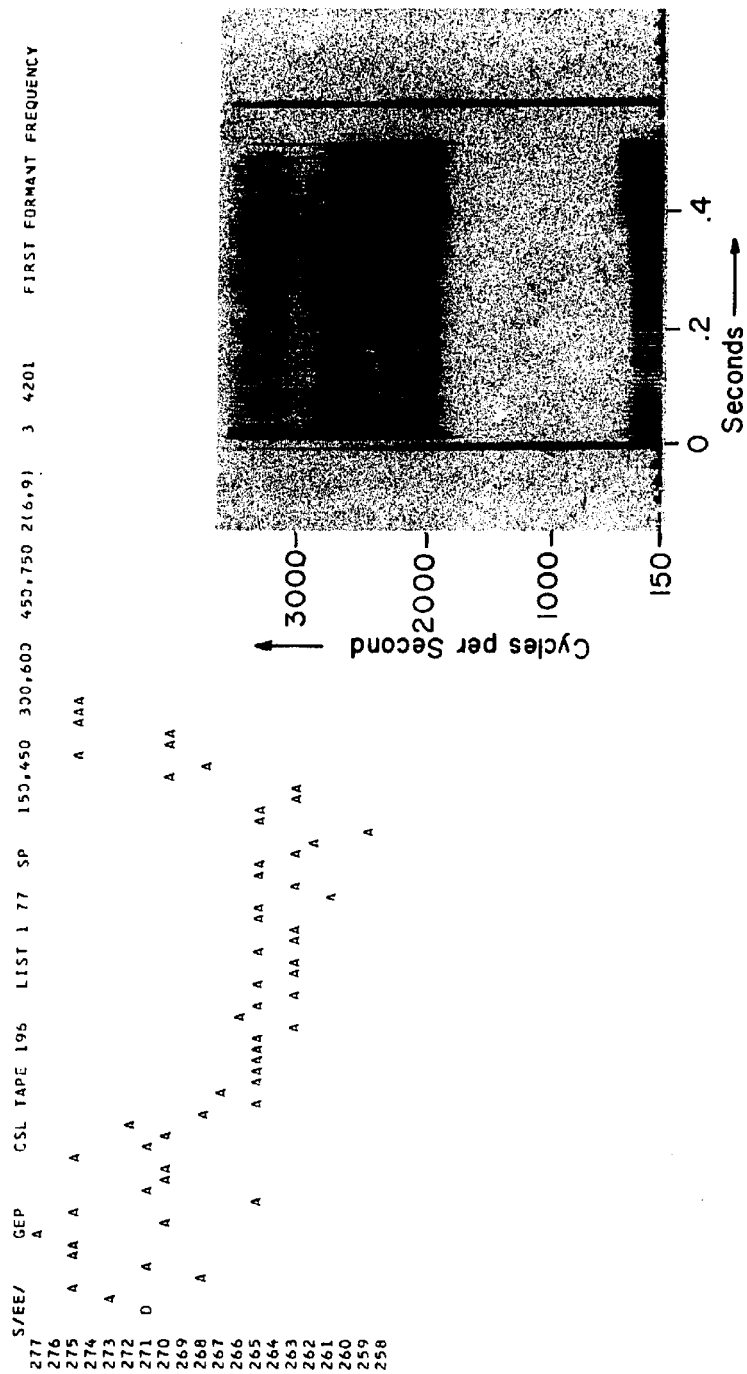
In order to show the results of some of the analyses by the formant tracker, a series of figures will be presented for each speech input discussed in the chapter. These figures include the following:

- 1) the graph drawn by the computer to represent the first formant, and
- 2) a spectrogram of the speech signal.

The pulses used to activate the digital processing equipment have been mixed with the speech wave and can be seen on the spectrograms. There is a slight delay from the first pulse until the digital tape is started.

1. Analysis of Isolated Vowels

One of the most easily isolated first formants occurs in the phone type, [i]. Figure 8-1 shows a spectrogram of an [i] and the results of estimating the first formant for this phone type. The spectrogram was made with a 300 cps analyzing filter. A slow sweep was employed so that 3400 cps are displayed in four inches. The graph drawn by the computer has a very unusual scale, namely, one cycle per line. Close inspection of Figure 8-1 will show that the average estimate of formant frequency is 267 cps and that all estimates fall within ± 10 cps of the average. The range of the estimate is only 18 cps, which is, of course, only one cycle per second more than the spacing between successive lines on a standard two-inch spectrogram. Because of the crudeness of the scale on the sound spectrogram relative to such calculations, there is obviously little to be gained from plotting the formant estimates on the spectrogram.



Graph and Spectrogram for an Isolated [i]

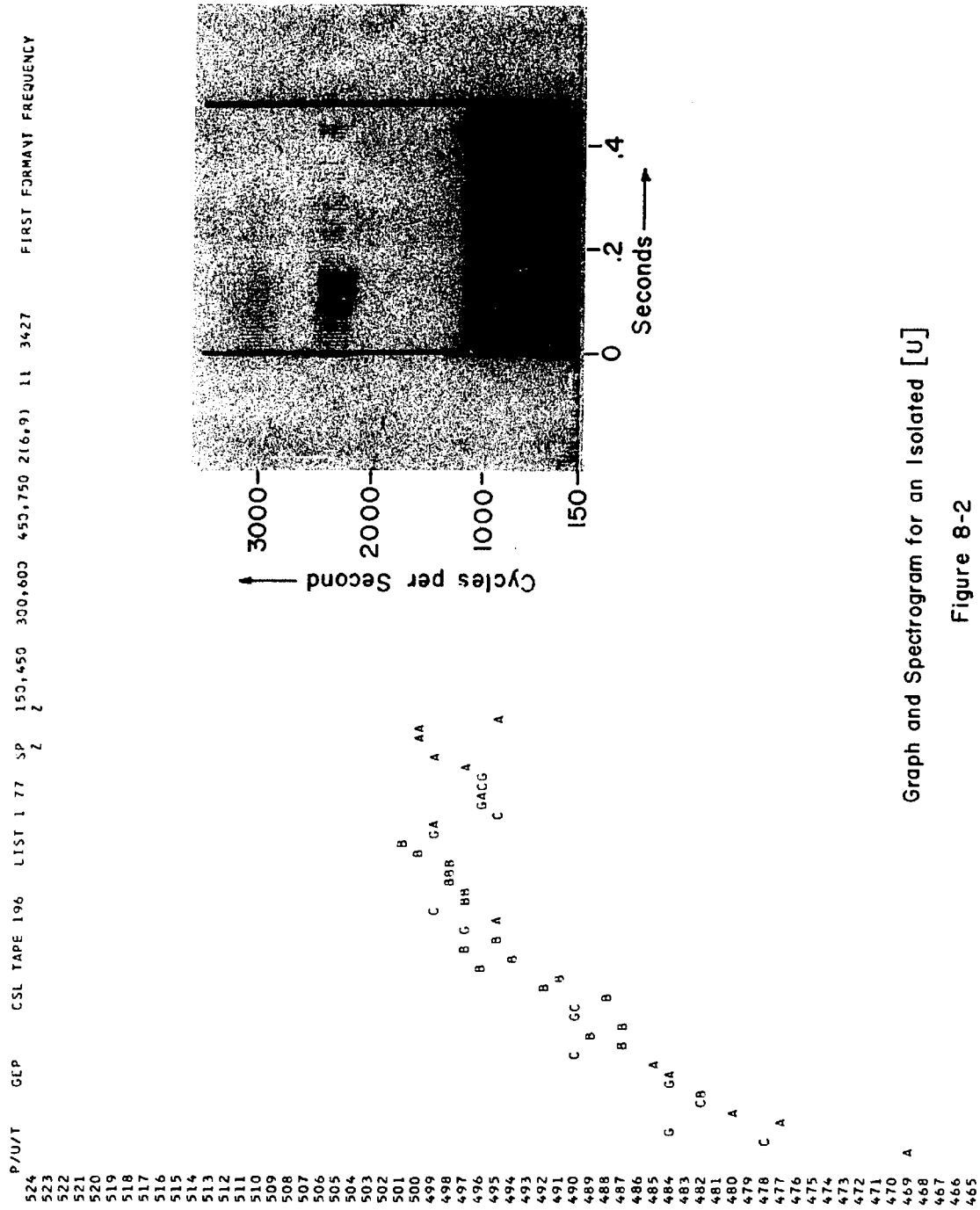
Figure 8-1

The second vowel to be considered is [U], for which analyses are presented in Figure 8-2. A rather marked rise can be seen on the graph drawn by the computer, again at one cps per line. One is tempted to consider the rise to be attributable to a natural frequency characteristic of the filter; that is, to the fact that the center frequency of the filter is 450 cycles. There is certainly nothing one can say, using only Figure 8-2, to refute this possibility. The variation, however, is most likely not due to the filter and will be further discussed when the results of processing [I] are presented. The two "Z's" at the top of the graph signify points that were off scale; they correspond to frequencies of 554 and 530 cps, respectively.

The results of processing the vowel [I] are shown in Figure 8-3. The formant tracking program was instructed to accept all frequencies from filter 3 in the range of 300 to 525 cps, which is a shift from its normal range of 375 to 525 cps. This shift in range was used to show that a first formant transition very much like the transition in [U] also occurred in [I]. The analysis was made using the same filter with the same settings used for analyzing the [U]. However, in this case, the transition is not from the center frequency of 450 cps; it is more nearly from a frequency of 350 cps. It seems reasonable to conclude that this type of transition is not due to a natural resonance of the filter.

2. Analysis of an Isolated Diphthong

The graph of the first formant of the diphthong [OI] is presented



Graph and Spectrogram for an Isolated [U]

Figure 8-2

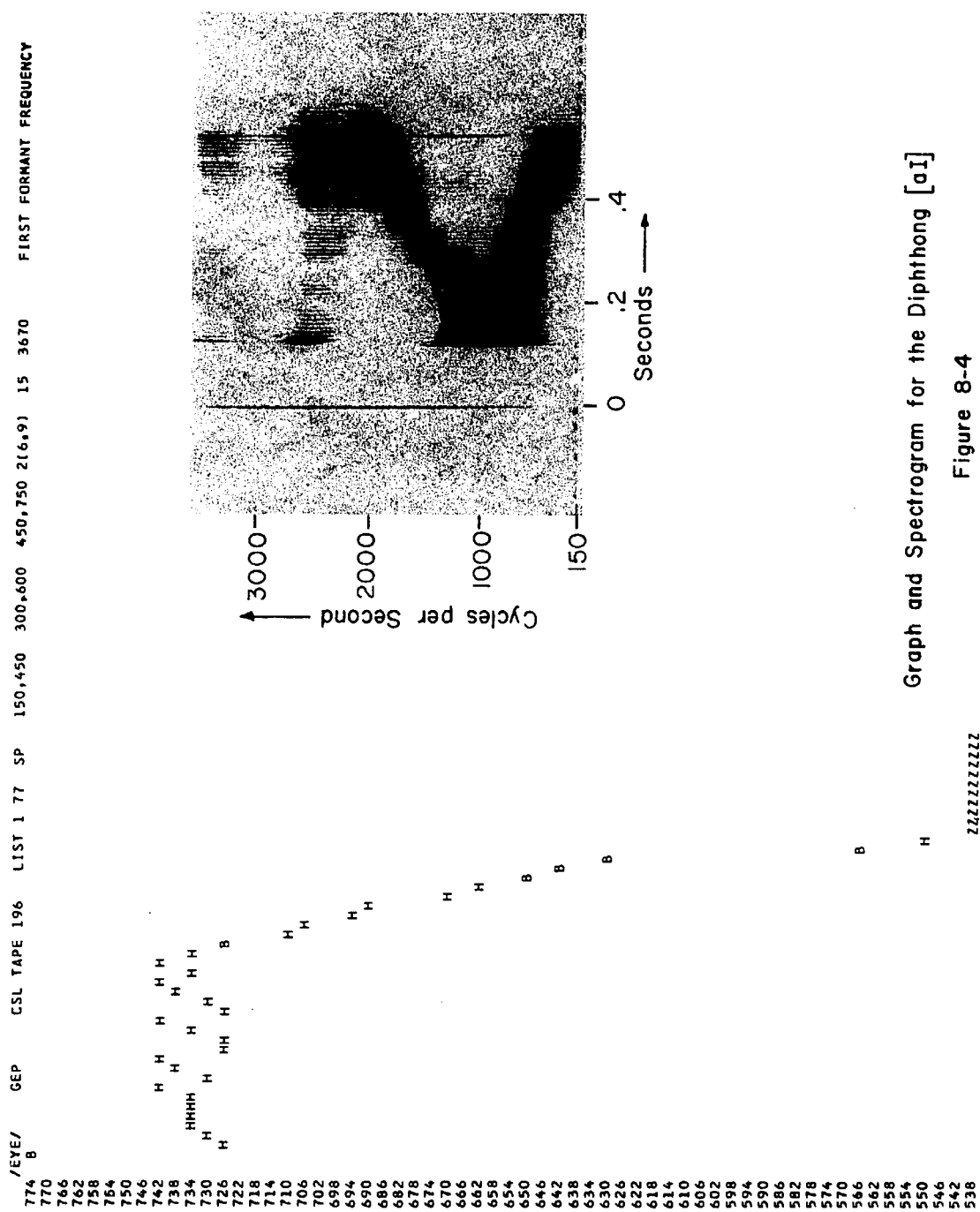
as Figure 6-7 in Chapter 6.

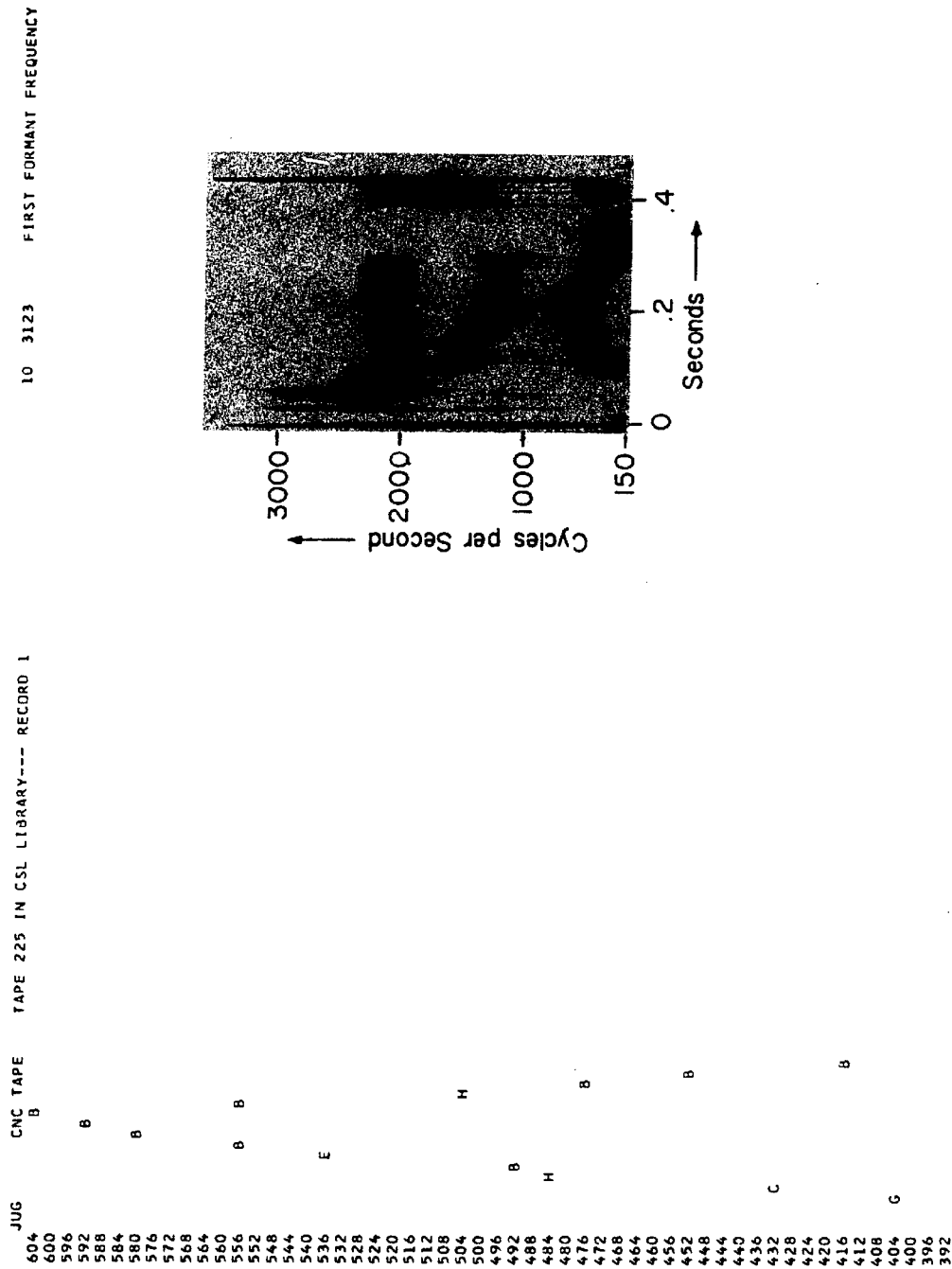
A spectrogram and analysis of [aI] are presented in Figure 8-4. The final "Z" entries correspond to periods 35 to 45 in the output table and their respective frequencies go from 508 down to 405 cps. Some of the final periods of the [aI] were not analyzed due to the falling intensity.

3. Analysis of CNC words

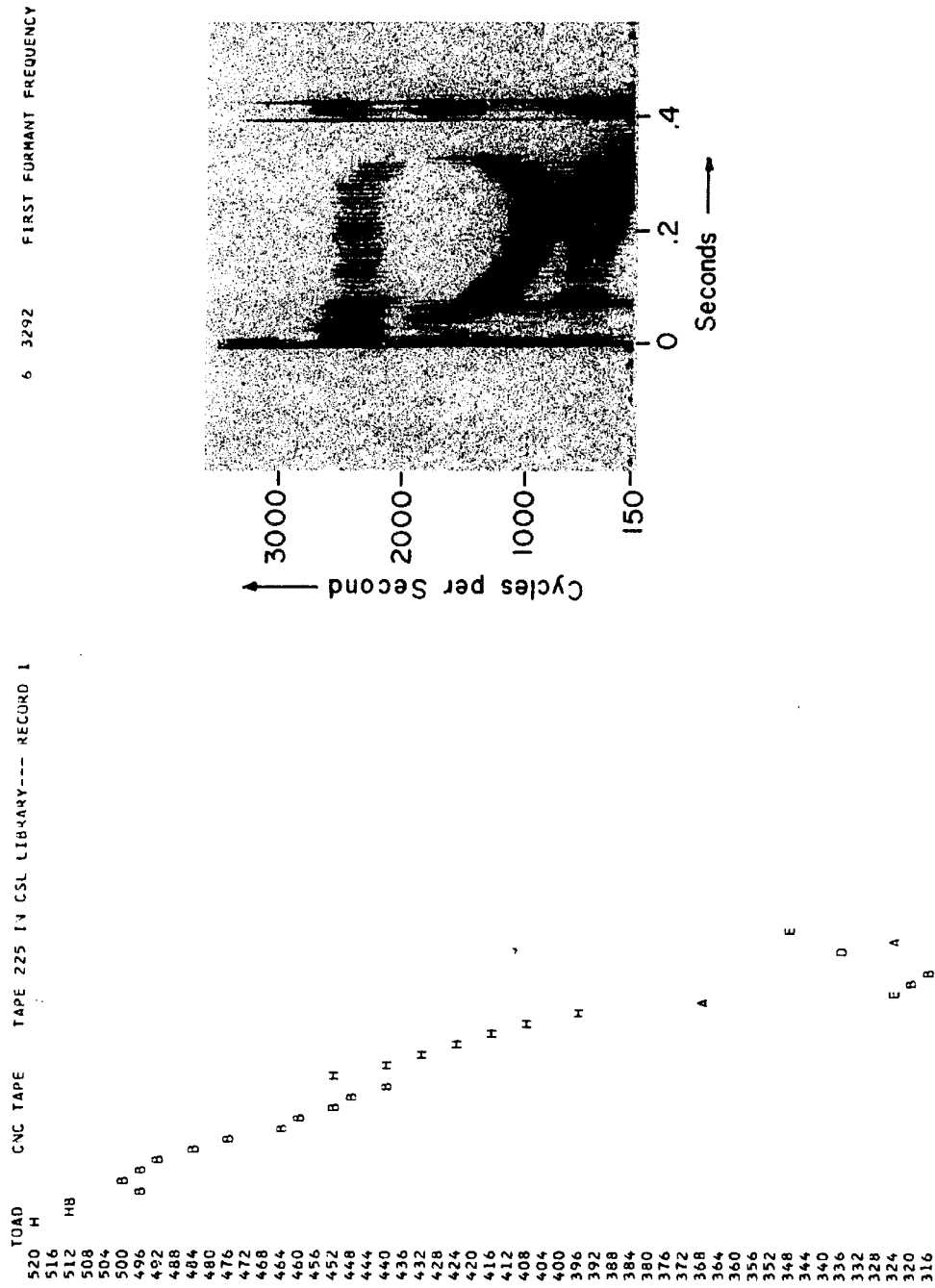
Figure 8-5 shows the spectrogram and first formant estimates for the CNC word jug. The level of the first formant dropped low enough to cause a problem, even for the AVC on the spectrograph, immediately after the peak of the first formant transition, as can be observed on the spectrogram. The listing of the formant had the "tenth" pulse at time 1350 and the "eleventh" pulse at time 1639. Thus, the formant tracker missed several periods at this point. (The tenth pulse is at the top of the graph drawn by the computer.) This occasional pulse omission is a recurring event and occurs for many reasons. It does not seem to be a major problem if formant continuity is assumed. The omission would, of course, be emphasized if a linear time scale were used for drawing the computer graphs; since a gap would be present in the plotted values.

The spectrogram and analysis of the CNC word toad is shown in Figure 8-6. The scale on the graph drawn by the computer is 4 cps per line.





Graph and Spectrogram for the Word jug

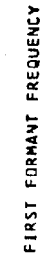
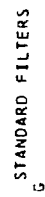


Graph and Spectrogram for the Word Toad

Figure 8-6

4. Analysis of Speech Recorded in Normal Laboratory Noise

The spectrogram and analysis presented in Figure 8-7 are for the word well recorded in the presence of conventional laboratory blowers. Enough of the spectrogram has been presented in Figure 8-7 to show the background noise. It will be noted that small perturbations of the spectrum do not create large errors with this type of formant tracking system. This would be expected since only the spectral areas in the region of local maxima are of interest, and these areas are relatively unchanged by low level background noise. Background noise has a stronger effect, of course, upon spectral areas in the region of local minima. One advantage of the system described here, however, is that the formant frequency measure is relatively insensitive to the magnitude and perturbation of the spectral minima.



Chapter 9

Summary and Discussion

This investigation has been an inquiry into the nature of vowel formants in speech and into some of the problems involved in their measurement. Although vowel formants have been a subject of research for many years, it is hoped that the present study represents a contribution to the description of their basic properties. The application of various types of glottal input waveforms to formant generating networks has been considered in some detail in the study. The various components of the output of a formant generating network resulting from various types of inputs have been represented analytically.

Expressions are derived for the output of a resonator when the input is a recurrent sequence of exponential pulses and when the input is a recurrent sequence of critically damped pulses. Interval representations are derived for the output associated with successive pulses at the input. The output of the resonator resulting from an input of periodically recurring exponentially damped pulses was described by a periodic term and a transient term. When the input is a periodic series of critically damped pulses, the output on any particular interval consists of an exponentially damped pulse, a critically damped pulse, and a damped sinusoid. As the number of input pulses increases the limiting steady-state condition is approached. The behavior of the output of the

resonator in the limiting case is also described in terms of an interval representation.

The problem of deriving the formants from the speech wave has also been considered in detail. Numerous approaches to formant analysis have been investigated in the past, and these are reviewed in this study. From an examination of the previous work, the author concluded that a combination of analog and digital processing would be effective for formant analysis. Analog filters are employed for the initial formant separation, and thereafter digital techniques are used to accomplish the formant frequency estimation. This procedure made it possible to utilize the convenience and speed of analog filters, as well as the power of a digital computer to perform logical operations on the output waveforms of the filters.

A method of deriving the frequency of the first formant was implemented in the experimental part of the investigation. The method has two main distinctions which separate it from all previous formant tracking schemes:

- a) wide bandwidth overlapping filters are used redundantly, and
- b) an envelope characteristic which corresponds to the pulsed time domain realization of a pair of conjugate poles is used for determining period boundaries.

In a study recently reported by Crystal [13], logical operations were performed on consecutive intervals in the output of a simulated basilar membrane model. The current investigation differs from that study in several ways, including the use of pitch synchronous techniques to determine the intervals upon which the logical operations are to be

performed.

In the formant tracker [5, 6] developed at Melpar, pitch synchronous tracking of the first formant is employed. In that system repeated differentiation is used to determine the period boundaries, after which the formant frequency is determined by axis crossing rates. The pitch detection circuitry used in the formant tracker discussed in this present study is different in that the formant envelope, not the discontinuity, is used for period determination. Thus, the Melpar formant tracker should determine more exactly the real period boundaries of the fundamental laryngeal excitation. While detailed investigations of combined analog and digital techniques for deriving successive periods of the fundamental voice frequency are in progress in the Communication Sciences Laboratory, the period boundary determination employed in the present study seemed entirely adequate to investigate the effectiveness of the proposed method of formant frequency measurement. It is the opinion of the author that better formant isolation can be achieved by the fixed filter technique described in this report than by positioning a single filter by counting axis crossing rates. If a formant isolating filter is positioned according to previous measurements, then an error in measurement may considerably affect the formant isolation. It must be recognized, however, that when a bank of fixed filters is employed, more digital computer processing is required than for a single (variable) filter.

In the calibration tests made on the formant tracker used in the study, no attempt was made to determine the accuracy of the system when the second formant is close to the first. This would require a

study of the interactions among the first and second formant frequencies, amplitudes, and bandwidths as observed in the filter outputs. The results would depend considerably upon the particular set of filters used. It is anticipated that a different set of filters will be used to implement a more complete formant-tracker based on this present research, and formant interaction will be investigated in that system. The following results indicate, however, that the present system has considerable ability to isolate first and second formants. The first formant in the /u/ which was used in testing the system showed no indication of second formant contamination. When only a single filter for the range 600-900 cps was used there was wild variation in the estimation of the first formant frequency of /a/. After two cascaded filters (each with slopes of 12 db/octave) were used to isolate the range of 600-900 cps, the variance in the first formant frequency of /a/ dropped to 311 ($\sigma \approx 17$ cps) with a mean formant frequency of 725 cps. In the above cases the first formant frequency estimate was in good agreement with human measurements of the corresponding spectrogram. The variation is expected for the first formant frequency since two filters do not provide the same degree of isolation that are achieved by the standard Bell Telephone Laboratories wide-band spectrograph filter. Thus, in future models of the formant tracker, even sharper filter slopes will be used. As would be expected from the formant locations, /o/ is one of the worst vowels for analysis with this system; (recall the filter slopes are 12 db/octave for the range 450-750 cps). Again, a far superior filter bank would be needed in a real system.

The current system used a multiplexer with a 50 Kc rate. This rate, which also was limited by the A/D converter and the tape unit, would have to be increased considerably to track all formants effectively. In addition, the tape unit only allowed six bits/sample of accuracy at the current operating speeds. The six bits/sample could be tolerated if very high quality equalization circuits were used in connection with good AGC (automatic gain control) circuits. It is more likely, however, that more bits/sample, high quality equalization, and AGC would all be needed for a highly reliable formant tracker.

Smoothing functions would need to be used on the output for certain applications. The particular smoothing functions to be used would, however, be strongly dependent upon the purpose of the measurements. None of the data presented in this report were smoothed by any such functions. It would also be expected that the normal output would be expressed as a function of time rather than by a period number. Since the time associated with each period is carried along in the output, such a presentation would only require a trivial change.

This study has shown the feasibility of an automatic pitch synchronous formant tracking scheme using fixed, wide bandwidth overlapping filters. Highly accurate formant measurements are possible with the system. While there were several practical constraints on the system developed, it appears that these can be surmounted in a more general system. Such a system should provide repeated and precise measurements of the first few formants under most speaking conditions of interest in speech analysis. One major interest of the author is to implement a full formant tracking system based upon the techniques described in this study.

Appendix A

A Simple Resonator

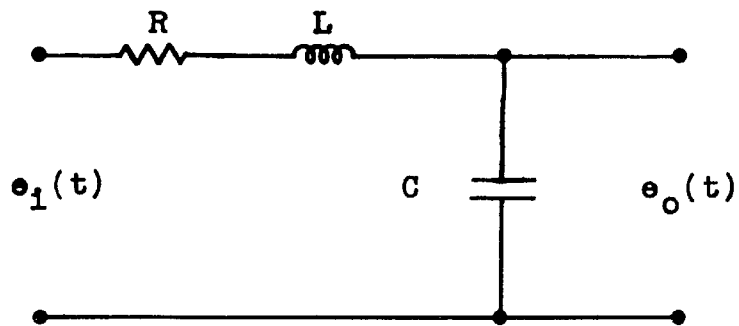
The purpose of this appendix is to summarize previously known results about the standard model of a single speech formant. In addition, the response of the model is determined for single pulses as well as for recurrent sequences of pulses. The results for the recurrent sequences of pulses are presented as infinite sums: this is necessary since there are no further restrictions placed upon the nature of the recurrence.

1. Frequencies Associated with a Simple Resonator

The voltage transfer function of the simple electrical resonant circuit shown in Figure A-1 corresponds to a pair of conjugate poles in the volume velocity transfer function of the vocal tract: that is, to a single formant.

The remainder of this section is devoted to a brief discussion of the resonator shown in Figure A-1, as well as to a derivation of the simple facts regarding its frequency characteristics. The Laplace transformation of the voltage transfer function of the resonator is given by

$$H(s) = \frac{\frac{1}{LC}}{s^2 + \frac{R}{L}s + \frac{1}{LC}} . \quad A.1$$



Simple Resonant Circuit

Figure A-1

Letting

$$k = \frac{R}{2L} \quad A.2$$

$$x = \frac{1}{LC} \quad A.3$$

Equation A.1 can be rewritten as:

$$H(s) = \frac{x}{s^2 + 2ks + x} \quad A.4$$

If it is the case that

$$x \geq k^2, \quad A.5$$

then $H(s)$ has poles given by:

$$s = -k \pm j[x - k^2]^{1/2}. \quad A.6$$

Inequality A.5 is assumed, since Figure 3-3 shows that this inequality is valid for vowels. The following equation defines v :

$$v = [x - k^2]^{1/2}. \quad A.7$$

Thus, the poles of $H(s)$ can be written as

$$s = -k \pm j v. \quad A.8$$

Letting $s = j\omega$ in Equation A.4 and then using the standard techniques for determining the absolute value of a complex function, the amplitude density spectrum of $H(s)$ on the imaginary axis is found to be:

$$\begin{aligned} |H(j\omega)| &= \left[\frac{x^2}{[x + 2jk\omega - \omega^2][x - 2jk\omega - \omega^2]} \right]^{1/2} \\ &= \frac{x}{[(x - \omega^2)^2 + 4k^2\omega^2]^{1/2}}. \end{aligned} \quad A.9$$

By taking the derivative with respect to ω and equating it to zero, the roots of the equation for the derivative are found to be:

$$\omega = 0 \quad \text{A.10}$$

$$\omega^2 = x - 2k^2 \quad \text{A.11}$$

It can be shown that in the case where

$$x \leq 2k^2 \quad \text{A.12}$$

the spectral amplitude density maximum occurs at

$$\omega_{\max} = 0, \quad \text{A.13}$$

whereas in the case where

$$x > 2k^2 \quad \text{A.14}$$

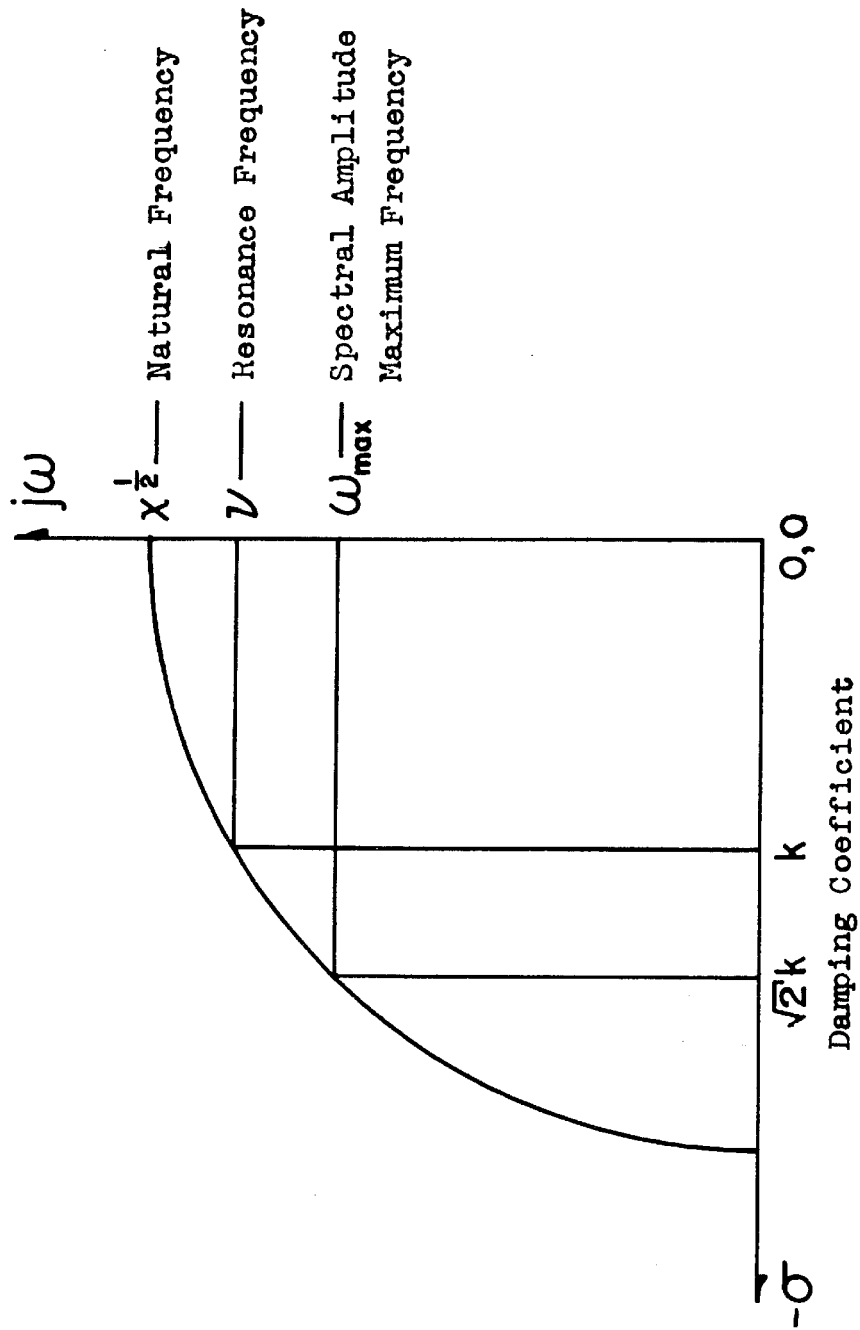
the spectral amplitude density maximum occurs at

$$\begin{aligned} \omega_{\max} &= \pm [x - 2k^2]^{1/2} \\ &= \pm [v^2 - k^2]^{1/2}, \end{aligned} \quad \text{A.15}$$

and the root $\omega=0$ corresponds to a local minimum.

The symbols of the above equation are generally identified as follows: $x^{1/2}$ is called the natural frequency, v is called the resonance frequency (and in the case of speech is defined as the formant frequency), $\omega_{\max}$ is called the spectral amplitude density maximum, and k is called the damping coefficient.

The relationships between the natural frequency, $x^{1/2}$, the resonance frequency, v , the spectral amplitude density maximum frequency, $\omega_{\max}$, and the damping coefficient, k , are shown in Figure A-2. It is clear,



Geometric Representation of the Relationships Between the Frequencies of a Simple Resonator When $X > 2k^2$

Figure A-2

both from the figure and the associated equations that

$$\omega_{\max} \leq v \leq x^{1/2}. \quad \text{A.16}$$

It is an easy matter to determine the amplitude at the spectral amplitude density maximum. One needs only to evaluate Equation A.9 at the frequency given by A.15:

$$\begin{aligned} |H(j\omega_{\max})| &= \frac{x}{[4k^4 + 4k^2x - 8k^4]^{1/2}} \\ &= \frac{x}{2kv}. \end{aligned} \quad \text{A.17}$$

A graph of the amplitude density spectrum of the transfer function of the resonator has been given in Figure A-3 for three cases: the case where $v = k$, the case where $v = 2k$, and the case where $v = 10k$. The figure shows that if the frequency v is held constant and the damping factor k is decreased, then the maximum value of the amplitude density spectrum increases. The case for $v = k$ is the point at which the amplitude maximum changes from

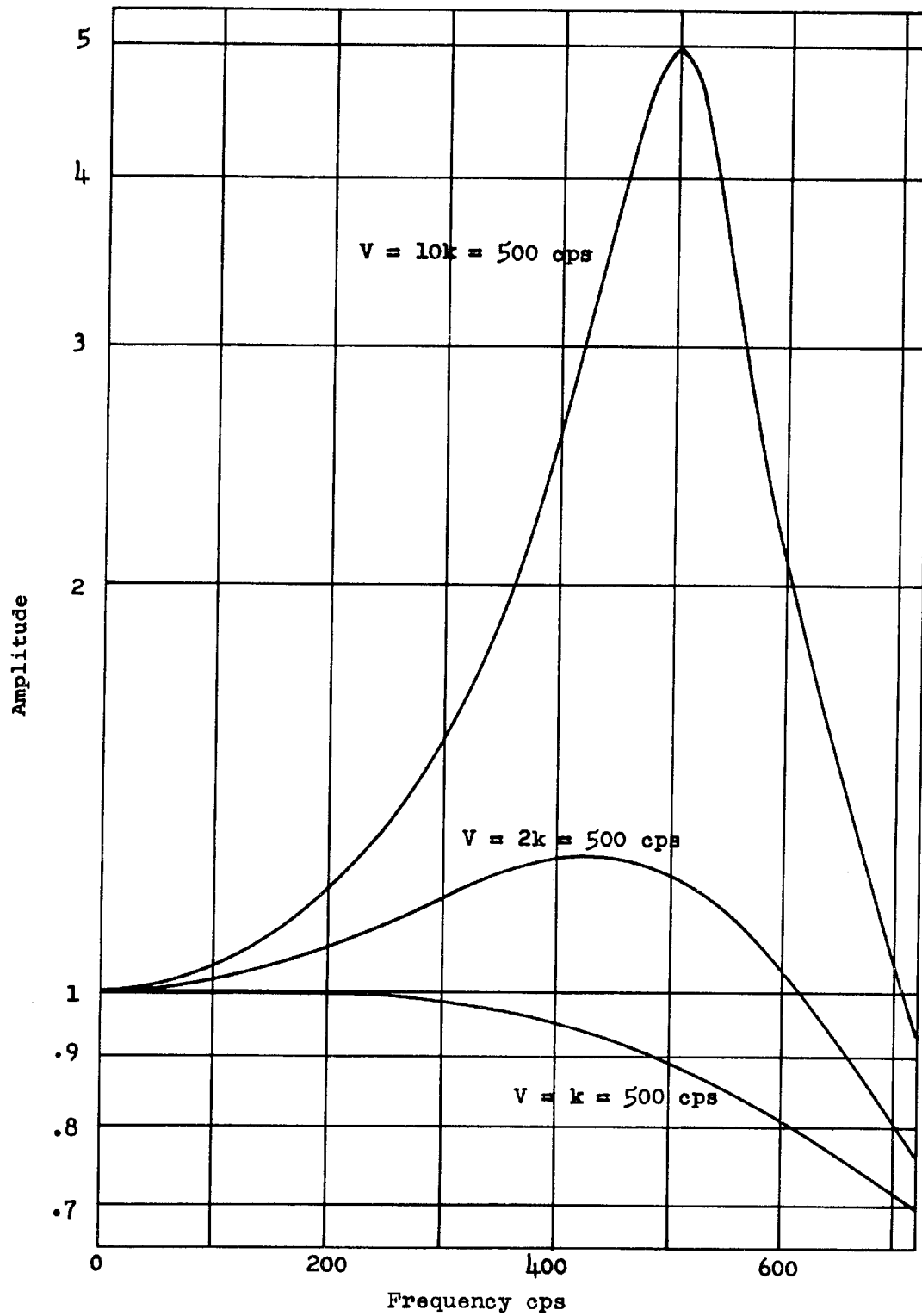
$$\begin{aligned} \omega_{\max} &= [x - 2k^2]^{1/2} \\ &= [v^2 - k^2]^{1/2} \end{aligned} \quad \text{A.18}$$

to

$$\omega_{\max} = 0.$$

This transition can be seen in the graph which illustrates the case $v = k$.

The inequality



Resonator Amplitude Response for
Fixed Resonance Frequency and Three Different
Conditions of Damping

Figure A-3

$$| H(j\omega) | \geq 1 \quad \text{A.19}$$

is satisfied for some values of ω . Thus, if a sinusoid of frequency ω is passed through the filter, and if ω is such that Inequality A.20 is satisfied, then the output sinusoid will have a larger amplitude than the input sinusoid. Accordingly, for sinusoids in this frequency range the filter acts as a voltage amplifier.

Equating Equation A.9 to unity and solving for ω , one finds that Inequality A.19 is satisfied when ω meets the conditions of the following inequalities:

$$0 \leq \omega \leq \sqrt{2} \omega_{\max}. \quad \text{A.20}$$

The above equations show that the resonator acts as a voltage amplifier for all frequencies greater than d.c. and less than $\sqrt{2} \omega_{\max}$.

It has been shown in the literature that k , the damping coefficient, is $1/2$ the bandwidth in radians per second if one either

- a) observes the output across the resistor, or
- b) allows a 6 db/octave rise as an approximation to the radiation resistance at the output of the vocal tract.

If B is the bandwidth in radians per second, then:

$$2k = B. \quad \text{A.21}$$

In summary, we note that there are 4 basic frequencies associated with the simple resonator investigated:

- (1) the natural frequency, $x^{1/2}$,
- (2) the resonant frequency, v ,

- (3) the spectral amplitude density maximum frequency, $\omega_{\max}$, and
- (4) the damping coefficient, k .

2. Response of a Simple Resonator to Three Different Types of Pulses

A pulse which has often been used in speech synthesis is an exponential pulse: here specified as $e(t)$. This pulse can be formed as the output of a circuit with a single pole. Its time domain representation is

$$e(t) = U(t)e^{-\lambda t} \quad \text{A.22}$$

where λ is a constant and $U(t)$ is the unit step function.

A second pulse which is commonly used in speech synthesis is a critically damped pulse: denoted $c(t)$. This pulse can be formed by a critically damped resonator and has a frequency domain characterization given by a double real pole. Its time domain representation is given by

$$c(t) = U(t)te^{-\lambda t}, \quad \text{A.23}$$

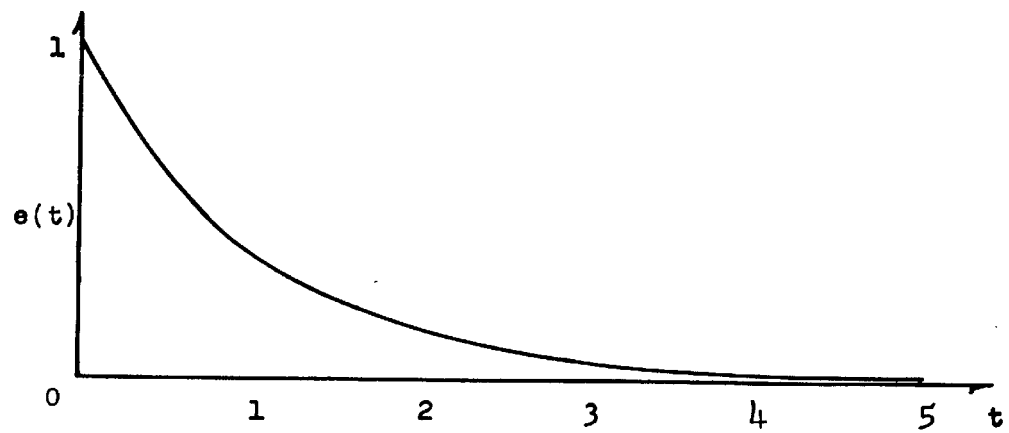
where $U(t)$ is the Heavyside unit step function, e is the base of the naperian logarithms and λ is a constant.

A third pulse, which is of interest for other reasons, is the rectangular pulse, $r(t)$. In the time domain, it is characterized by

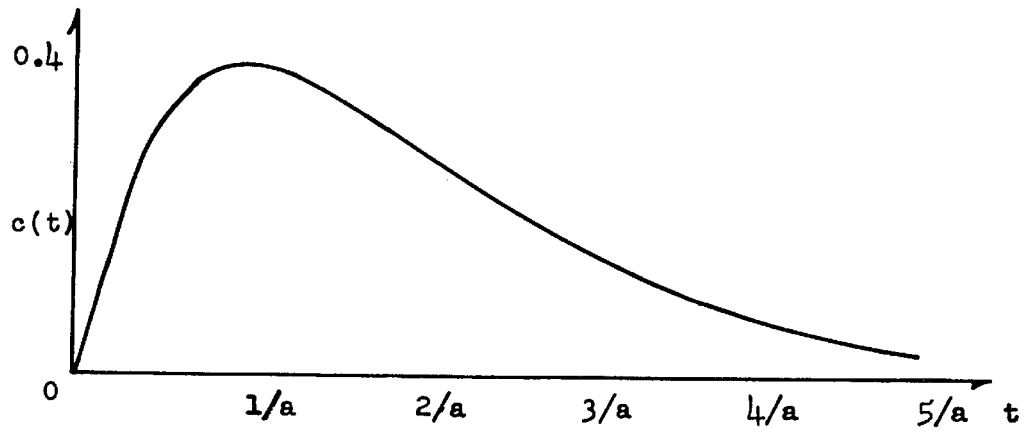
$$r(t) = U(t)U(W-t) \quad \text{A.24}$$

where W is the duration of the pulse.

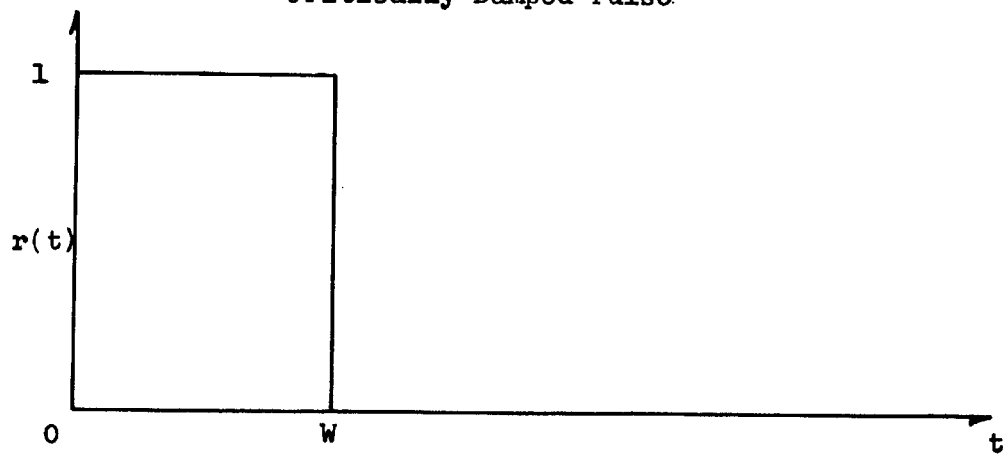
The time domain representation of these pulses are shown in Figure A-4.



Exponential Pulse



Critically Damped Pulse



Rectangular Pulse

Basic Input Pulses Investigated

Figure A-4

Since a simple (ideal) resonator is a linear device, there is no need to add an arbitrary amplitude coefficient to each of the pulses described above. The output corresponding to one pulse with amplitude coefficient A is simply A times the output corresponding to the pulse.

The Laplace transforms of the above pulses are:

$$\mathcal{L}[e(t)] = \frac{1}{s + \lambda}, \quad \text{A.25}$$

$$\mathcal{L}[c(t)] = \frac{1}{[s + \lambda]^2}, \quad \text{A.26}$$

$$\mathcal{L}[r(t)] = \frac{1}{s} [1 - e^{-sW}]. \quad \text{A.27}$$

The Laplace transformation of the voltage output of the simple resonator corresponding to a particular input is given by the product of the Laplace transformation of the voltage transfer function of the resonator and the Laplace transformation of the input. In the following development a subscript is used with the voltages to identify the type of input pulse involved. The Laplace transformation of the voltage output corresponding to an exponential pulse, $\mathcal{L}[e_e(t)]$, is given by:

$$\begin{aligned} \mathcal{L}[e_e(t)] &= H(s) \mathcal{L}[e(t)] \\ &= \frac{1}{[s + \lambda]} \frac{x}{[s^2 + 2ks + x]} \\ &= \frac{x}{[\lambda^2 + x - 2k\lambda]} \left[\frac{1}{[s + \lambda]} - \frac{[s + k] + [k - \lambda]}{[s + k]^2 + v^2} \right]. \quad \text{A.28} \end{aligned}$$

According to the essential uniqueness of the transformation, the output of a simple filter corresponding to a single exponential input pulse is given by:

$$e_e(t) = \frac{xU(t)}{[(\lambda-k)^2 + v^2]} \left[e^{-\lambda t} - e^{-kt} \left[\cos vt + \frac{k-\lambda}{v} \sin vt \right] \right]. \quad A.29$$

Using a simple trigonometric identity, the above equation can be rewritten as:

$$e_e(t) = U(t)[E_1 e^{-\lambda t} - E_2 e^{-kt} \sin(vt + \varphi_e)], \quad A.30$$

where

$$E_1 = \frac{x}{[(\lambda - k)^2 + v^2]} \quad A.31$$

$$E_2 = \frac{x}{v[v^2 + (\lambda - k)^2]^{1/2}} \quad A.32$$

$$\varphi_e = \tan^{-1} \left[\frac{v}{k - \lambda} \right]. \quad A.33$$

The Laplace transform of the voltage output corresponding to a single critically damped pulse is given by:

$$\begin{aligned} \mathcal{L}[e_e(t)] &= H(s) \mathcal{L}[c(t)] \\ &= \frac{1}{[s + \lambda]^2} \frac{x}{[s^2 + 2ks + x]} \\ &= \frac{2x[\lambda - k]}{[\lambda^2 + x - 2k\lambda]^2} \frac{1}{[s + \lambda]} + \frac{x}{[\lambda^2 + x - 2k\lambda]} \frac{1}{[s + \lambda]^2} \\ &\quad + x \left[\frac{2[k - \lambda][s + k] + [\lambda - k]^2 - v^2}{[\lambda^2 + x - 2k\lambda]^2 [s^2 + 2ks + x]} \right]. \quad A.34 \end{aligned}$$

The time domain representation of $e_c(t)$ is thus:

$$e_c(t) = \frac{2x[\lambda - k] U(t)}{[(\lambda - k)^2 + v^2]^2} e^{-\lambda t} + \frac{x U(t)}{[(\lambda - k)^2 + v^2]} t e^{-\lambda t} \\ + \frac{x e^{-kt} U(t)}{[(\lambda - k)^2 + v^2]^2} \left[2(k-\lambda) \cos vt + \frac{[(\lambda-k)^2 - v^2]}{v} \sin vt \right]. \quad A.35$$

This time domain representation of the output of the resonator corresponding to a single critically damped input pulse can be rewritten as:

$$e_c(t) = U(t) \left[C_1 e^{-\lambda t} + C_2 t e^{-\lambda t} + C_3 e^{-kt} \sin(vt + \varphi_c) \right], \quad A.36$$

where

$$C_1 = \frac{2x[\lambda - k]}{[(\lambda - k)^2 + v^2]^2}, \quad A.37$$

$$C_2 = \frac{x}{[(\lambda - k)^2 + v^2]}, \quad A.38$$

$$C_3 = \frac{x}{v[(\lambda - k)^2 + v^2]^2}, \quad A.39$$

$$\varphi_c = \tan^{-1} \left[\frac{2v[k - \lambda]}{[(\lambda - k)^2 - v^2]} \right]. \quad A.40$$

The Laplace transform of the voltage output corresponding to a single rectangular input pulse is given by:

$$\mathcal{L}[e_r(t)] = H(s) \mathcal{L}[r(t)]$$

$$\begin{aligned}
\mathcal{L}[e_r(t)] &= \left[\frac{x}{[(s+k)^2 + v^2]} \quad \frac{1}{s} \right] [1 - e^{-sW}] \\
&= \left[\frac{-[s+k] - k}{[(s+k)^2 + v^2]} + \frac{1}{s} \right] [1 - e^{-sW}].
\end{aligned} \tag{A.41}$$

The time domain representation of $r(t)$ is given by:

$$\begin{aligned}
e_r(t) &= U(t) \left[1 - e^{-kt} \left\{ \cos vt + \frac{k}{v} \sin vt \right\} \right] \\
&\quad - U(t-W) \left[1 - e^{-k(t-W)} \left\{ \cos[v(t-W)] + \frac{k}{v} \sin[v(t-W)] \right\} \right].
\end{aligned} \tag{A.42}$$

From the above, it follows that the output of the simple resonator corresponding to a single rectangular input pulse can be written:

$$\begin{aligned}
e_r(t) &= U(t) \left[1 - R_1 e^{-kt} \sin[vt + \varphi_r] \right] \\
&\quad - U(t-W) \left[1 - R_1 e^{-k(t-W)} \sin[v(t-W) + \varphi_r] \right],
\end{aligned} \tag{A.43}$$

where

$$R_1 = \left[1 - \left(\frac{k}{v} \right)^2 \right]^{1/2}, \tag{A.44}$$

$$\varphi_r = \tan^{-1} \frac{v}{k} \tag{A.45}$$

Expressions have now been derived which represent the output of a simple resonator as a function of time for each of the three input pulses of interest. The next section is devoted to a representation of the output corresponding to a recurrent input train of each of these

three different types of pulses.

3. Response of a Simple Resonator to a Recurrent Sequence of Pulses

The pulses in a recurrent series of pulses will be considered to occur at times t_1 . Figure A-5 shows three different recurrent series of pulses, all of which have the same pulse initiation times. Capital letters will be used to designate pulse trains which correspond to the recurrence of the pulses designated above with small letters, that is

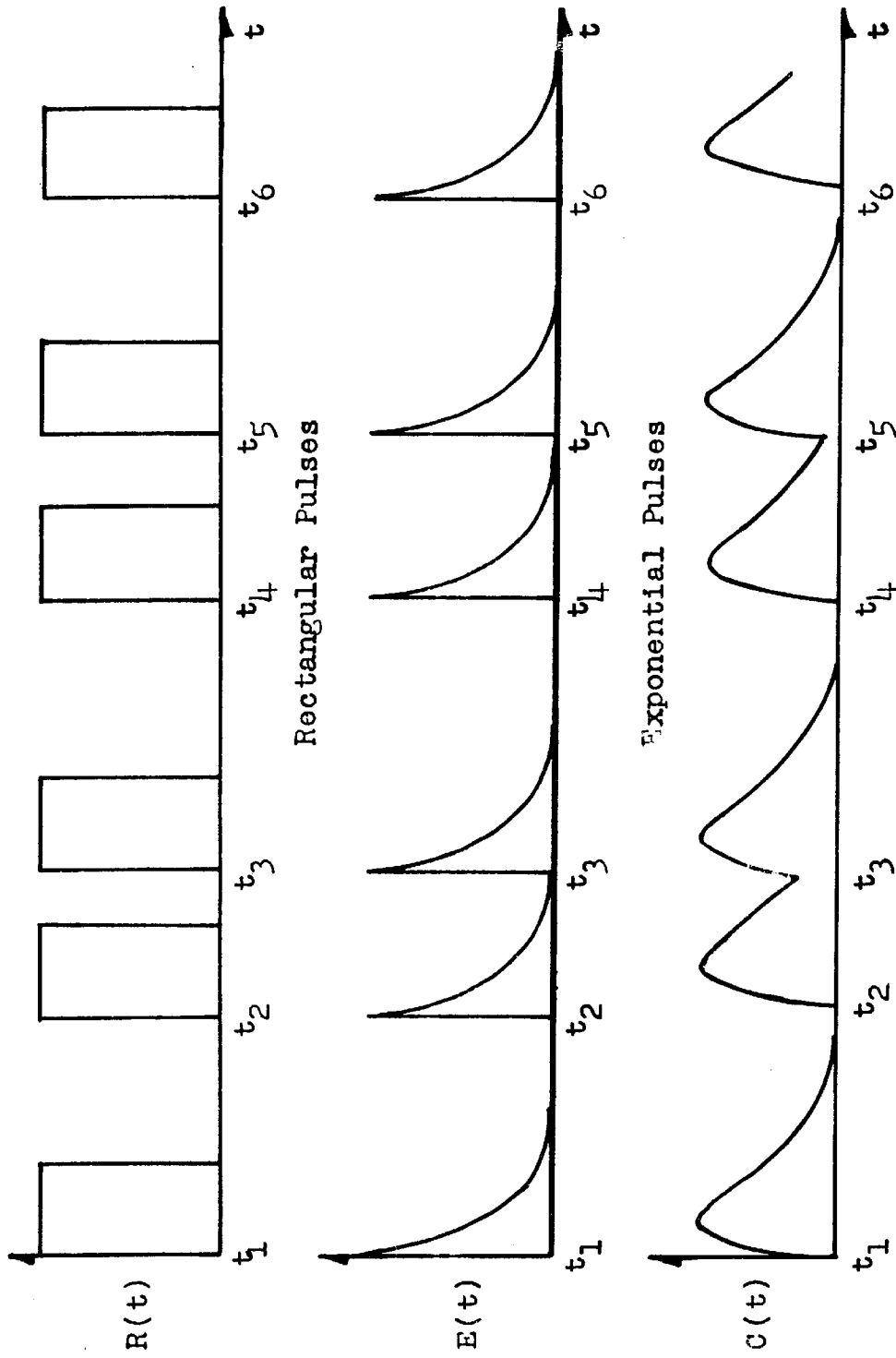
$$\begin{aligned} E(t) &= \sum_{n=0}^{\infty} U(t-t_1)e(t-t_1) \\ &= \sum_{n=0}^{\infty} U(t-t_1)e^{-\lambda(t-t_1)} \end{aligned} \quad A.46$$

is a recurring sequence of exponential pulses;

$$\begin{aligned} C(t) &= \sum_{n=0}^{\infty} U(t-t_1)c(t-t_1) \\ &= \sum_{n=0}^{\infty} U(t-t_1)[t-t_1]e^{-\lambda(t-t_1)} \end{aligned} \quad A.47$$

is a recurring sequence of critically damped pulses; and

$$R(t) = \sum_{n=0}^{\infty} U(t-t_1)r(t-t_1)$$



Critically Damped Pulses
Recurrent Sequences of Pulses
Figure A-5

$$R(t) = \sum_{n=0}^{\infty} [U(t-t_1) - U(t-t_1-W)] \quad \text{A.48}$$

is a recurring sequence of rectangular pulses.

By knowing the output corresponding to the individual pulses and by using the linearity and time invariant properties of the simple resonator, one can easily obtain a representation for the output of the resonator corresponding to each of the above recurrent sequences of pulses. Each output voltage will again be subscripted to denote its corresponding input.

The output of the simple resonator resulting from each of the recurrent inputs is given by:

$$\begin{aligned} e_E(t) &= \sum_{i=0}^{\infty} e_e(t-t_1) \\ &= \sum_{i=0}^{\infty} U(t-t_1) \left[E_1 e^{-\lambda(t-t_1)} - E_2 e^{-k(t-t_1)} \sin[v(t-t_1)+\varphi_e] \right]; \end{aligned} \quad \text{A.49}$$

$$\begin{aligned} e_C(t) &= \sum_{i=0}^{\infty} e_c(t-t_1) \\ &= \sum_{i=0}^{\infty} U(t-t_1) \left[C_1 e^{-\lambda(t-t_1)} + C_2(t-t_1) e^{-\lambda(t-t_1)} \right. \\ &\quad \left. + C_3 e^{-k(t-t_1)} \sin[v(t-t_1)+\varphi_c] \right]; \end{aligned} \quad \text{A.50}$$

$$\begin{aligned}
e_R(t) &= \sum_{i=0}^{\infty} U(t-t_i) r(t-t_i) \\
&= \sum_{i=0}^{\infty} U(t-t_i) \left[1-R_1 e^{-k(t-t_i)} \sin[v(t-t_i)+\varphi_r] \right] \\
&\quad - \sum_{i=0}^{\infty} U(t-t_i-W) \left[1-R_1 e^{-k(t-t_i-W)} \sin[v(t-t_i-W)+\varphi_r] \right]. \quad A.51
\end{aligned}$$

Expressions have now been exhibited which correspond exactly to the output of the simple filter for the three different types of recurrent input pulses. Three sums play a major role in each of the above representations, namely the sums:

$$C(t) = \sum_{i=0}^{\infty} U(t-t_i) [t-t_i] e^{-\lambda(t-t_i)}, \quad A.52$$

$$E(t) = \sum_{i=0}^{\infty} U(t-t_i) e^{-\lambda(t-t_i)}, \quad A.53$$

and a sum which shall be denoted by $S_{\varphi}(t)$:

$$S_{\varphi}(t) = \sum_{i=0}^{\infty} U(t-t_i) e^{-k(t-t_i)} \sin[v(t-t_i)+\varphi_1]. \quad A.54$$

It is possible to obtain closed form representations for the above sums by making a further restriction on the recurring sequences. If the sequences have the additional property that for all $i \geq 0$

$$t_i - t_{i+1} = T$$

A.55

where T is a constant, then the sums can be simplified. When Equation A.55 is satisfied the sequences will be called periodic, even though they are not periodic in the strict mathematical sense. Various coefficients for the periodic case are determined in Appendix B.

Appendix B

Mathematical Representations of Periodically Recurring Pulses

The purpose of this appendix is to obtain several representations for the following functions:

$$0 \leq t < \infty \quad E(t) = \sum_{n=0}^{\infty} U(t-nT)e^{-\lambda(t-nT)}, \quad B.1$$

$$0 \leq t < \infty \quad C(t) = \sum_{n=0}^{\infty} U(t-nT)(t-nT)e^{-\lambda(t-nT)}, \quad B.2$$

$$0 \leq t < \infty \quad S_{\varphi}(t) = \sum_{n=0}^{\infty} U(t-nT)e^{-k(t-nT)}\sin[v(t-nT)+\varphi]. \quad B.3$$

The function, $E(t)$, is a periodically recurring exponential pulse, the function, $C(t)$, is a periodically recurring critically damped pulse, and the function $S_{\varphi}(t)$, is a periodically recurring damped sinusoid. The various representations for these pulses are needed for the discussion in the body of the text. Since the representations do not seem to exist in the literature, they are included for reference in this paper. Since they are long, they have been relegated to this appendix.

1. Interval Representations

For the remainder of this section it is assumed that the variable t

satisfies the following inequality:

$$[m-1]T \leq t < mT. \quad B.4$$

The interval from $[m-1]T$ to mT will be called the m^{th} period. Thus, this section will be devoted to deriving the representations of the three functions for the m^{th} period.

a. Periodically recurring exponential pulses

Assuming that t is in the m^{th} period, Equation B.1 can be rewritten as

$$E(t) = \sum_{n=0}^{m-1} e^{-\lambda(t-nT)} \quad B.5$$

where

$$[m-1]T \leq t < mT.$$

Since

$$\sum_{n=0}^{m-1} z^n = \frac{[1 - z^m]}{[1 - z]} \quad B.6$$

Equation B.5 may be rewritten as

$$E(t) = \frac{[1 - e^{\lambda mT}]}{[1 - e^{\lambda T}]} e^{-\lambda t}. \quad B.7$$

The above equation shows that a sum of exponential pulses can be exactly represented during the m^{th} period by a single exponential pulse with the same decay as the original exponential pulse. The amplitude coefficient for the pulse is determined for the representation on the m^{th} period.

Equation B.7 expresses the coefficients of the pulses as a function of the period.

The next step is to transform the equation so that one can make comparisons of the function according to period number. Recalling that:

$$t = \tau + [m-1]T, \quad \text{B.8}$$

Equation B.7 may be rewritten as:

$$\begin{aligned} E(\tau+[m-1]T) &= \frac{[1 - e^{\lambda mT}]}{[1 - e^{\lambda T}]} e^{-\lambda(\tau+[m-1]T)} \\ &= \frac{[1 - e^{-\lambda mT}]}{[1 - e^{-\lambda T}]} e^{-\lambda\tau} \\ &= \xi_m e^{-\lambda\tau}, \end{aligned} \quad \text{B.9}$$

where

$$\xi_m = \frac{1 - e^{-\lambda mT}}{1 - e^{-\lambda T}}. \quad \text{B.10}$$

The above type of representation will be useful when we consider the behavior of the function according to the period number.

b. Periodically recurring critically damped pulses

For values of t defined by Equation B.4, it is possible to rewrite Equation B.2 as follows:

$$\begin{aligned} C(t) &= \sum_{n=0}^{m-1} (t-nT) e^{-\lambda(t-nT)} \\ &= \left[t e^{-\lambda t} \sum_{n=0}^{m-1} e^{\lambda nT} \right] - \left[T e^{-\lambda t} \sum_{n=0}^{m-1} n e^{\lambda nT} \right]. \end{aligned} \quad \text{B.11}$$

The first term in the above sum can be simplified by the use of Equation B.6, and the second term can be simplified by use of the following:

$$\begin{aligned}
 \sum_{n=0}^{m-1} nz^n &= z \frac{d}{dz} \sum_{n=0}^{m-1} z^n \\
 &= z \frac{d}{dz} \frac{[1 - z^m]}{[1 - z]} \\
 &= \frac{z[1 - mz^{m-1} + [m-1]z^m]}{[1 - z]^2}.
 \end{aligned} \tag{B.12}$$

Letting

$$z = e^{\lambda T}$$

Equation B.11 can be rewritten as:

$$\begin{aligned}
 C(t) &= \frac{[1 - e^{\lambda m T}]}{[1 - e^{\lambda T}]} te^{-\lambda t} \\
 &\quad - \frac{Te^{\lambda T}[1 - me^{\lambda [m-1]T} + [m-1]e^{\lambda m T}]}{[1 - e^{\lambda T}]^2} e^{-\lambda t}.
 \end{aligned} \tag{B.13}$$

The above equation shows that a periodically recurring sum of critically damped pulses can be exactly represented on any period as a sum of a critically damped and an exponential pulse. In addition the coefficients can be evaluated as functions of λ , T , and m . The time shift defined by Equation B.8 will now be made so that we may later consider the representation as a function of period number. Thus:

$$\begin{aligned}
C(\tau+[m-1]T) &= \frac{[1 - e^{\lambda mT}]}{[1 - e^{\lambda T}]} (\tau + [m-1]T) e^{-\lambda(\tau+[m-1]T)} \\
&\quad + \frac{-Te^{\lambda T}[1 - me^{\lambda[m-1]T} + [m-1]e^{\lambda mT}]}{[1 - e^{\lambda T}]^2} e^{-\lambda(\tau+[m-1]T)} \\
&= \xi_m \tau e^{-\lambda \tau} + \rho_m e^{-\lambda \tau},
\end{aligned} \tag{B.14}$$

where

$$\rho_m = \frac{Te^{-\lambda T}[1 - me^{-\lambda[m-1]T} + [m-1]e^{-\lambda mT}]}{[1 - e^{-\lambda T}]^2}, \tag{B.15}$$

and where ξ_m is defined by Equation B.10.

c. Periodically recurring damped sinusoids

The sum of two damped sinusoids with the same frequency and damping factor is a damped sinusoid which in turn has the same frequency and damping factor. For any interval of the form given by Equation B.4, the sum, $S_\varphi(t)$, (which becomes a finite sum of damped sinusoids), can be expressed as a single damped sinusoid. That is, when t satisfies Equation B.4 we can write Equation B.3 as:

$$S_\varphi(t) = \sum_{n=0}^{m-1} e^{-k(t-nT)} \sin[v(t-nT)+\varphi]. \tag{B.16}$$

In addition, we know there exist both a coefficient A_m and a phase angle B_m such that

$$S_\varphi(t) = A_m e^{-kt} \sin(vt+\varphi+B_m).$$

We shall now evaluate the coefficients A_m and B_m as functions of v , k , m , and T .

By using the identity

$$\sin z = \frac{e^{jz} - e^{-jz}}{2j} \quad \text{B.17}$$

Equation B.16 can be rewritten as

$$\begin{aligned} S_\varphi(t) &= \sum_{n=0}^{m-1} e^{-kt} e^{knT} \left[\frac{e^{j(vt - vnT + \varphi)} - e^{-j(vt - vnT + \varphi)}}{2j} \right] \\ &= \frac{e^{-kt + jvt + j\varphi}}{2j} \sum_{n=0}^{m-1} e^{(k - jv)nT} - \frac{e^{-kt - jvt - j\varphi}}{2j} \sum_{n=0}^{m-1} e^{(k + jv)nT} \end{aligned} \quad \text{B.18}$$

Recalling that

$$\sum_{n=0}^{m-1} z^n = \frac{[1 - z^m]}{[1 - z]}, \quad \text{B.19}$$

it is possible to eliminate the finite sums which occur in Equation B.18 and express the relation as follows:

$$\begin{aligned} S_\varphi(t) &= \frac{e^{-kt} e^{j(vt + \varphi)} [1 - e^{(k - jv)mT}]}{2j [1 - e^{(k - jv)T}]} - \frac{e^{-kt} e^{-j(vt + \varphi)} [1 - e^{(k + jv)mT}]}{2j [1 - e^{(k + jv)T}]} \\ &= \frac{e^{-kt} e^{j(vt + \varphi)} [1 - e^{(k - jv)mT}] [1 - e^{(k + jv)T}]}{2j [1 - e^{(k + jv)T}] [1 - e^{(k - jv)T}]} \\ &\quad + \frac{-e^{-kt} e^{-j(vt + \varphi)} [1 - e^{(k + jv)mT}] [1 - e^{(k - jv)T}]}{2j [1 - e^{(k + jv)T}] [1 - e^{(k - jv)T}]} \end{aligned} \quad \text{B.20}$$

Multiplying the terms in the numerator and then grouping them according to Equation B.17 allows the numerator to be written as a sum of sinusoidal terms. Using the fact that

$$\cos z = \frac{e^{jz} + e^{-jz}}{2} \quad \text{B.21}$$

allows the denominator of B.20 to be simplified. Thus

$$\begin{aligned} S_{\phi}(t) = & \frac{e^{-k(t-mT)} \left[e^{kT} \sin[v(t-[m-1]T)+\phi] - \sin[v(t-mT)+\phi] \right]}{1 - 2e^{kT} \cos vT + e^{2kT}} \\ & + \frac{-e^{-kt} \left[e^{kT} \sin[v(t+T)+\phi] - \sin[vt+\phi] \right]}{1 - 2e^{kT} \cos vT + e^{2kT}}. \end{aligned} \quad \text{B.22}$$

One may define two variables μ_1 and γ_1 so that for all z :

$$\mu_1 \sin(z - \gamma_1) = e^{kT} \sin(z + vT) - \sin(z), \quad \text{B.23}$$

in which case it may be easily verified that:

$$\mu_1^2 = [1 - 2e^{kT} \cos vT + e^{2kT}] \quad \text{B.24}$$

and

$$\tan \gamma_1 = \frac{\sin vT}{e^{-kT} - \cos vT}. \quad \text{B.25}$$

Using the above, Equation B.22 can be simplified to:

$$S_{\phi}(t) = \frac{e^{-kt}}{\mu_1} \left[e^{kmT} \sin[vt + \phi - \gamma_1 - vmT] - \sin[vt + \phi - \gamma_1] \right]. \quad \text{B.26}$$

It is easily seen that if:

$$\mu_m \sin(z + \gamma_m) = e^{kmT} \sin[z - vmT] - \sin[z], \quad \text{B.27}$$

then

$$\mu_m^2 = [1 - 2e^{kmT} \cos vmT + e^{2kmT}] \quad \text{B.28}$$

and

$$\tan \gamma_m = \frac{\sin vmT}{e^{-kmT} - \cos vmT}. \quad \text{B.29}$$

It should be noted that although Equation B.27 is different in form from Equation B.23, the equation for μ_1 and γ_1 given by Equations B.24 and B.25 are compatible with the values of μ_m and γ_m that are obtained by substituting $m=1$ into Equations B.28 and B.29. Substituting the above results into Equation B.26 gives:

$$S_\varphi(t) = \frac{\mu_m}{\mu_1} e^{-kt} \sin[vt + \varphi - \gamma_1 + \gamma_m]. \quad \text{B.30}$$

Equation B.30 is an equation for the sum of the displaced sinusoids during the m^{th} period; that is, for

$$[m-1]T \leq t < mT, \quad \text{B.31}$$

we have shown that

$$\frac{\mu_m}{\mu_1} e^{-kt} \sin[vt + \varphi - \gamma_1 + \gamma_m] = \sum_{n=0}^{\infty} U(t-nT) e^{-k(t-nT)} \sin(v(t-nT) + \varphi). \quad \text{B.32}$$

The above equation shows the coefficients on the m^{th} period. A more enlightening representation is possible if one lets τ be defined as

$$\tau + [m-1]T = t, \quad \text{B.33}$$

in order to achieve a representation which has the property that $\tau=0$ at the start of the period. Thus:

$$S_{\varphi}(\tau+[m-1]T) = \frac{\mu_m}{\mu_1} e^{-k(\tau+[m-1]T)} \sin[v(\tau+[m-1]T) + \varphi - \gamma_1 + \gamma_m]. \quad \text{B.34}$$

It is convenient to define:

$$\begin{aligned} \alpha_m &= \frac{\mu_m}{\mu_1} e^{-k[m-1]T} \\ &= \frac{[1 - 2e^{-kmT} \cos v m T + e^{-2kmT}]^{1/2}}{[1 - 2e^{-kT} \cos v T + e^{-2kT}]^{1/2}}. \end{aligned} \quad \text{B.35}$$

It should be clear that α_m is the amplitude coefficient of the representation on the m^{th} period with the shifted time base. Defining

$$-\beta_m = v m T + \gamma_m \quad \text{B.36}$$

and recalling that

$$\tan(z+y) = \frac{\tan z + \tan y}{1 - (\tan z)(\tan y)} \quad \text{B.37}$$

the tangent of β_m can be written as follows:

$$-\tan \beta_m = \frac{\tan v m T + \tan \gamma_m}{1 - (\tan v m T)(\tan \gamma_m)}$$

$$\begin{aligned}
-\tan \beta_m &= \frac{\frac{\sin v m T}{\cos v m T} + \frac{\sin v m T}{e^{-k m T} - \cos v m T}}{1 - \frac{\sin^2 v m T}{\cos v m T [e^{-k m T} - \cos v m T]}} \\
&= \frac{e^{-k m T} \sin v m T}{e^{-k m T} \cos v m T - 1} \\
&= \frac{-\sin v m T}{e^{k m T} - \cos v m T} .
\end{aligned} \tag{B.38}$$

It then follows that:

$$\tan \beta_m = \frac{\sin v m T}{e^{k m T} - \cos v m T} . \tag{B.39}$$

Noticing that

$$\beta_1 = -\gamma_1 - v T \tag{B.40}$$

Equation B.34 can be rewritten as follows

$$S_\varphi(\tau + [m-1]T) = \alpha_m e^{-k\tau} \sin[v\tau + \varphi + \beta_1 - \beta_m] . \tag{B.41}$$

This latter representation is particularly useful when comparing the form of the sum on different periods.

2. Interval Steady-State Representation

There are simple arguments which show heuristically that one might expect the three functions which have been considered to have an interval steady-state representation. We shall simply investigate the existence

of a limit (for increasing m) for the three interval representations which have been found. The limits are relatively easy to evaluate.

a. Periodically recurring exponential pulses

In Equation B.33 it was found that

$$E(\tau + [m-1]T) = \xi_m e^{-\lambda\tau}.$$

The limit of the coefficients is given by

$$\begin{aligned} \xi &= \lim_{m \rightarrow \infty} \xi_m \\ &= \lim_{m \rightarrow \infty} \frac{1 - e^{-\lambda mT}}{1 - e^{-\lambda T}} \\ &= \frac{1}{1 - e^{-\lambda T}} \end{aligned} \quad \text{B.42}$$

from which it follows immediately that

$$\begin{aligned} P_E(\tau) &= \lim_{m \rightarrow \infty} E(\tau + [m-1]T) \\ &= \xi e^{-\lambda\tau}. \end{aligned} \quad \text{B.43}$$

b. Periodically recurring critically damped pulses

In Equation B.14 it was found that:

$$C(\tau + [m-1]T) = \xi_m \tau e^{-\lambda\tau} + \rho_m e^{-\lambda\tau}$$

where the limit of ξ_m is defined by Equation B.42 and the limit of ρ_m is:

$$\begin{aligned}
\rho &= \lim_{m \rightarrow \infty} \rho_m \\
&= \lim_{m \rightarrow \infty} \frac{T e^{-\lambda T} [1 - m e^{-\lambda [m-1]T} + [m-1] e^{-\lambda m T}]}{[1 - e^{-\lambda T}]^2} \\
&= \frac{T e^{-\lambda T}}{[1 - e^{-\lambda T}]^2} .
\end{aligned}
\tag{B.44}$$

It follows that:

$$\begin{aligned}
P_C(\tau) &= \lim_{m \rightarrow \infty} C(\tau + [m-1]T) \\
&= \xi \tau e^{-\lambda \tau} + \rho e^{-\lambda \tau} .
\end{aligned}
\tag{B.45}$$

c. Periodically recurring damped sinusoids

In Equation B.41 it was shown that a periodically recurring damped sinusoid could be written as

$$S_\varphi(\tau + [m-1]T) = \alpha_m e^{-k\tau} \sin(v\tau + \varphi + \beta_1 - \beta_m).$$

Consider, for a moment, the limit of the amplitude coefficients α_m :

$$\begin{aligned}
\alpha &= \lim_{m \rightarrow \infty} \alpha_m \\
&= \lim_{m \rightarrow \infty} \frac{[1 - 2e^{-k m T} \cos v m T + e^{-2k m T}]^{1/2}}{[1 - 2e^{-k T} \cos v T + e^{-2k T}]^{1/2}} \\
&= \frac{1}{[1 - 2e^{-k T} \cos v T + e^{-2k T}]} .
\end{aligned}
\tag{B.46}$$

Noticing that

$$\alpha[1 - e^{-kmT}] \leq \alpha_m \leq \alpha[1 + e^{-kmT}], \quad \text{B.47}$$

one can see that the convergence of the sequence α_m is quite rapid as a function of m . A more detailed discussion of the convergence for speech is shown in Chapter 3, section 2, using typical values for natural vowels.

Recalling from Equation B.39 that:

$$\tan \beta_m = \frac{\sin vmT}{e^{kmT} - \cos vmT}$$

it is easy to see that

$$\frac{-1}{e^{kmT} - 1} < \tan \beta_m < \frac{1}{e^{kmT} - 1} \quad \text{B.48}$$

from which it follows that

$$\lim_{m \rightarrow \infty} \tan \beta_m = 0 \quad \text{B.49}$$

and thus

$$\begin{aligned} \beta &= \lim_{m \rightarrow \infty} \beta_m \\ &= 0. \end{aligned} \quad \text{B.50}$$

Heuristically it follows that:

$$\begin{aligned} P_{S_\varphi} &= \lim_{m \rightarrow \infty} S_\varphi(\tau + [m-1]T) \\ &= \alpha e^{-k\tau} \sin[v\tau + \varphi + \beta_1]. \end{aligned} \quad \text{B.51}$$

The easiest formal justification of Equation B.51 is to use Equation B.26 after time translation. The purpose of the above was to derive the limit for use in the next section and to have the form available for a later discussion of the rate of convergence in the case of speech. It should be mentioned that the above expression is an exact mathematical representation of the steady-state for a sum of periodically recurring damped sinusoids. Clearly, the damping factor and the resonance frequency of the decaying sinusoid both occur in this representation.

3. Representations on the Interval $(0, \infty)$

Looking at Equations B.1, B.2, and B.3, it is not clear that one should expect to be able to determine a representation which is valid for the interval $(0, \infty)$ as the sum of a periodic and transient term. A moment's reflection, however, shows that it would be possible, since B.1 is the output of an R-C circuit which has been impulsed with a sequence of delta functions. B.2 is the output of a critically damped R-L-C network which has been pulsed with a periodic sequence of delta functions, and B.3 is the output of a simple underdamped resonator which has been pulsed with a periodic sequence of Dirac delta functions. In each case we know that the output can be represented in terms of a transient solution and a periodic solution, although the periodic solution is generally expressed as a Fourier series. The limiting behavior of each of the three equations has been found on each period. Since the limiting behavior and the periodic solution are one and the same it is possible to find closed form representations for each of the three equations on the interval $(0, \infty)$.

a. Periodically recurring exponential pulses

In order to obtain the decomposition of Equation B.2, we shall begin with Equation B.7:

$$\begin{aligned}
 E(t) &= \frac{[1 - e^{-\lambda m T}]}{[1 - e^{-\lambda T}]} e^{-\lambda t} \\
 &= \frac{1}{1 - e^{-\lambda T}} e^{-\lambda t} - \frac{1}{1 - e^{-\lambda T}} e^{-\lambda t} e^{\lambda m T} \\
 &= T_E(t) + P_E(t),
 \end{aligned} \tag{B.52}$$

where

$$\begin{aligned}
 T_E(t) &= \frac{1}{1 - e^{-\lambda T}} e^{-\lambda t} \\
 &= -e^{-\lambda T} \xi e^{-\lambda t},
 \end{aligned} \tag{B.53}$$

$$P_E(t) = -\frac{1}{1 - e^{-\lambda T}} e^{-\lambda t} e^{\lambda m T}. \tag{B.54}$$

Translating time to the m^{th} period

$$\begin{aligned}
 P_E(\tau + [m-1]T) &= \frac{-1}{1 - e^{-\lambda T}} e^{-\lambda(\tau + [m-1]T)} e^{\lambda m T} \\
 &= \xi e^{-\lambda \tau}
 \end{aligned}$$

which is the limiting behavior determined in Equation B.43. It has been shown that

$$0 \leq t < \infty \quad E(t) = T_E(t) + P_E(t) \tag{B.55}$$

where $T_E(t)$ is a transient term and $P_E(t)$ is a periodic term. The

periodic term is the limiting behavior which was found previously.

b. Periodically recurring critically damped pulses

In order to determine the decomposition of $C(t)$ we shall begin with Equation B.14 and rearrange its terms in the following manner:

$$\begin{aligned}
 C(\tau+[m-1]T) &= \frac{[1 - e^{\lambda mT}]}{[1 - e^{\lambda T}]} (\tau+[m-1]T) e^{-\lambda(\tau+[m-1]T)} \\
 &+ \frac{-Te^{\lambda T}[1 - me^{\lambda[m-1]T} + [m-1]e^{\lambda mT}]}{[1 - e^{\lambda T}]^2} e^{-\lambda(\tau+[m-1]T)} \\
 &= \frac{(\tau+[m-1]T)e^{-\lambda(\tau+[m-1]T)}}{1 - e^{\lambda T}} - \frac{\tau e^{-\lambda(\tau-T)}}{1 - e^{\lambda T}} - \frac{[m-1]Te^{-\lambda(\tau-T)}}{1 - e^{\lambda T}} \\
 &\quad - \frac{-Te^{\lambda T}e^{-\lambda(\tau+[m-1]T)}}{[1 - e^{\lambda T}]^2} - \frac{Te^{\lambda T}(me^{\lambda[m-1]T} + [m-1]e^{\lambda mT})e^{-\lambda(\tau+[m-1]T)}}{[1 - e^{\lambda T}]^2} \\
 &= \frac{(\tau+[m-1]T)e^{-\lambda(\tau+[m-1]T)}}{1 - e^{\lambda T}} - \frac{Te^{\lambda T}e^{-\lambda(\tau+[m-1]T)}}{[1 - e^{\lambda T}]^2} \\
 &+ \frac{Te^{\lambda T}}{[1 - e^{\lambda T}]^2} e^{-\lambda\tau} - \frac{e^{\lambda T}}{[1 - e^{\lambda T}]} \tau e^{-\lambda\tau}.
 \end{aligned} \tag{B.56}$$

Letting

$$\begin{aligned}
 T_C(\tau+[m-1]T) &= \frac{(\tau+[m-1]T)e^{-\lambda(\tau+[m-1]T)}}{1 - e^{\lambda T}} \\
 &- \frac{Te^{\lambda T}e^{-\lambda(\tau+[m-1]T)}}{[1 - e^{\lambda T}]^2}
 \end{aligned} \tag{B.57}$$

we find

$$T_C(t) = -e^{-\lambda T} \xi t e^{-\lambda t} - \rho e^{-\lambda t}. \quad B.58$$

Let

$$\begin{aligned} P_C(t) &= \frac{T e^{-\lambda T} e^{-\lambda t}}{[1 - e^{-\lambda T}]^2} - \frac{e^{\lambda T}}{[1 - e^{-\lambda T}]} t e^{-\lambda t} \\ &= \rho e^{-\lambda t} + \xi t e^{-\lambda T}. \end{aligned} \quad B.59$$

Thus $P_C(t)$ is the limiting interval representation defined in Equation B.45. Extending $P_C(t)$ so that it is periodic for the positive real numbers we can then write:

$$0 \leq t < \infty \quad C(t) = T_C(t) + P_C(t) \quad B.60$$

where $T_C(t)$ is the transient term and $P_C(t)$ is the periodic term.

c. Periodically recurring damped sinusoids

To determine the periodic and transient components of the first equation, B.1, we shall decompose Equation B.26 as follows:

$$\begin{aligned} S_\varphi(t) &= \frac{e^{-kt}}{\mu_1} (e^{kmT} \sin[vt + \varphi - \gamma_1 - vmT] - \sin[vt + \varphi - \gamma_1]) \\ &= \frac{e^{-kt} e^{kmT}}{\mu_1} \sin[vt + \varphi - \gamma_1 - vmT] - \frac{e^{-kt}}{\mu_1} \sin[vt + \varphi - \gamma_1] \\ &= T_{S_\varphi}(t) + P_{S_\varphi}(t), \end{aligned} \quad B.61$$

where

$$T_{S_{\varphi}}(t) = \frac{e^{-kt}}{\mu_1} \sin[vt + \varphi - \gamma_1], \quad B.62$$

$$P_{S_{\varphi}}(t) = \frac{e^{-kt} e^{kmT}}{\mu_1} \sin[vt + \varphi - \gamma_1 - vmT]. \quad B.63$$

Clearly

$$\begin{aligned} P_{S_{\varphi}}(\tau + [m-1]T) &= \frac{e^{kmT} e^{-k(\tau + [m-1]T)}}{\mu_1} \sin[v(\tau + [m-1]T) + \varphi - \gamma_1 - vmT] \\ &= \frac{e^{kT}}{\mu_1} e^{-k\tau} \sin[v\tau + \varphi - \gamma_1 - vT] \\ &= \alpha e^{-k\tau} \sin[v\tau + \varphi + \beta_1] \\ &= P_{S_{\varphi}}(\tau) \end{aligned}$$

which was the periodic solution determined in Equation B.51. Thus we have found a periodic and a transient term such that:

$$0 \leq t \quad P_{S_{\varphi}}(t) + T_{S_{\varphi}}(t) = \sum_{n=0}^{\infty} \mu(t-nT) e^{-k(t-nT)} \sin[v(t-nT) + \varphi]. \quad B.64$$

The two main uses of this equation are:

- a) to consider the physical variables, k , v , T and how they affect the steady-state, and
- b) to observe the convergence to the steady-state condition.

4. Summary of Representations

This appendix has given a mathematical derivation of some closed form representations which are useful in discussing simple resonators. Even in the non-periodic case, there exist exact closed form (not infinite series) interval representations. The reason for assuming periodicity was to make it possible to study analytically the convergence to steady-state and other properties of the output of a simple resonator.

Bibliography

1. C. G. Bell, H. Fujisaki, J. M. Heinz, K. N. Stevens, and A. S. House, Reduction of Speech Spectra by Analysis-by-Synthesis Techniques, The Journal of the Acoustical Society of America, Volume 33, Number 12, December, 1961, pp. 1725-1736.
2. W. R. Bennett, The Correlatograph, The Bell System Technical Journal, Volume 32, Number 5, September, 1953, pp. 1173-1185.
3. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, Incorporated, New York, New York (1958).
4. B. Bogert, The Vobanc-A Two-to-One Speech Band-Width Reduction System, The Journal of the Acoustical Society of America, Volume 28, Number 3, May, 1956, pp. 399-403.
5. S. J. Campenella, D. C. Coulter, D. M. Early, R. E. Irons, and L. A. Barnes, Formant Tracking Vocoder System, Second Quarterly Report, December 1962—February 1963, Melpar, Incorporated. (AD 409099).
6. S. J. Campenella and D. C. Coulter, Formant-Period Tracker, The Journal of the Acoustical Society of America, Volume 35, Number 5, May, 1963, p. 806.
7. S. J. Campenella and D. Coulter, OVE II Simulation of Errors in Formant Tracking, STL-QPSR-3/1962, The Royal Institute of Technology, Stockholm, Sweden, 1962, pp. 28-29.
8. S. H. Chang, Two Schemes of Speech Compression System, The Journal of the Acoustical Society of America, Volume 28, Number 4, July, 1956, pp. 565-572.
9. S. H. Chang, R. E. Bach, Jr., C. R. Howard, and R. Sukys, The Formoder as a Tool for Speech Studies, Proceedings of the Seminar on Speech Compression and Processing, Volume 1, B-6, (1959). (AD 233717).
10. E. C. Cherry and V. J. Phillips, Some Possible Uses of Single Sideband Signals in Formant-Tracking Systems, The Journal of the Acoustical Society of America, Volume 33, Number 8, August, 1961, pp. 1067-1077.
11. P. R. Clement, On Completeness of Basis Functions Used for Signal Analysis, Society for Industrial and Applied Mathematics Review, Volume 5, Number 2, April, 1963, pp. 131-139.
12. R. Courant and D. Hilbert, Methods of Mathematical Physics, Interscience Publishers, Incorporated, New York, New York, Volume 1 (1953).

13. T. H. Crystal, Speech Analysis with Computer-Simulated Filter Bank, The Journal of the Acoustical Society of America, Volume 35, Number 11, November, 1963, p. 1912.
14. J. Dammann, Studies in the Application of Adaptive Threshold Devices to the Recognition of Acoustic-Phonetic States, Ph.D. dissertation, The University of Michigan, 1964.
15. E. E. David, Jr., Computer-Catalyzed Speech Research, Fourth International Congress on Acoustics, Congress Report II, (1962), pp. 69-88.
16. E. E. David, Jr., Artificial Auditory Recognition in Telephony, IBM Journal of Research and Development, Volume 2, Number 4, October, 1958, pp. 294-309.
17. E. E. David, Jr., Signal Theory in Speech Transmission, Institute of Radio Engineers Transactions on Circuit Theory, Volume CT-3, December, 1956, pp. 232-244.
18. K. H. Davis, Automatic Recognition of Phonetic Elements in Speech, The Journal of the Acoustical Society of America, Volume 25, Number 4, July, 1953, p. 832.
19. K. H. Davis, R. Biddulph, and S. Balashek, Automatic Recognition of Spoken Digits, The Journal of the Acoustical Society of America, Volume 24, Number 6, November, 1952, pp. 637-642.
20. P. Denes and M. V. Mathews, Spoken Digit Recognition Using Time-Frequency Pattern Matching, The Journal of the Acoustical Society of America, Volume 32, Number 11, November, 1960, pp. 1450-1455.
21. J. Descloux, Approximations in L^p and Chebyshev Approximations, Journal of the Society for Industrial and Applied Mathematics, Volume 11, Number 4, December, 1963, pp. 1017-1026.
22. L. O. Dolansky, Choice of Base Signals in Speech Signal Analysis, Institute of Radio Engineers Transactions on Audio, Volume AU-8, Number 6, November-December, 1960, pp. 221-229.
23. L. O. Dolansky, Electronically-Controlled Audio Filters, Proceedings of the Institute of Radio Engineers, Volume 43, Number 11, November, 1955, pp. 1580-1586.
24. H. Dudley, R. R. Riesz, and S. Watkins, A Synthetic Speaker, Journal of the Franklin Institute, Volume 227, June, 1939, pp. 739-764.
25. H. K. Dunn, Methods of Measuring Vowel Formant Bandwidths, The Journal of the Acoustical Society of America, Volume 33, Number 12, December, 1961, pp. 1737-1746.

26. H. K. Dunn, The Calculation of Vowel Resonances, and an Electrical Vocal Tract, The Journal of the Acoustical Society of America, Volume 22, Number 6, November, 1950, pp. 740-753.
27. R. M. Fano, The Information Theory Point of View in Speech Communication, The Journal of the Acoustical Society of America, Volume 22, Number 6, November, 1950, pp. 691-696.
28. R. M. Fano, Short-Time Autocorrelation Functions and Power Spectra, The Journal of the Acoustical Society of America, Volume 22, Number 5, September, 1950, pp. 546-550.
29. G. Fant, Speech Analysis and Synthesis, Technical Final Report, Royal Institute of Technology, Stockholm, Sweden (1962).
30. G. Fant, Acoustic Theory of Speech Production, Mouton and Company, 's-Gravenhage (1960).
31. G. Fant, Acoustic Analysis and Synthesis of Speech with Applications to Swedish, Ericsson Technics, No. 1 (1959), pp. 3-108.
32. G. Fant, Modern Instruments and Methods for Acoustic Studies of Speech, Royal Institute of Technology, STL, Report 8, Stockholm, Sweden (1957). (AD 146813).
33. G. Fant, Transmission Properties of the Vocal Tract, Massachusetts Institute of Technology, Acoustic Laboratory, Report No. TR 12 (1952).
34. G. Fant, K. Fintoft, J. Liljencrants, B. Lindblom, and J. Martony, Formant-Amplitude Measurements, The Journal of the Acoustical Society of America, Volume 35, Number 11, November, 1963, pp. 1753-1761.
35. J. L. Flanagan, Some Influences of the Glottal Wave Upon Vowel Quality, Proceedings of the Fourth International Congress on Phonetic Sciences, (1961), pp. 126-141.
36. J. L. Flanagan, Analog Measurements of Sound Radiation from the Mouth, The Journal of the Acoustical Society of America, Volume 32, Number 12, December, 1960, pp. 1613-1623.
37. J. L. Flanagan, Estimates of the Maximum Precision Necessary in Quantizing Certain "Dimensions" of Vowel Sounds, The Journal of the Acoustical Society of America, Volume 29, Number 4, April, 1957, pp. 533-534.
38. J. L. Flanagan, Note on the Design of "Terminal-Analog" Speech Synthesizers, The Journal of the Acoustical Society of America, Volume 29, Number 2, February, 1957, pp. 306-310.

39. J. L. Flanagan, Band Width and Channel Capacity Necessary to Transmit the Formant Information of Speech, The Journal of the Acoustical Society of America, Volume 28, Number 4, July, 1956, pp. 592-596.
40. J. L. Flanagan, Evaluation of Two Formant-Extracting Devices, The Journal of the Acoustical Society of America, Volume 28, Number 1, January, 1956, pp. 118-125.
41. J. L. Flanagan, Automatic Extraction of Formant Frequencies from Continuous Speech, The Journal of the Acoustical Society of America, Volume 28, Number 1, January, 1956, pp. 110-118.
42. J. L. Flanagan, A Difference Limen for Vowel Formant Frequency, The Journal of the Acoustical Society of America, Volume 27, Number 3, May, 1955, pp. 613-617.
43. J. L. Flanagan and A. S. House, Development and Testing of a Formant-Coding Speech Compression System, The Journal of the Acoustical Society of America, Volume 28, Number 6, November, 1956, pp. 1099-1106.
44. J. W. Forgie and C. D. Forgie, Results Obtained from a Vowel Recognition Computer Program, The Journal of the Acoustical Society of America, Volume 31, Number 11, November, 1959, pp. 1480-1489.
45. D. B. Fry, Duration and Intensity as Physical Correlates of Linguistic Stress, The Journal of the Acoustical Society of America, Volume 27, Number 4, July, 1955, pp. 765-768.
46. O. P. Gandhi, G. E. Peterson, and F. Yu, Recurrently Impulsed Resonators in Speech and Psychophysical Studies, Language and Speech, Volume 3, Part 3, July-September, 1960, pp. 140-154.
47. B. Gold, Computer Program for Pitch Extraction, The Journal of the Acoustical Society of America, Volume 34, Number 7, July, 1961, pp. 916-921.
48. M. Halle and K. Stevens, Speech Recognition: A Model and a Program for Research, Institute of Radio Engineers Transactions on Information Theory, Volume IT-8, Number 2, February, 1962, pp. 155-159.
49. C. M. Harris and W. M. Waite, Response of Spectrum Analyzers of the Bank-of-Filters Type to Signals Generated by Vowel Sounds, The Journal of the Acoustical Society of America, Volume 35, Number 12, December, 1963, pp. 1972-1977.
50. K. S. Harris, H. S. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants, The Journal of the Acoustical Society of America, Volume 29, Number 2, February, 1958, pp. 122-126.

51. J. M. Heinz, Analysis of Acoustic Speech Signals into Articulatory Parameters, The Journal of the Acoustical Society of America, Volume 35, Number 5, May, 1963, p. 779.
52. G. A. Hellwarth, An Automatic Speech Formant Tracking Filter, The University of Michigan, Communication Sciences Laboratory Report No. 10 (1962). (AD 282147).
53. L. Hermann, Nachtrag zur Untersuchung der Vocalcurven, Archives Gestalt Physiology, Volume 58 (1894), pp. 264-279.
54. J. N. Holmes, An Investigation of the Volume Velocity Waveform at the Larynx during Speech by Means of an Inverse Filter, Paper G13, Congress Report 1, Fourth International Congress on Acoustics, Copenhagen (1962).
55. R. A. Houde, W. D. Larkin, and L. C. Stewart, Performance Analysis of the General-Dynamics-Electronics Formant Vocoder, The Journal of the Acoustical Society of America, Volume 35, Number 11, November, 1963, p. 1912.
56. C. R. Howard, Speech Analysis-Synthesis Scheme Using Continuous Parameters, The Journal of the Acoustical Society of America, Volume 28, Number 6, November, 1956, pp. 1091-1098.
57. W. H. Huggins, Representation and Analysis of Signals, Part I. The Use of Orthogonalized Exponentials, Johns Hopkins University, AFRC TR-57-357, (1957). (AD 208133).
58. W. H. Huggins, A Phase Principle for Complex-Frequency Analysis and Its Implications in Auditory Theory, The Journal of the Acoustical Society of America, Volume 24, Number 6, November, 1952, pp. 582-589.
59. G. W. Hughes, The Recognition of Speech by Machine, Massachusetts Institute of Technology, Research Laboratory of Electronics, Report No. TR-395, May, 1961. (AD 268489).
60. D. Jackson, Fourier Series and Orthogonal Polynomials, Mathematical Association of America, Fourth Edition, The Collegiate Press, Menasha, Wisconsin, 1957.
61. J. E. Karlin and S. N. Alexander, Communication Between Man and Machine, Proceedings of the Institute of Radio Engineers, Volume 50, Number 5, May, 1962, pp. 1124-1128.
62. L. G. Kersta, Voiceprint Identification, Nature, Volume 196, Number 4861, December, 1962, pp. 1253-1257.
63. W. Koenig, H. K. Dunn, and L. Y. Lacy, The Sound Spectrograph, The Journal of the Acoustical Society of America, Volume 17, Number 4, July, 1946, pp. 19-49.

64. W. Koenig and A. E. Ruppel, Quantitative Amplitude Representation in Sound Spectrograms, The Journal of the Acoustical Society of America, Volume 20, Number 6, November, 1948, pp. 787, 795.
65. A. Krokstad, Formant Tracking, Royal Institute of Technology, Stockholm, Sweden, Report Number STL-QPSR-2/1960, January, 1961, p. 6.
66. Krohn-Hite Corporation, Operating and Maintenance Manual, Model 310-AB, 580 Massachusetts Avenue, Cambridge 39, Massachusetts.
67. P. Ladefoged and D. E. Broadbent, Information Conveyed by Vowels, The Journal of the Acoustical Society of America, Volume 29, Number 1, January, 1957, pp. 98-104.
68. W. Lawrence, Formant Tracking by Self Adjusting Inverse Filtering, Paper C5, Stockholm Speech Communications Seminar, Royal Institute of Technology (1962).
69. I. Lehiste and G. E. Peterson, Studies of Syllable Stress, The University of Michigan, Speech Research Laboratory, Report No. 3, November, 1959.
70. A. M. Liberman, Some Results of Research on Speech Perception, The Journal of the Acoustical Society of America, Volume 29, Number 1, January, 1957, pp. 117-123.
71. P. Lieberman, Perturbations in Vocal Pitch, The Journal of the Acoustical Society of America, Volume 33, Number 5, May, 1961, pp. 597-603.
72. B. Lindblom, Spectrographic Study of Vowel Reduction, The Journal of the Acoustical Society of America, Volume 35, Number 11, November, 1963, pp. 1773-1781.
73. B. Lindblom, Accuracy and Limitations of Sona-Graph Measurements, Proceedings of the Fourth International Congress of Phonetic Sciences, Mouton and Company, The Hague, pp. 188-200. (Plus comments on pp. 200-202). (1962).
74. H. J. Manley, Analysis-Synthesis of Connected Speech in Terms of Orthogonalized Exponentially Damped Sinusoids, The Journal of the Acoustical Society of America, Volume 35, Number 4, April, 1963, pp. 464-474.
75. M. V. Mathews, J. E. Miller, and E. E. David, Jr., Pitch Synchronous Analysis of Voiced Sounds, The Journal of the Acoustical Society of America, Volume 33, Number 2, February, 1961, pp. 179-186.
76. W. Meyer-Eppler, Grundlagen und Anwendungen der Informationstheorie, Springer-Verlag, Berlin (1959).

77. G. A. Miller, Decision Units in the Perception of Speech, Institute of Radio Engineers Transactions on Information Theory, Volume IT-8, Number 2, February, 1962, pp. 81-83.
78. R. L. Miller, Nature of the Vocal Cord Wave, The Journal of the Acoustical Society of America, Volume 31, Number 6, June, 1959, pp. 667-679.
79. W. A. Munson and H. C. Montgomery, A Speech Analyzer and Synthesizer, The Journal of the Acoustical Society of America, Volume 22, Number 5, September, 1950, p. 678.
80. K. Nakata, J. Suzuki, and K. Maezono, Speech Analysis by Computer, The Journal of the Radio Research Laboratories, (Japan), Volume 9, Number 43, May, 1962, pp. 277-291.
81. A. M. Noll, Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection, The Journal of the Acoustical Society of America, Volume 36, Number 2, February, 1964, pp. 296-302.
82. H. F. Olson and H. Belar, Phonetic Typewriter III, The Journal of the Acoustical Society of America, Volume 33, Number 11, November, 1961, pp. 1610-1615.
83. H. F. Olson and H. Belar, Phonetic Typewriter, The Journal of the Acoustical Society of America, Volume 28, Number 6, November, 1956, pp. 1072-1081.
84. H. F. Olson, H. Belar, and R. de Sobrino, Demonstration of a Speech Processing System Consisting of a Speech Analyzer, Translator, Typer, and Synthesizer, The Journal of the Acoustical Society of America, Volume 34, Number 10, October, 1962, pp. 1535-1538.
85. A. P. Paul, A. S. House, and K. N. Stevens, Automatic Reduction of Vowel Spectra: An Analysis-by-Synthesis Method and Its Evaluation, The Journal of the Acoustical Society of America, Volume 36, Number 2, February, 1964, pp. 303-308.
86. E. Peterson, Frequency Detection and Speech Formants, The Journal of the Acoustical Society of America, Volume 23, Number 6, November, 1951, pp. 668-674.
87. G. E. Peterson, Automatic Speech Recognition Procedures, Language and Speech, Volume 4, Number 4, October-December, 1961, pp. 200-219.
88. G. E. Peterson, Parameters of Vowel Quality, Journal of Speech and Hearing Research, Volume 4, Number 1, March, 1961, pp. 10-29.
89. G. E. Peterson, Vowel Formant Measurements, Journal of Speech and Hearing Research, Volume 2, Number 2, June, 1959, pp. 173-183.

90. G. E. Peterson, The Information-Bearing Elements of Speech, The Journal of the Acoustical Society of America, Volume 24, Number 6, November, 1952, pp. 629-637.
91. G. E. Peterson and H. L. Barney, Control Methods Used in a Study of the Vowels, The Journal of the Acoustical Society of America, Volume 24, Number 2, March, 1952, pp. 175-184.
92. E. N. Pinson, Pitch-Synchronous Time-Domain Estimation of Formant Frequencies and Bandwidths, The Journal of the Acoustical Society of America, Volume 35, Number 8, August, 1963, pp. 1264-1273.
93. S. Pruzansky, Pattern-Matching Procedure for Automatic Talker Recognition, The Journal of the Acoustical Society of America, Volume 35, Number 3, March, 1963, pp. 354-358.
94. J. R. Rice, Chebyshev Approximation by Exponentials, Journal of the Society for Industrial and Applied Mathematics, Volume 10, Number 1, March, 1962, pp. 149-161.
95. H. Rusten, Synthesis of R-L-C Networks by Discrete Tschebyscheff Approximations in the Time Domain, The University of Michigan, Electronic Defense Group, Report No. TR-107, April, 1960 (AD 238392).
96. T. Sakai, The Phonetic Typewriter: Its Fundamentals and Mechanism, Studia Phonologica, Volume 1 (1961), pp. 140-152. (Institution for Phonetic Sciences, University of Kyoto, Japan).
97. T. Sakai and S. Doshita, An Automatic Recognition System of Speech Sounds, Studia Phonologica, Volume 2 (1962), pp. 83-95. (Institution for Phonetic Sciences, University of Kyoto, Japan).
98. T. Sakai and S. I. Inoue, New Instruments and Methods for Speech Analysis, The Journal of the Acoustical Society of America, Volume 32, Number 4, April, 1960, pp. 441-450.
99. M. R. Schroeder and B. S. Atal, Generalized Short-Time Power Spectra and Autocorrelation Functions, The Journal of the Acoustical Society of America, Volume 34, Number 11, November, 1962, pp. 1679-1683.
100. E. W. Scripture, Researches in Experimental Phonetics. The Study of Speech Curves, The Carnegie Institution of Washington, Publication 44, Gibson Brothers Press (1906).
101. G. Sebestyen, Pattern Recognition by an Adaptive Process of Sample Set Construction, Institute of Radio Engineers Transactions on Information Theory, Volume IT-8, Number 5, September, 1962, pp. 82-91.
102. G. S. Sebestyen, Recognition of Membership in Classes, Institute of Radio Engineers Transactions on Information Theory, Volume IT-7, Number 1, January, 1961, pp. 44-50.

103. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana (1959).
104. P. Simmons, Automation of Speech, Speech Synthesis and Synthetic Speech, A Bibliographical Survey from 1950-1960, Institute of Radio Engineers Transactions on Audio, Volume AU-9, Number 6, November-December, 1961, pp. 191-196.
105. L. R. Spogen, H. N. Shaver, D. E. Baker, and B. V. Blom, Speech Processing by the Selective Amplitude Sampling System, The Journal of the Acoustical Society of America, Volume 32, Number 12, December, 1960, pp. 1621-1625.
106. K. N. Stevens, Toward a Model for Speech Recognition, The Journal of the Acoustical Society of America, Volume 32, Number 1, January, 1960, pp. 47-55.
107. K. N. Stevens, Autocorrelation Analysis of Speech Sounds, The Journal of the Acoustical Society of America, Volume 22, Number 6, November, 1950, pp. 769-772.
108. K. N. Stevens and A. S. House, An Acoustical Theory of Vowel Production and Some of Its Implications, The Journal of Speech and Hearing Research, Volume 4, Number 4 (1961), pp. 303-320.
109. J. Suzuki, Y. Kadokawa, and K. Nakata, Formant-Frequency Extraction by the Method of Moment Calculation, The Journal of the Acoustical Society of America, Volume 35, Number 9, September, 1963, pp. 1345-1353.
110. L. R. Talbert, G. F. Groner, and J. S. Koford, Real-Time Adaptive Speech-Recognition Scheme, The Journal of the Acoustical Society of America, Volume 35, Number 5, May, 1963, p. 807.
111. R. Timcke, H. von Leden, and P. Moore, Laryngeal Vibrations: Measurements of the Glottic Wave; Part 1. The Normal Vibratory Cycle, Archives of Otolaryngology, Volume 68, July, 1958, pp. 1-19.
112. J. V. Tobias, Relative Occurrence of Phonemes in American English, The Journal of the Acoustical Society of America, Volume 31, Number 5, May, 1959, p. 631.
113. G. Ungeheuer, Elemente Einer Akustischen Theorie der Vokalartikulation, Springer-Verlag, Berlin (1962).
114. F. Vilbig and K. Haase, Some Systems for Speech-Band Compression, The Journal of the Acoustical Society of America, Volume 28, Number 4, July, 1956, pp. 573-577.
115. E. S. Weibel, Vowel Synthesis by Means of Resonant Circuits, The Journal of the Acoustical Society of America, Volume 27, Number 5, September, 1955, pp. 858-865.

116. M. R. Weiss and C. M. Harris, Computer Technique for High-Speed Extraction of Speech Parameters, The Journal of the Acoustical Society of America, Volume 35, Number 2, February, 1963, pp. 207-214.
117. N. Wiener, Speech Language and Learning, The Journal of the Acoustical Society of America, Volume 22, Number 6, November, 1960, pp. 696-697.
118. J. Wiren and H. L. Stubbs, Electronic Binary Selection System for Phoneme Classification, The Journal of the Acoustical Society of America, Volume 28, Number 6, November, 1956, pp. 1082-1091.