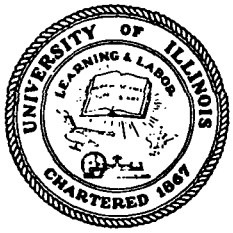
Report NASA-6
UILU-ENG-72-2542

Asynchronous Sampling Of Speech With Some Vocoder Experimental Results



M. L. Babcock

Final Report

Grant NGR 14-005-111
National Aeronautics and Space Administration

Biological Computer Laboratory
Electrical Engineering Research Laboratory
Engineering Experiment Station
University of Illinois
Urbana, Illinois 61801

BCL

Report NASA-6

UILU-ENG-72-2542

ASYNCHRONOUS SAMPLING OF SPEECH WITH SOME VOCODER EXPERIMENTAL RESULTS

by

M. L. Babcock

FINAL REPORT

June 1972

National Aeronautics and Space Administration

Grant NGR 14-005-111

BIOLOGICAL COMPUTER LABORATORY
ELECTRICAL ENGINEERING RESEARCH LABORATORY
ENGINEERING EXPERIMENT STATION
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS

TABLE OF CONTENTS

Page Intentionally Left Blank

LIST OF FIGURES

ASYNCHRONOUS SAMPLING OF SPEECH WITH SOME VOCODER EXPERIMENTAL RESULTS

1. ABSTRACT

A speech waveform can be sampled asynchronously, resulting in a sequence of samples which are not at all periodic nor indicative of the amplitude of the speech signal but which definitely carry sufficient information for both intelligibility and speaker identification. This method of asynchronously sampling speech is based upon the derivatives of the acoustical speech signal.

The following results are apparent from experiments to date:

(1) It is possible to represent speech by a string of pulses of uniform amplitude, where the only information contained in the string is the spacing of the pulses in time.

(2) The string of pulses may be produced in a simple analog manner.

(3) The first derivative of the original speech waveform is the most important for the encoding process.

(4) The resulting pulse train can be utilized to control an acoustical signal production system to regenerate the "intelligence" of the original speech.

This paper describes the present processes involved in obtaining the pulses, some results of analysis of the pulse

distributions, and a description of a simple attempt to produce the original speech from the pulse code.

2. INTRODUCTION

2.1 Spectrographic Analysis

The analysis and synthesis of speech has had an extensive and long history. During much of this history, the concepts and underlying principles of Fourier and his ideas of the representation of mathematical functions by sinusoidal components have dominated the field. In the earlier history, these concepts produced an understanding of speech which are based upon the sinusoidal components and the conceptual framework provided by the Fourier representation. The work of Von Kempelen, Wheatstone, and Willis, as exemplified by their mechanical models of the vocal tract which produced vowel-like speech sounds, served to give substance to the frequency based approach. Additionally, Ohm and his "law of acoustics" and Helmholtz and his resonators were instrumental in structuring the total framework for the analysis of speech.

With the invention of the sound spectrograph in the early thirties, technology contributed an instrument which has become widely used for the analog spectra analysis of speech. The perfection of this instrument has reached such a state and it is so simple and easy to use that it has become the reference standard and dominant analysis method for the study of auditory sounds. This is so even though the sound spectrograph largely ignores phase information and so at best gives only the amplitude of the sinusoidal components. Thus it and most

related instruments are deficient in at least one aspect of a general purpose analyzer, i.e., the phase or time relations among the various components.

This deficiency can be accepted if one also accepts the common misinterpretation of some of Ohm's results. Ohm's "law of acoustics" states that the subjective response to a particular sound is independent of the phase relationship of the frequency components of that sound. This has been disproven a number of times, but the misconception that the ear is insensitive to phase persists in many forms. A very simple but adequate illustration that the ear can detect phase is the comparison of the phase of impulse function with white noise. These signals theoretically have identical power spectra but differ only in the phase relationships among the components. The "ear-brain system" distinguishes between these signals ridiculously easily. (It should be emphasized that in this example it is much more meaningful to consider the time-domain descriptions of these signals than the frequency-domain characterizations.)

Aside from the area of speech analysis and synthesis, there are two ways of describing the reaction of a system to an impulse function. The classical physicist's view has been somewhat along the lines of the concepts of frequency analysis. He has contended that an impulse function contains all the frequencies which are possible under any circumstance and that the system which is excited by the impulse merely selects those which are "natural" to the system. From this philosophical

point of view, an interesting question is "What happened to the other frequencies?" Of course, they can all be identified in the infinite spectra of the resulting damped sinusoidal response.

In contrast, the modern view of systems is that a finite amount of energy is delivered to the system in zero time interval and at a specified time. The system stores this energy and dissipates it over an infinite time with a response waveform depending upon the time-constants of the system. This is the free response of the system. The parameters which characterize the system and the response are both time and frequency based.

Another factor contributing to the dominance of spectrographic analysis was the concept of Helmholtz resonators and their production of certain types of waveforms at various resonant frequency combinations. From this came the concept that a Helmholtz resonator reinforces the sound at various resonant frequencies. It does this, but only at the expense of energy at some other time--commonly attributed to other frequencies. More properly, that is, the reinforcement amounts to a redistribution of the source energy with respect to time. The total energy produced by the source remains unchanged. The interval over which the energy is redistributed is dependent upon the time constant or "Q" of the resonators in combination with the energy distribution in the energy source. Thus, as a conceptual model of the speech source, the Helmholtz resonators were satisfactory for the formants that could be detected and

examined by similar types of resonators (filters) which were used in the sound spectrograph.

When the sound spectrograph is being excited by a nearly repetitive signal, then the filters will closely indicate the frequency of the source signal. When the signal is non-repetitive or changing rapidly in frequency, then the filter cannot respond rapidly enough, and the result is a redistribution of the energy over an interval in time which is both indicative of the frequencies present in the signal and of the natural response of the filters.

As a result of the analysis by filters, there is an apparent dominance of vowel-like energy in the signal source. In addition to this emphasis of quasi-steady-state periodic energy, the transient non-periodic energy is effectively redistributed in time so that the transients appear smoothed and diminished in amplitude; something akin to an integration process is effected. Thus the consonants of speech, which contribute most of the information to speech communication, are deemphasized and the vowels are emphasized.

Other examples of factors contributing to the dominance of frequency analysis can be cited. Most of these are from disciplines other than acoustics where successes were obtained by the use of frequency analysis. However, these successes often were not due to frequency analysis alone, but rather to the instrumentation methods in conjunction with the particular phenomena being analyzed. (For example, often the excitation signals were sinusoidal in nature; thus the sinusoidally based

analysis was very intrinsically related to the signals being investigated.) But the successes were often attributed to the method itself, and the result was that frequency analysis became the panacea of many unanalyzed phenomena, speech included. It can therefore be stated that even though the frequency spectrum representation of certain signals and systems has proved to be useful for specifying certain characteristics of systems, it does not necessarily follow that the same type of representation is the most desirable for specifying the parameters of all signals, or more specifically, for specifying those of speech.

Thus it may well be that the dominance of one conceptual scheme and the existence of a simple menas of measurement using this reference base have contributed a false sense of effort and direction to the analysis of speech. Certainly the results of spectrographic analysis over the last two decades have failed to contribute sufficiently to the knowledge of speech to adequately abstract the parameters needed for acceptable speech "recognition".

2.2 Other Techniques

It therefore seems reasonable to consider more novel techniques which can be based upon some unified theoretical understanding of the acoustical speech signal, its production, and its perception; these techniques will have as a goal the automated processing of speech, including specification, translation, transmission, synthesis, and "recognition". Consequently, study at the University of Illinois concerns the fundamental

properties of speech which are significant for automatic speech processing. Since such properties are affected to some degree by subsequent "recognition algorithms", this study cannot completely ignore the recognition procedures to be eventually employed; however, the initial emphasis must be on the discovery of promising parameter (feature) extraction techniques.

The fundamental thesis of the present work asserts that the sustained, quasi-stationary relationships and spectral aspects of speech are less important than are the dynamic qualities and the timed sequences. The attempt is thus made to explore the time-domain of speech signals and the relationships which may exist there.

The only major similarly oriented work which has been done in recent years is that by Licklider in the early 1950's, who explored the zero-axis crossing information of clipped speech. His interest was essentially of a behavioral psycho-acoustic perceptual nature, however, and not concerned with the basic characterization of parameters. Licklider found that normal speech could be infinitely clipped so that essentially only rectangular pulses remained. The resulting speech signal was then heard by trained listeners, the result being an articulation index of 70%. The overall naturalness was low. When the normal speech was replaced with differentiated speech, Licklider found that the naturalness improved considerably and the intelligence level increased to 90% from the previous 70%. Incidentally, when the speech was deprived by center clipping, keeping the peaks only, the results were disastrous, the speech becoming nothing but static.

If one models the vocal tract as a cascaded set of resonant circuits and hypothesizes that the excitation from the glottis is very impulsive in nature, then one can easily visualize that the intrinsic components of speech are damped sinusoids. This intuition is even closer to reality during the consonant and non-vowel sounds where the glottis is not the only source of excitation for the vocal tract. In practice, it seems to be a good hypothesis that, regardless of how and where the acoustic signal is generated, it appears to be made up of a considerable number of damped sinusoidal waveforms, at least to the first approximation.

One can relate this hypothesis to the axis-crossing work and measurements. Whether or not a signal is a damped sinusoid or an undamped sinusoid cannot be generally established by simply monitoring the axis-crossings alone, since these axis-crossings occur at the same times for both signals. However, if the signal is differentiated, the axis-crossings of the derivatives occur at different times, and indeed, the damping constant and the frequency may be found to be a function of the time intervals between axis-crossings of the signal and its derivatives. This certainly suggests the value of differentiating the signal and investigating the sequences of axis-crossing times for the derivatives.

If one does just this and also uses the axis crossings to produce short interval, uniform amplitude pulses, then a sequence of equal amplitude pulses occurring at irregular intervals results. This technique is called "asynchronous

sampling". The "axis" crossed by the waveform need not be the zero level axis, but may be any axis between the zero level and the peak level of the waveform. If these pulse positions are examined and analyzed in terms of pitch-synchronous histograms, there appears to be enough "patternicity" present to indicate the possibility of limited recognition capability.

However, the most striking result thus far is the following: When such pulse trains are used to excite a limited bandwidth amplifier and loud speaker system, the resulting output is speech of reasonably high intelligibility. This indicates that speech may be represented by sequences of "samples" or time markers which are not independently time-periodic but are intrinsically speech-periodic. These markers do not explicitly indicate the amplitude of the speech signals, but they definitely carry sufficient information for both intelligibility and talker identification. That is, it is the sequence and epochical time of the samples which is important.

3. ASYNCHRONOUS SAMPLING PROCESS

3.1 Experiment

Figure 1 is a functional diagram of the Speech Signal Processing System. The concepts and methods involved in asynchronously sampling speech can be illustrated by the following simple experimental procedure.

The differentiator was set for the zero'th derivative and the bias variable was changed over the range of maximum signal rejection to zero-axis rejection. That is, with the bias set very high, none of the signal was present; with the

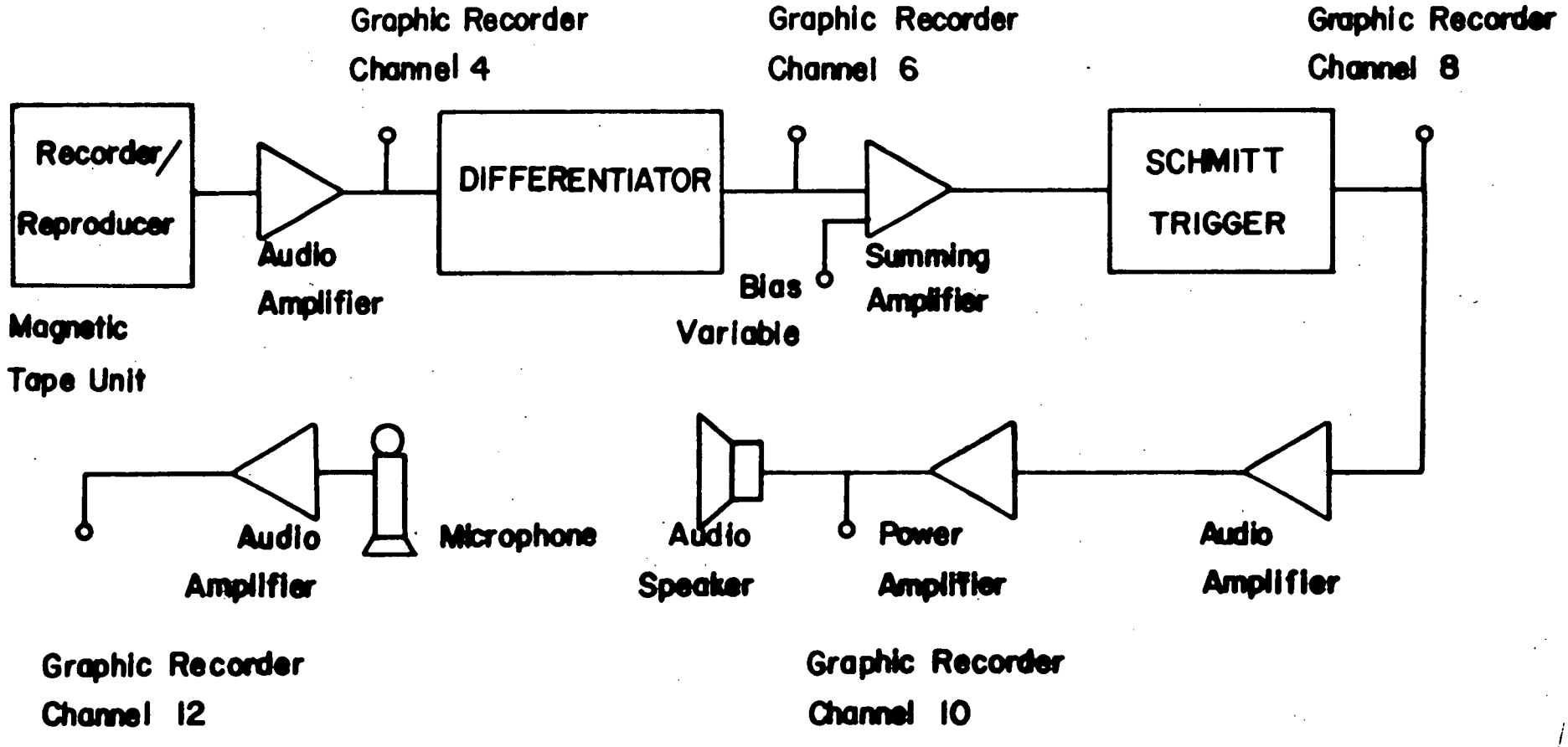


Figure 1. Speech Signal Processing System

bias at zero, the signal above the zero axis was present. It should be emphasized that the signal into the speaker was not the semi-rectified speech waveform, but rather was a set of pulses of standard amplitude but occurring at the times of the bias axis crossing points. The results of this experiment were that the bias must be very close to the zero level for the signal to be interpreted as intelligible speech. A bias of 10% of the maximum signal extinction bias produced very unintelligible speech. Also, speaker identification was low.

The differentiator was then set for the first derivative and the bias again varied from maximum signal rejection to zero-axis signal rejection. The results of this experiment indicated that a much higher bias was acceptable for the signal to be interpreted as intelligible speech. In some instances, the bias could be as high as 30% of the maximum signal extinction bias for intelligible speech.

Moreover, if the signal was monitored aurally as the bias was reduced from extinction, at a value where approximately 10-15 glottis pulse groups were present in the signal to the audio speaker, the vowels became intelligible. In fact, the differences among the vowels appeared distinctly to be different distributions of the pulses within a glottal pulse group when viewed on the graphic recorder trace. In contrast, before the bias had reached the value of proper vowel identification when one was adjusting its values downward from maximum extinction, frequently a vowel was misunderstood. As an example of this, the vowel /e/ in "head" was often confused with the

vowel /i/ in "hid" until the bias was lowered sufficiently for proper identification.

Proceeding with higher order derivatives through the fifth, exactly analogous results were obtained. For a given intelligibility, the bias value can be the highest for the second derivative, with the value dropping off for higher order derivatives. (It should, however, be noted in this respect that the data for these higher derivatives is not too reliable and should be considered only as indicative of the results; it is hoped that future work will produce more reliable data.)

3.2 Equipment

The actual sampling device operates as follows. The speech event, in the form of an electrical representation (e.g., voltage), is fed into a series of differentiators and amplifiers. A D.C. bias is added, so that the average value of the wave is other than zero.

The wave is then fed into the trigger of a General Radio pulser, which acts on it in the following manner: every time there is a negative transition through zero voltage, the pulser produces one pulse, of specified amplitude and duration, as shown in Figure 2.

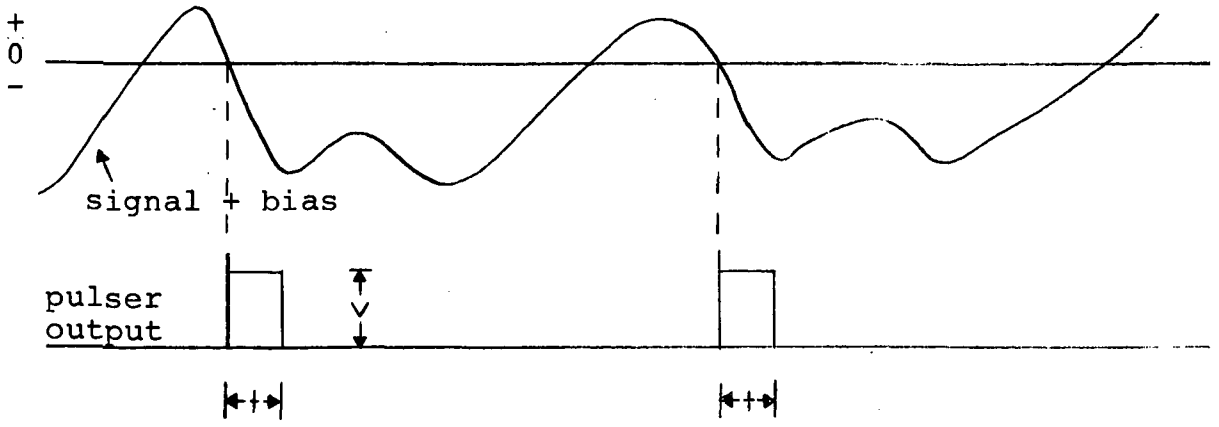


Figure 2. A pulse coding of speech.

The resulting string of pulses is fed into an audio amplifier and then into a loud speaker, allowing the experimenter to monitor the pulser output aurally.

Using this technique, it was found that standardized pulse representations of speech events could indeed be recognized, and that the intelligibility of the audio signal produced at the system output depended upon the positioning of the cutting axis for the trigger of the pulser, the order of the derivative, and the pulse duration chosen as standard.

3.3 Intelligibility Tests

The intelligibility test of the system was performed in the following manner. The 11 vowels listed in Table I were placed in an h-d environment by pronouncing the vocabulary words shown.

Table I

Vowel	Vocabulary Word
u	who'd
ʊ	hood
o	hoed
ɔ	hawed
ʌ	hud
a	hod
æ	had
ɛ	head
e	hade
I	hid
i	heed

These words were placed in the larger context phase, "the first word who'd had finished", etc. so that they were not spoken in isolation. The tape was then edited so that only the vocabulary word remained for the intelligibility test.

The 11 resulting words were normalized in amplitude in the following manner. They were fed through the pulse train generator system shown in Figure 1 and the d.c. level shifter was changed until only a small number of pulses resulted at the output for only one of the words. This word thus had the greatest individual peak amplitude in its waveform. Using this word as the "standard normalized peak value", the other words were adjusted in amplitude on an individual basis, using the same fixed reference value for the d.c. level shifter. The result, when all the words had been normalized, was a set of 11 vowels in the h-d environment, each word spoken in isolation and each producing a small number (approximately five) of pulses at the output of the pulser when it was used to excite the pulse train generator system. These 11 words then became the standard vocabulary of vowels.

Next, since the intelligibility of these vowels under various levels of zero-axis shift was to be determined, the range of the d.c. level shift from the zero value to the maximum value just used for the normalizing value was determined. This range was arbitrarily partitioned into 10 levels. Thus at the d.c. shift level 1 the waveform was being sampled at the zero-axis level (0% peak value) of the normal speech acoustical signal and at the d.c. shift level 10, the waveform was being sampled practically at its 90% peak value.

Therefore, considering a total of 10 levels for each of the 11 vocabulary words, there resulted a total test sample of 110 words to be tested in isolation for their intelligibility. Since the test was to be a critical test of the intelligibility present in the sampled speech, the test was made deliberately severe by randomizing these 110 samples and selecting six naive listeners to evaluate the intelligibility contained in the samples. That is, the listeners had not heard any of the samples previously, had not been trained in the test procedures involved, were not screened in any manner (they were not speech trained subjects; in fact, two of them were high school students), and were told only to write the word which they heard as each sample was played to them through the system. The subjects were in a quiet environment when they evaluated the samples.

The results were recorded and evaluated by means of a computerized evaluation program. In fact, the entire randomization and evaluation process was done by means of a computer program. This program, which may be used for any number of speech evaluation tests, will appear in a forthcoming report. It automatically conducts all computations required for confusion matrix evaluations and points the resulting matrices in standard format form. One such confusion matrix is shown in Figure 3. This is for a 30% peak level sampling of normal undifferentiated speech.

The results of the intelligibility tests to date are as follows:

PERCENTAGE SCORE

STIMULUS	RESPONSE 1	2	3	4	5	6	7	8	9	10	11	12	13
1	50.0	16.7	16.7			16.7							
2		16.7			16.7		33.3			33.3			
3	16.7	50.0	16.7					16.7					
4							83.3			16.7			
5							100.						
6	16.7		16.7					16.7		50.0			
7						16.7	83.3						
8	33.3		16.7				16.7	16.7		16.7			
9	16.7	16.7		16.7						50.0			
10	50.0								16.7	33.3			
11	16.7					16.7	33.3	16.7					16.7
12													
13													

Figure 3. Confusion matrix for the 11 vowels sampled at the 30 percent peak level.

(1) The intelligibility of normal speech sampled at the 0 to 30% peak level is approximately 60%. Considering the severe restrictions placed upon the tests by the methods and procedures used, this seems very good, but certainly not of "vocoder" quality.

(2) The intelligibility of first order differentiated speech sampled at the 0 to 30% peak level is greater than that of normal speech, the intelligibility being in the range of 80% for words spoken in isolation, again a very severe restriction.

When samples of speech in context are used and informally evaluated by listeners, the indications are that the pulse train generator system can be used to produce highly intelligible speech in the "vocoder" sense. With this background, and with the results of other work utilizing this system, it was decided to investigate the preceeding system at the synthesis and vocoder level. These experiments will be described next. The vocabulary and basic intelligibility test procedures were greatly relaxed for these tests and follow the more normally accepted testing procedures for vocoders.

4. THREE CHANNEL VOCODER USING THRESHOLD ASYNCHRONOUS SAMPLING

4.1 Equipment

The system developed at the Biological Computer Laboratory of the University of Illinois is an extension of the one channel vocoder discussed above. In this system the input speech signal is passed through a series of differentiators

which produce the zeroth, first and second derivatives. Each derivative is then fed into a separate General Radio pulse generator which operates in the manner previously described in Figure 2. The outputs of the pulse generators are fed through bandpass filters, which have adjustable resonant frequencies, and into an audio amplifier-speaker combination. Thus there are three separate channels, each consisting essentially of a differentiator, pulse generator, filter, and loudspeaker. Switches on the loudspeakers allow the monitoring of any one channel or any combination of channels. A block diagram of this system and its circuit diagrams are shown in Figure 4. The object of this experiment was to determine the intelligibility as the filters of each channel are adjusted and also as the different channels are monitored acoustically in the various combinations.

Filters were included because it was felt that the addition of them to the system would increase the intelligibility. Figure 5 shows the response of the speaker both with and without the filters. The pulse width for each speaker is the same as that used during the testing. It can be seen that the resonant frequency of the speaker with the filter is nearly the same as that without the filter. Hence, the reduction in bandwidth resulting from the filters makes that system less intelligible. It should be noted that the filters are set to a position which the author feels yields maximum intelligibility. This is a personal value judgment and must be weighed proportionally.

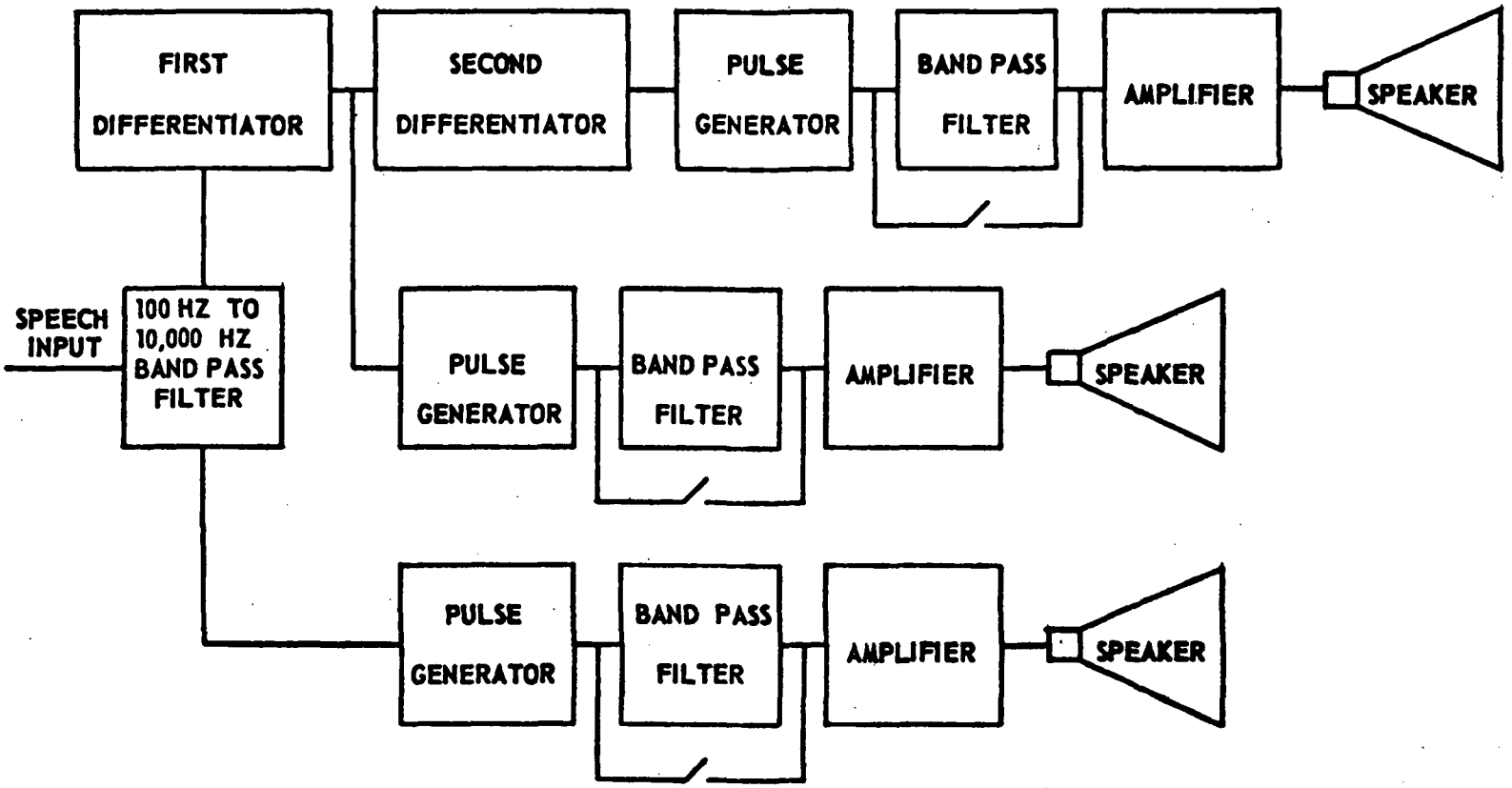
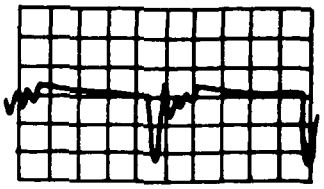
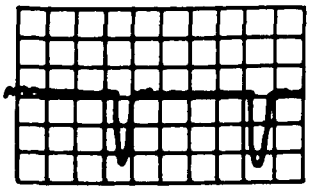


Figure 4. Block diagram of three channel sampler



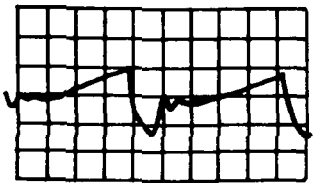
1 v/div.

Zeroth Derivative
No Filter
20 μsec./div.
20 μsec. pulse in.



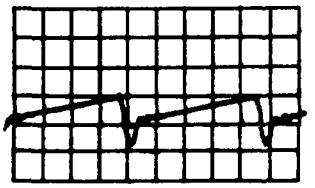
1 v/div.

First Derivative
No Filter
20 μsec./div.
7 μsec. pulse in.



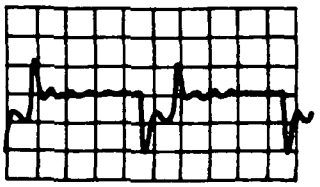
1 v/div.

Zeroth Derivative
1700-2050 Hz Bandpass Filter
1800 Hz Resonant Freq.
20 μsec./div.
20 μsec. pulse in.



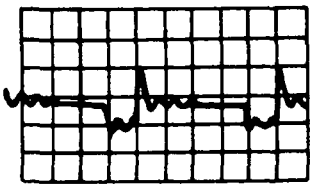
1 v/div.

First Derivative
1430-1700 Hz Bandpass Filter
1600 Hz Resonant Freq.
20 μsec./div.
7 μsec. pulse in.



1 v/div.

Second Derivative
No Filter
20 μsec./div.
17 μsec. pulse in.



0.1 v/div.

Second Derivative
1800-2300 Hz Bandpass Filter
2100 Hz Resonant Freq.
20 μsec./div.
17 μsec. pulse in.

Figure 5. Speaker response.

4.2 Tests

The word choice tests in Tables II and III were used to test the intelligibility of the system without the filters. The test was conducted in the following manner. One of the two words in each word pair was chosen by the author and recorded on tape with the phrase "I will write..." preceeding that word. The subject taking the test listens and then chooses the word that he believes comes through the speaker. He chooses the word and goes on to the next pair. Words in the pairs are somewhat similar in sound to see just how efficient the system is. The only difference between Test I and Test II is the derivative associated with the various words. This was controlled by the author operating the switches on the loudspeakers. The test was divided into seven parts in order to test the derivatives and their combinations. Those derivatives being tested are:

Zeroth derivative alone
First derivative alone
Second derivative alone
Zeroth and first derivatives
Zeroth and second derivatives
First and second derivatives
Zeroth and first and second derivatives

Ten subjects were divided evenly into two groups for each test with about one half of each group being somewhat familiar with a system of this type (it appears that listening experience increases the intelligibility for this or any other system). One subject was given the test listening solely to the unprocessed speech directly from the tape recorder. His results

TABLE II

ZEROTH DERIVATIVE

1	gauze	-	cause
2	moot	-	boot
3	feet	-	peat
4	champ	-	camp
5	pop	-	top
6	goal	-	dole
7	tilt	-	kilt
8	dent	-	tent
9	moss	-	boss
10	thew	-	too
11	cheese	-	keys
12	map	-	nap
13	cod	-	pod
14	so	-	show
15	big	-	pig
16	net	-	debt

SECOND DERIVATIVE

17	thong	-	tong
18	juice	-	goose
19	peach	-	teach
20	cast	-	past
21	sot	-	shot
22	dote	-	tote
23	mitt	-	bit
24	fend	-	pend
25	chalk	-	caulk
26	moon	-	noon
27	keep	-	peep
28	bass	-	gas
29	vault	-	fault
30	dot	-	tot
31	moan	-	bone
32	thin	-	tin

FIRST DERIVATIVE

33	jet	-	get
34	pall	-	tall
35	you	-	woo
36	teal	-	keel
37	vast	-	fast
38	mom	-	bomb
39	fort	-	port
40	chink	-	kink
41	pest	-	test
42	yawl	-	wall
43	boon	-	goon
44	veal	-	feel
45	mat	-	bat
46	fond	-	pond
47	choke	-	coke
48	pip	-	tip
49	shed	-	said

FIRST & SECOND DERIVATIVE

50	ball	-	gall
51	zoo	-	sue
52	need	-	deed
53	fan	-	pan
54	chock	-	cock
55	post	-	toast
56	shift	-	sift
57	pen	-	ken
58	boom	-	doom
59	jaw	-	chaw
60	nude	-	dude
61	thee	-	dee
62	sad	-	thad
63	mob	-	nob
64	coal	-	pole
65	bill	-	gill
66	bent	-	pent

TABLE II (continued)

ZEROTH & FIRST & SECOND DERIVATIVE

67	gnaw	-	daw
68	fool	-	pool
69	seem	-	theme
70	bad	-	dad
71	shock	-	sock
72	bold	-	gold
73	gin	-	chin
74	mend	-	bend
75	fawn	-	pawn
76	chew	-	coo
77	weed	-	reed
78	can	-	tan
79	bob	-	gob
80	goad	-	code
81	nip	-	dip
82	vest	-	best
83	sought	-	thought

ZEROTH & SECOND DERIVATIVE

84	womb	-	room
85	she	-	see
86	ram	-	yam
87	tee	-	knee
88	box	-	pox
89	nose	-	doze
90	fin	-	pin
91	jest	-	guest
92	bong	-	dong
93	coot	-	toot
94	wield	-	yield
95	gaff	-	calf
96	knock	-	dock
97	vote	-	boat
98	sink	-	think
99	bed	-	dead

ZEROTH & FIRST DERIVATIVE

100	caught	-	taught
101	ruse	-	use
102	bean	-	peen
103	nab	-	dab
104	von	-	bon
105	joe	-	go
106	bid	-	did
107	yen	-	wen
108	raw	-	yaw
109	dune	-	tune
110	meat	-	beat
111	than	-	dan
112	jot	-	cot
113	mode	-	node
114	kit	-	pit
115	ted	-	ked

TABLE III

FIRST DERIVATIVE

1	gauze	-	cause
2	moot	-	boot
3	feet	-	peat
4	champ	-	camp
5	pop	-	top
6	goal	-	dole
7	tilt	-	kilt
8	dent	-	tent
9	moss	-	boss
10	thew	-	too
11	cheese	-	keys
12	map	-	nap
13	cod	-	pod
14	so	-	show
15	big	-	pig
16	net	-	debt

FIRST & SECOND DERIVATIVE

17	thong	-	tong
18	juice	-	goose
19	peach	-	teach
20	cast	-	past
21	sot	-	shot
22	dote	-	tote
23	mitt	-	bit
24	fend	-	pend
25	chalk	-	caulk
26	moon	-	noon
27	keep	-	peep
28	bass	-	gas
29	vault	-	fault
30	dot	-	tot
31	moan	-	bone
32	thin	-	tin

ZEROTH DERIVATIVE

33	jet	-	get
34	pall	-	tall
35	you	-	woo
36	teal	-	keel
37	vast	-	fast
38	mom	-	bomb
39	fort	-	port
40	chink	-	kink
41	pest	-	test
42	yawl	-	wall
43	boon	-	goon
44	veal	-	feel
45	mat	-	bat
46	fond	-	pond
47	choke	-	coke
48	pip	-	tip
49	shed	-	said

ZEROTH & SECOND DERIVATIVE

50	ball	-	gall
51	zoo	-	sue
52	need	-	deed
53	fan	-	pan
54	chock	-	cock
55	post	-	toast
56	shift	-	sift
57	pen	-	ken
58	boom	-	doom
59	jaw	-	chaw
60	nude	-	dude
61	thee	-	dee
62	sad	-	thad
63	mob	-	nob
64	coal	-	pole
65	bill	-	gill
66	bent	-	pent

TABLE III (continued)

ZEROTH & FIRST DERIVATIVE

67	gnaw	-	daw
68	fool	-	pool
69	seem	-	theme
70	bad	-	dad
71	shock	-	sock
72	bold	-	gold
73	gin	-	chin
74	mend	-	bend
75	fawn	-	pawn
76	chew	-	coo
77	weed	-	reed
78	can	-	tan
79	bob	-	gob
80	goad	-	code
81	nip	-	dip
82	vest	-	best
83	sought	-	thought

ZEROTH & FIRST & SECOND DERIVATIVE

84	womb	-	room
85	she	-	see
86	ram	-	yam
87	tee	-	knee
88	box	-	pox
89	nose	-	doze
90	fin	-	pin
91	jest	-	guest
92	bong	-	dong
93	coot	-	toot
94	wield	-	yield
95	gaff	-	calf
96	knock	-	dock
97	vote	-	boat
98	sink	-	think
99	bed	-	dead

SECOND DERIVATIVE

100	caught	-	taught
101	ruse	-	use
102	bean	-	peen
103	nab	-	dab
104	von	-	bon
105	joe	-	go
106	bid	-	did
107	yen	-	wen
108	raw	-	yaw
109	dune	-	tune
110	meat	-	beat
111	than	-	dan
112	jot	-	cot
113	mode	-	node
114	kit	-	pit
115	ted	-	ked

and comments are shown in Table IV and should be taken into account during the evaluation of the other tests.

The pulse width and amplitude are set to what the author believes yields maximum intelligibility. There is a two to three microsecond range over which the pulse width may vary without affecting the intelligibility of the system. The pulse widths and amplitudes for the various derivatives are listed below.

Zeroth	-	15 microseconds	-	5 volts
First	-	7 microseconds	-	5 volts
Second	-	17 microseconds	-	7 volts

The dc bias voltage is 30 percent of the signal level for the first derivative while the other pulsers have a zero voltage threshold level.

4.3 Results

The opinion of the subjects tested was that although the second derivative was least intelligible by itself, when added to any other derivatives, it made the resulting combination more intelligible. This seems to be supported by Table V, which shows the results of the tests given to 10 people. Listed in Column 1 is a number assigned each person. The headings of the other columns are the combinations of derivatives being tested. The listings in each column are the number of mistakes made by each person for the various derivatives. When the second derivative was added to the zeroth derivative, the number of mistakes decreased in five cases, increased in one case, and remained the same in four. When the second

TABLE IV

1	gauze	-	cause	
2	moot	-	boot	
3	feet	-	peat	
4	champ	-	camp	
5	pop	-	top	
6	goal	-	dole	
7	tilt	-	kilt	
8	dent	-	tent	
9	moss	-	boss	stretched
10	thew	-	too	
11	cheese	-	keys	
12	map	-	nap	
13	cod	-	pod	
14	so	-	show	
15	big	-	pig	
16	net	-	debt	
17	thong	-	tong	
18	juice	-	goose	
19	peach	-	teach	
20	cast	-	past	
21	sot	-	shot	
22	dote	-	tote	
23	mitt	-	bit	
24	fend	-	pend	
25	chalk	-	caulk	
26	moon	-	noon	phone noise
27	keep	-	peep	
28	bass	-	gas	
29	vault	-	fault	
30	dot	-	tot	
31	moan	-	bone	stretched
32	thin	-	tin	
33	jet	-	get	
34	pall	-	tall	
35	you	-	woo	
36	teal	-	keel	
37	vast	-	fast	
38	mom	-	bomb	
39	fort	-	port	
40	chink	-	kink	
41	pest	-	test	
42	yawl	-	wall	
43	boon	-	goon	
44	veal	-	feel	
45	mat	-	bat	
46	fond	-	pond	
47	choke	-	coke	
48	pip	-	tip	
49	shed	-	said	
50	ball	-	gall	
51	zoo	-	sue	
52	need	-	deed	
53	fan	-	pan	
54	chock	-	cock	
55	post	-	toast	
56	shift	-	sift	
57	pen	-	ken	slightly
58	boom	-	doom	confused
59	jaw	-	chaw	
60	nude	-	dude	
61	thee	-	dee	
62	sad	-	thad	
63	mob	-	nob	
64	coal	-	pole	
65	bill	-	gill	
66	bent	-	pent	

TABLE IV (continued)

67	gnaw	-	daw	
68	fool	-	pool	
69	seem	-	theme	not clear
70	bad	-	dad	
71	shock	-	sock	
72	bold	-	gold	
73	gin	-	chin	
74	mend	-	bend	
75	fawn	-	pawn	
76	chew	-	coo	
77	weed	-	reed	
78	can	-	tan	
79	bob	-	gob	
80	goad	-	code	
81	nip	-	dip	
82	vest	-	best	
83	sought	-	thought	
84	womb	-	room	
85	she	-	see	
86	ram	-	yam	
87	tee	-	knee	
88	box	-	pox	
89	nose	-	doze	
90	fin	-	pin	
91	jest	-	guest	
92	bong	-	dong	
93	coot	-	toot	
94	wield	-	yield	
95	gaff	-	calf	
96	knock	-	dock	
97	vote	-	boat	
98	sink	-	think	
99	bed	-	dead	
100	caught	-	taught	
101	ruse	-	use	
102	bean	-	peen	
103	nab	-	dab	
104	von	-	bon	incorrect
105	joe	-	go	
106	bid	-	did	
107	yen	-	wen	
108	raw	-	yaw	
109	dune	-	tune	
110	meat	-	beat	
111	than	-	dan	
112	jot	-	cot	
113	mode	-	node	
114	kit	-	pit	
115	ted	-	ked	

MISTAKES IN DERIVATIVES

Subject No. Tested	DERIVATIVES USED						
	0	1	2	0 + 2	1 + 2	0 + 1	0 + 1 + 2
1	1	2	2	0	0	0	2
2	0	4	2	0	1	0	3
3	1	3	1	1	3	1	3
4	0	1	2	0	1	0	1
5	2	0	3	2	0	0	2
6	3	2	7	1	3	3	0
7	2	4	4	1	0	0	5
8	1	3	5	3	2	3	0
9	1	5	4	0	3	5	1
10	2	3	5	1	0	1	1

TABLE V

Commonly mistaken word pairs	Mistakes made by five subjects tested	Derivative used
Moot - Boot	4	1
Champ - Camp	4	1
Cast - Past	4	1 + 2
Cast - Past	4	2
Fort - Port	5	0
Bold - Gold	4	0 + 2
Bold - Gold	4	0 + 1 + 2

TABLE VI

derivative was added to the first derivative, the number of mistakes decreased in six instances, increased in one, and remained the same in three. Similarly, when the first derivative was added to the zeroth, the number of mistakes was reduced in four cases, increased in two cases, and remained the same in four. When the first derivative was added to the second derivative, the number of mistakes decreased in nine cases and increased in only one. The average test scores ranged from a low of 78.1 percent with the second derivative to a high of 94.38 percent with the zeroth and second derivatives.

Care must be taken, however, in interpreting these results; two factors in particular should be considered. First, the fact that this is a forced choice test (i.e., the listener is forced to choose one of two words), introduces a certain amount of guesswork; the listener may guess an answer when he really doesn't know. Second, the number of tests given was insufficient to make any general conclusions possible.

Shown in column 1 of Table VI are the most commonly missed word pairs. The second column gives the number of mistakes made by the five subjects tested. Listed in the last column is the derivative used in that test. Further study should include word pairs of these types in an attempt to establish any definite trends and possible reasons for the apparent confusion.

It appears that a combination of derivatives produces greater intelligibility than any one derivative alone. This

increase in intelligibility might be further increased with the addition of higher derivatives. A reasonable approximation to the speech wave which is sought may indeed be realized using the principle of adaptive sampling on the various derivatives of speech.

5. PROGNOSIS

These results indicate trends that should be studied further. The combination of derivatives does seem to improve the intelligibility. Future research will include a study of the pulse distributions to see if word recognition is possible. In addition, the acoustic output of the various derivatives and their combinations will be recorded and compared with the original signal to determine any possible correlation.

REFERENCE

1. Babcock, M. L., J. W. Atwood, J. R. Cohen, M. P. Hoffman: Search and Evaluation of Significant Event Sequences in Automated Speech Analysis, NASA Report NGR 1400S-111, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, (1969).

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
University of Illinois
Biological Computer Laboratory
Urbana, Illinois 61801

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
ASYNCHRONOUS SAMPLING OF SPEECH WITH SOME VOCODER EXPERIMENTAL RUSULTS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Final Report 1 July 1967 - 31 January 1971

5. AUTHOR(S) *(Last name, first name, initial)*
M. L. Babcock

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1972	34	1

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
NASA-NGR 14-005-111	
b. PROJECT AND TASK NO.	NASA-6
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	UILU-ENG-72-2542

10. AVAILABILITY/LIMITATION NOTICES
Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	National Aeronautics and Space Administration Washington, D. C. 20546

13. ABSTRACT

A speech waveform is sampled asynchronously, resulting in a sequence of samples which are not at all periodic nor indicative of the amplitude of the speech signal but which contain sufficient information for both intelligibility and speaker identification. This method of asynchronously sampling speech is based upon the derivatives of the acoustical speech signal.

The following are the results from experiments to date:

(1) It is possible to represent speech by a string of uniform amplitude pulses, where the spacing of the pulses contains the only information.

(2) The string of pulses may be produced in a simple analog man manner.

(3) The first derivative of the original speech waveform is the most important for the encoding process.

(4) The resulting pulse train can be utilized to regenerate the the "intelligence" of the original speech.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Speech Analysis Speech Synthesis Speech Transmission Time Domain of Speech Coded Speech Vocoder Pulse Coding of Speech						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U).*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.