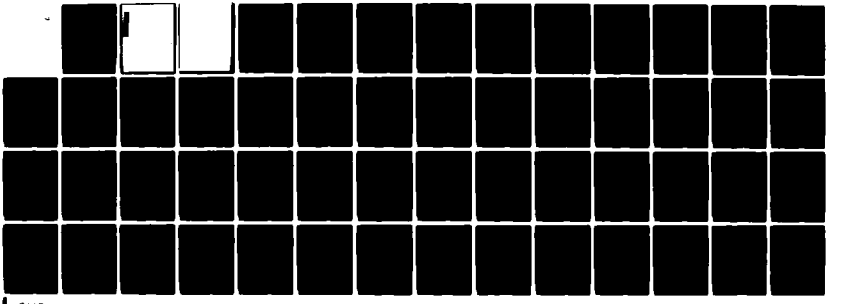
AD-A138 650 EVALUATION OF THE COMPREHENSION OF NON-CONTINUOUS SPED-UP VOCODED SPEECH:..(U) MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB J T LYNCH ET AL. 05 JAN 84 1/1

UNCLASSIFIED TR-671 ESD-TR-83-230 F19628-80-C-0002 F/G 17/2 NL

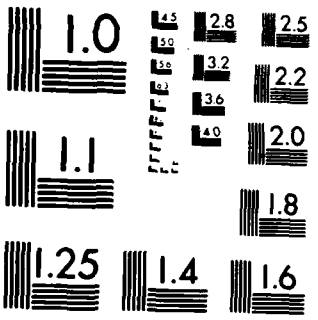


END

DATE FILMED

4-84

DTIC



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

EVALUATION OF THE COMPREHENSION OF NON-CONTINUOUS, SPED-UP VOCODED SPEECH: A STRATEGY FOR COPING WITH FADING HF CHANNELS

J.T. LYNCH
B. GOLD
J. TIERNEY
C.J. BOWERS
Group 24

TECHNICAL REPORT 671

5 JANUARY 1984

LEXINGTON MASSACHUSETTS

ABSTRACT

A novel technique for coping with fading and burst noise on ~~HF~~ high frequency channels used for digital voice communication has been developed and evaluated. The technique transmits digital voice only during the high signal-to-noise ratio time intervals, i.e., channel "on" times, and speeds up the speech when necessary in order to avoid delays which would hinder conversation. The technique was evaluated using a model of the human speech comprehension process, which was tested using a spoken version of a reading comprehension test. The test involved fifteen spoken, two-minute paragraphs processed by a real-time channel vocoder simulation which had been modified to also simulate the on/off characteristics of a fading ~~HF~~ channel. Using a speed-up factor of 1.5, the percentage of correct test responses verified the comprehension model. If the average "on" time is longer than about two seconds or the average channel "off" time is shorter than about one-half second, then the speech is comprehensible. Since these conditions are met for most disturbed and undisturbed ionospheric conditions, it is concluded that the sped-up speech technique is appropriate for ~~HF~~ high frequency digital voice communication systems.

CONTENTS

	Abstract	iii
	Figure Captions	vi
I.	INTRODUCTION	1
	A. HF Channel Characteristics	1
	B. Advanced HF Communications Equipment	6
II.	SPED-UP SPEECH STRATEGY	10
III.	COGNITIVE SPEECH PERCEPTION MODEL	15
IV.	SPEECH COMPREHENSION TESTING	19
V.	SIMULATION AND EXPERIMENTAL FACILITY	23
VI.	TEST RESULTS	25
VII.	SUMMARY AND CONCLUSIONS	37
	References	44
	Appendix I	46
	Appendix II	47

FIGURE CAPTIONS

1. Multiple propagation modes supported by ionospheric refraction. 2

2. HF ionogram showing multiple modes. 4

3. Block diagram of sped-up speech transmission system. 11

4. Block diagram of system with input and output buffers. 14

5. Hypothesized speech intelligibility as a function of speech "on" time and "off" time. 18

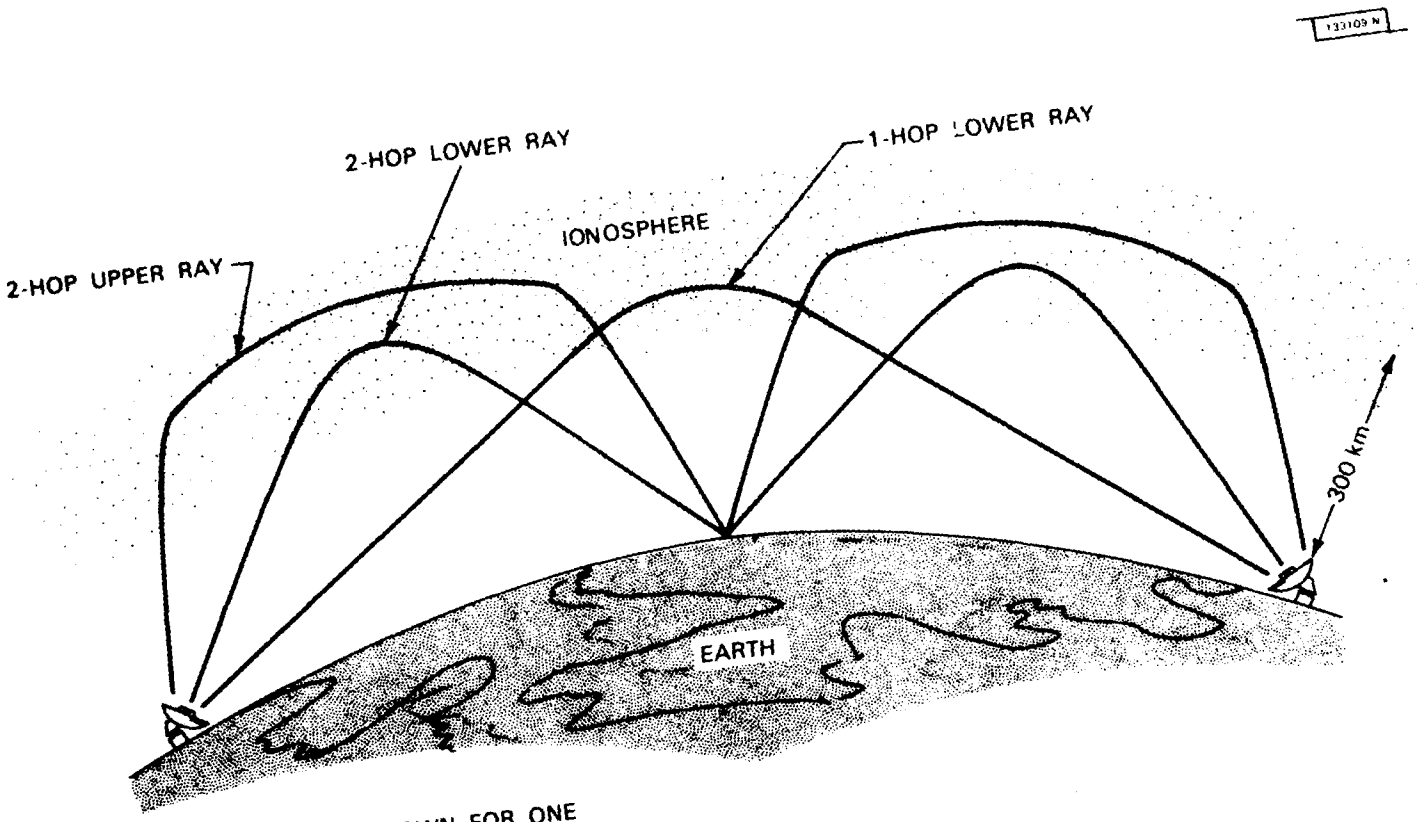
I. INTRODUCTION

High Frequency (HF) communications (3-30 MHz) channels are widely used because long-range paths (5000 km) are reliable and inexpensive. However, the ionosphere which supports the channel by refracting HF signals back towards the earth is responsible for producing time-varying fading caused by interference of multi-path signals. This fading together with natural and human generated burst noise produces errors in digital communications links. Because as little as a 5 percent bit-error-rate (BER) can severly degrade a digital speech system, such a system is far more vulnerable to HF fading and burst noise than is a conventional analog speech system. Surmounting this digital system difficulty is desirable in order to make the HF system compatible with global digital secure communication systems. This report describes the development and evaluation of a novel approach to making reliable a digital speech HF communication system.

A. HF Channel Characteristics

The ionosphere is a region in the earth's upper atmosphere (100-500 km) which has large numbers of ions and free electrons ($\simeq 10^6/cm^3$) produced by ultraviolet light, x-rays, and particle radiation from the sun [1]. These electron densities decrease by over an order of magnitude at night which results in lowering the maximum usable frequency (MUF) from about 30 MHz to about 10 MHz.

The daytime ionosphere supports several propagation modes for long distance paths. Figure 1 shows an F lower-ray path, an F upper-ray path, both double-hop; and a single-hop path. On very long paths (5000 km) six-hop signals may propagate.



NOTE: RAY PATHS SHOWN FOR ONE RADIO FREQUENCY, f_1 (See Figure 2)

Fig. 1. Multiple propagation modes supported by ionospheric refraction.

The particular ray-path depends on the signal frequency. Figure 2 shows the group-delay versus frequency for a typical channel in a format commonly referred to as an ionogram. Figure 2 indicates the radio frequency at which the propagation modes in Fig. 1 exist. Another source of signal multipath is caused by magneto-ionic splitting. Each propagation mode is split into two components called an ordinary-ray and an extraordinary-ray. The signals in these split modes are separated by microseconds rather than milliseconds. The two rays have different polarizations and consequently the coherent interference of the two signals is often called polarization fading. Under quiet ionospheric conditions polarization fading is much slower in its variation than is multi-hop fading.

Other propagation phenomena also contribute to signal fading. Large scale inhomogeneities in the ionosphere cause rays to focus their energy on parts of the earth at the expense of other parts. A lower region of the ionosphere called the D-layer has little refractive power but is the greatest source of absorption in the signals. Variations in space and time of this layer also produce changes in the signal strength.

Fading due to interference of the multi-hop and upper and lower ray signals varies considerably because of constantly varying ionospheric conditions. The ionosphere behaves much like clouds or the surface of the ocean in that there are waves of long extent as well as local variations.

Under quiet ionospheric conditions, the fading of long-path HF channels has time constants of about 10 seconds or more. Under disturbed conditions time constants as short as one second are common. These

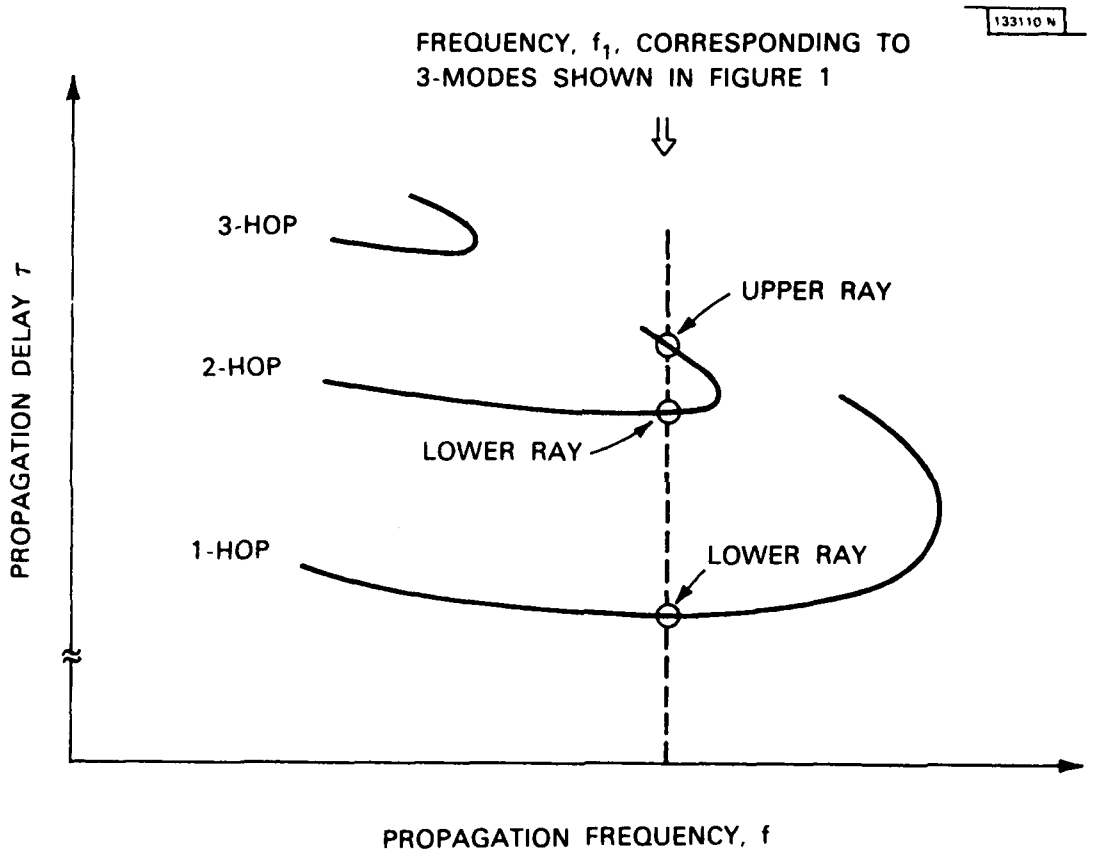


Fig. 2. HF ionogram showing multiple modes.

estimates derive from the traditional method of measuring these channel statistics which has been to spectrum analyze the amplitude (envelope) of a CW signal [1]. A record of frequency spectra versus time of day is thereby obtained. For magnetically quiet conditions, the spread of the spectra is about 0.1 Hz, while for magnetically disturbed conditions the average spread is about 1 Hz with 2 or 3 Hz spreads being possible. These records tend to hide slower varying fading due to focusing and D-layer absorotion and hence give upper bounds on the fading rate.

The details of the signal envelope depend on the number and relative amplitudes of the interfering signal modes. Two interfering signals of nearly identical strength produce an envelope characterized by strong intervals of high amplitude separated by short but very deep nulls (≈ -40 dB relative to the peak). Five or six interfering signals, on the other hand, would produce a less organized envelope pattern and would seldom have extraordinary deep nulls but might have frequent shallow nulls (≈ -15 dB). In either case the fades have much shorter duration than the strong-signal intervals.

In subsequent sections of this report, it will be expedient to refer to typical fade durations. Such a characterization is now produced with the obvious caveat that it is a gross approximation since many subtleties are ignored. Under quiet conditions, fades of 1 to 4 seconds separated by about 10 to 40 seconds are herein defined as typical, while for disturbed conditions the intervals could be a factor of 10 shorter (0.1 to 0.4 s fades, 1 to 4 s apart).

Predictions of the fading time constants for various path geometries, and signal frequencies based on the degree of ionospheric disturbance are

not available. Furthermore, the experiences of HF communications personnel may underestimate the degree of future ionospheric disturbance because of increasing solar activity in the next several years.

In addition to the diurnal variations in the ionosphere due to the earth's rotation there are ionospheric variations due to sunspot activity. Sunspots are responsible for the variation in HF propagation conditions which occur on the time scales of minutes to seconds. The year 1985 is near a minimum in sunspot activity so that any recent HF channel tests were made under near optimum propagation conditions. In about seven years (1990-1991), the sunspot activity will be at a maximum for the 10 year cycle and near a maximum for the 50 year cycle. Consequently, future operational HF systems will have to use a disturbed ionosphere which will cause more rapid fading and have stronger and more frequent atmospheric noise bursts. It is important, therefore, that systems planned for future operational use be tested using simulated channels and that channel parameters be chosen to represent these expected disturbed ionospheric conditions.

B. Advanced HF Communications Equipment

Within the last five years, there have been considerable advances in the state of the art in HF communication equipment. These advances extend the range of propagation conditions which support reliable digital voice communication by reducing data error rates under fading channel conditions. This section contains a description of three different approaches, one of which is exploited in a speech transmission algorithm reported on here.

Advanced HF digital modems have either a parallel or a series structure. The parallel structured modem divides the available 3 kHz bandwidth into many, say n, parallel subbands. Each subband contains a single sine wave whose phase is modified by 0° or 180° every nT seconds, where T is roughly 3 kHz^{-1}. The parallel modem suppresses fading-induced channel errors because multipath fading is frequency selective and effects only one of the n parallel subbands. A simple error correction code is used to provide a signal which is ideally errorless. A time domain interpretation of the modem's operation is that if each of the subband signals is much longer than the multipath spread (1-5 ms on a 3-5 hop path), then the multipath does not significantly affect signalling.

The series-structured modem uses a single sine wave whose phase is changed every T seconds (rather than nT seconds), where again T is roughly 3 kHz^{-1}. The modem uses channel compensation techniques to extract the correct phase of the signal. The channel compensation filter coefficients can be obtained by frequently sending a reference signal or by an adaption procedure where the data itself is used to estimate channel characteristics.

The series structure gives better performance than the parallel structure for reasonable signal-to-noise ratios because multi-hop paths can cause several of the parallel channels to produce errors which cannot be corrected by simple coding. In addition, the signal level of each of the sine waves of the parallel modem must be less than the single sine wave of the series modem in order to constrain the peak power in the composite signal. Consequently, depending on the number of parallel channels,

approximately an extra 4-8 dB of signal power is required by the parallel modem to achieve the same performance as the serial modem.

On the other hand, in low signal-to-noise situations the series modem fails catastrophically because the required channel calibration can be inadequate. The parallel modem, on the other hand, fails gracefully and hence has superior performance in these infrequent but potentially significant situations.

These general issues are now used to describe recently developed HF modems. The Naval Research Laboratory (NRL) has developed a parallel modem designed especially for use with digital speech [2]. The Advanced Narrowband Voice Terminal (ANDVT) modem has 39 channels thereby providing much more immunity to fading than the conventional parallel modems which use 6 or 12 tones. The ANDVT uses a modem block length which is compatible with the vocoder frame length and employs redundant coding on the most important vocoder data parameters.

The Harris Corporation [3] has developed a series modem which uses a channel calibration signal as part of every data block. They can use binary-phase (0°-180°), quadruple phase (0°, 90°, 180°, 270°) or 8-ary phase (45° increments) and achieve 2.4, 4.8, or 7.2 kb/s data rates.

GTE Sylvania [4] has developed a series modem which uses an adaptive algorithm to determine the channel compensation filter coefficients after an initial channel estimation procedure is executed. The GTE modem also has an automatic repeat request (ARQ) feature which allows it to send virtually error-free data at the expense of a variable data rate resulting from repeated data blocks. The GTE modem has theoretically twice the data rate of the Harris modem because the latter uses half of the available data

bits to calibrate the channel. This advantage evaporates if the GTE modem's adaptive channel estimation algorithms fail due to low signal-to-noise ratio and the modem has to re-acquire and repeat data using the ARQ feature. The GTE modem requires a reliable low data-rate feedback channel which must be considered in deploying and testing this system.

These three advanced HF modems can make significant improvements in the usefulness of HF digital voice communications. There are two reasons, however, for seeking alternative techniques to improve HF digital voice systems. The first is that there will always be ionospheric propagation conditions for which the above modems will fail. Whether or not such a failure is important will depend on the function served by the communication link. How often such failures might occur depends on the state of the ionosphere. As discussed earlier, the 1982-1985 period will have a very quiet ionosphere because of a minimum of sunspot activity while the 1990-1991 period will have the highest level of sunspot activity seen in the last 50 years, and hence a very disturbed ionosphere.

The second reason for seeking alternate techniques for improving HF digital voice communications is to provide system designers with performance and cost trade-offs. The series modems require state-of-the-art digital processing equipment, and hence are more expensive in terms of size, power, weight, and procurement cost. Alternative approaches may be less expensive, and may be easier to incorporate into existing systems. The next section describes an approach which has the potential of satisfying these objectives.

II. SPED-UP SPEECH STRATEGY

The automatic-repeat-request (ARQ) feature of the GTE Sylvania modem suggests a very simple way to use a time-varying, fading HF channel: Only use it when it is good. This idea is easy to exploit for data communication but at first glance is not applicable to digital speech communication, especially for two-way conversations. It could be very bothersome for a speaker to have to wait for an intermittent "on" signal in order to talk. If the channel went on and off for time intervals as short as one or two seconds to tens of seconds, then it would be exceedingly difficult, if possible at all, to only speak during the channel "on" times. This problem can be solved by appropriately buffering and speeding up the speech so that the speaker can be unaware of the channel disruptions. The listener, however, may have difficulty understanding the speech if it is broken up too much. The issue of comprehending intermittant speech is theoretically and experimentally addressed in subsequent sections. This section describes how the sped-up speech algorithm works to take advantage of the channel "on" times.

A simple version of the system is shown in Fig. 3. The speech is assumed to be vocoded at 2400 b/s and feeds an input buffer. If the channel is "on," the speech is immediately transmitted over the channel. At the output a vocoder synthesizes the speech. If the channel is momentarily "off," that is it has an unacceptable error rate, the digital speech is stored in the input buffer. When the channel is "on" again, the buffer is emptied. However, if the speaker continues to speak, there will be a delay; in fact, an endlessly growing delay in the speech transmission

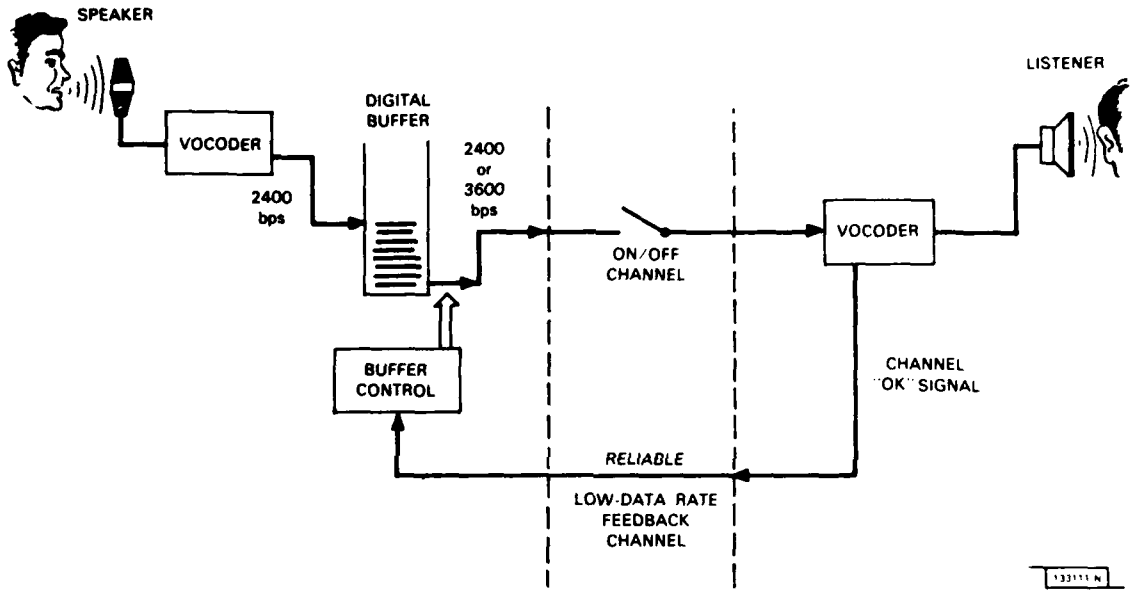


Fig. 3. Block diagram of sped-up speech transmission system.

unless the digital speech is transmitted at a faster rate than it was spoken. At the receiver the speech must also be synthesized at this faster rate in order to avoid accruing a limitless buffer and increasing delays.

A speed-up factor of 1.5 was used in the studies reported here. Many informal listeners in our laboratory were unaware that the speech was sped up at all at this rate, and appear to have no difficulty in understanding words or comprehending sentences. Research on reading aids for the blind [5] has shown that speed-up factors up to 2 produce comprehensible speech. Experimental data was obtained in this present study which verifies that the speed-up of 1.5 introduced no measurable degradation in the comprehensibility of the speech.

The increased transmission rate can be achieved in several ways. As noted earlier a modem using phase shift keying (FSK) can achieve faster data rates by employing smaller phase shift increments, i.e., 90° instead of 180°. This faster rate is obtained at the cost of a higher bit error-rate (BER). However, in a non-fading condition even standard modems have BER of better than 10^{-3} so that an order-of-magnitude increase in the BER would not degrade the vocoded speech because vocoders can withstand one percent error rates before the speech is perceptibly altered. Another way to speed up the speech transmission is to use more efficient (and less intelligable) vocoder algorithms such as 1800 b/s LPC instead of 2400 b/s LPC coding but continue to use a 2400 B/S data rate.

In the system discussed thus far, the transmission rate and the vocoder synthesis rate are identical. Several advantages accrue if a receiver buffer is used to allow these two rates to be independently set.

The block diagram of such a system is shown in Fig. 4. The first advantage is that the designer is free to choose transmitter equipment free of listener constraints. For example a 9,600 b/s link could be used at 25% duty cycle, thus allowing time-division multiplexing. The number of possible users would depend on the channel conditions but the quality of the transmission for each user would not depend on channel conditions. In the extreme case, such a communication system becomes a packet-switching network.

The second advantage of having independent transmission and synthesis rates is to be able to optimize the comprehensibility of the output speech. For example, if the channel turned on and off at a rate of one to ten transitions per second, we might expect the speech distortion to be intolerable. If a problem does exist, it is easily solved by smoothing the output speech rate. By buffering one-half or one second of speech the output speech would be continuous, although delayed. A variable speed-up factor could also be used. For example, the speed-up factor could be 1.5 if the buffer contained more than one second of speech but only 1.2 if it contained less than one second of speech. For two-way conversation, delays in excess of 1/4 second are considered to be very annoying. With such long delays the system might be used in a "push-to-talk" mode where each speaker signals the end of his segment by saying "over" or some such code word. The buffer at the receiver end allows the system designer and possibly even the system user to pick a read-out algorithm which is a compromise between smooth speech and minimum delayed speech. The appropriate compromise may depend on the channel characteristics and a given user's needs.

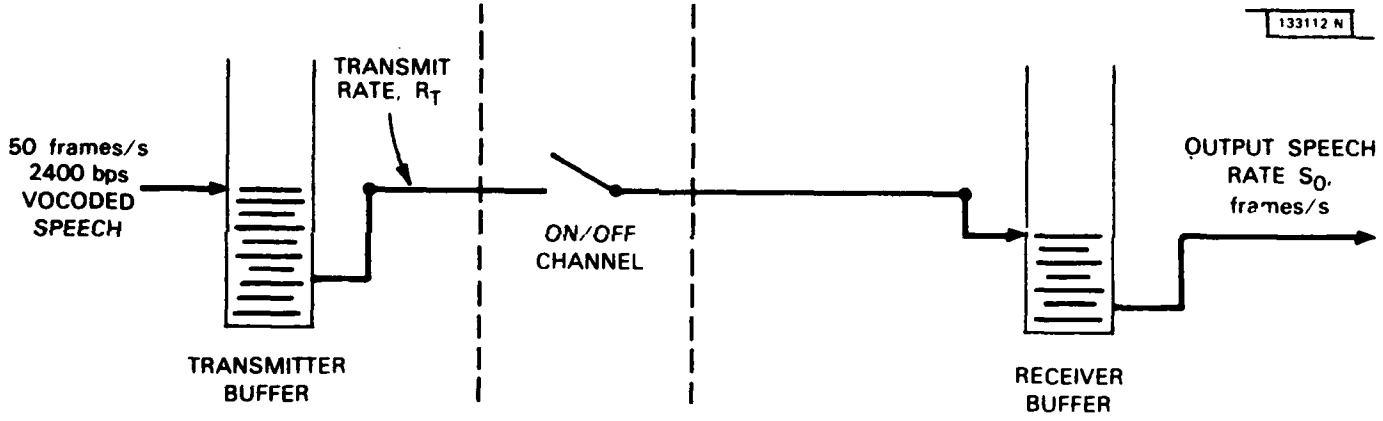


Fig. 4. Block diagram of system with input and output buffers.

As the above discussion indicates, it is easy to design a speech system which works on an intermittant channel. The key issue is how useful is such a speech system. The theoretical aspects of this issue are addressed in the next section as a way to prepare for an empirical investigation of it.

III. COGNITIVE SPEECH PERCEPTION MODEL

The advantage of a theoretical approach to the speech comprehension issue is to allow for generalization of the empirical results. In designing a system, one invariably uses parameters which are different than those used in laboratory test conditions. System design is thus greatly facilitated if the underlying principles are understood. Planning of an experiment is also facilitated if underlying principles are understood. In this study the planning of the experiment presented great difficulty because time and cost constraints prevented the testing of all the parameter values of interest. Four system parameters were identified and a decision to use 5 values for each parameter would have meant using $4^5=1024$ test conditions. Five test conditions were more realistic for the number of subjects and the time available on this project. A theoretical model should allow choosing the five test conditions to maximize the knowledge gained by the experiment.

The theoretical model developed here is based on a variation of a formulation by Miller and Licklider [6] to explain an experimental result. They used a periodic on-off speech signal and measured the intelligibility of the speech as a function of the duty cycle and period of the interruptions. In their experiment the subject heard normal speech, or no speech (silence).

They measured intelligibility by having the subject identify individual spoken words. Miller and Licklider found that for 50% duty cycle and interruption periods over 2 seconds, the subject identified about 50% of the words correctly. This was expected since the subject never heard half of the words. For periods on the order of 0.2 seconds the subjects identified about 90% of the words. This unexpected result was explained by the fact that the subject was able to hear part of every syllable in a word and this was often sufficient to identify the syllables and the words. For very short periods, about 2 ms, the speech became unintelligible because the multiplicative distortion introduced modulation artifacts into the speech spectrum. They also found a dip in the intelligibility curve for periods of about 0.5 seconds. This dip in intelligibility was attributed to an off segment deleting parts of two words (which are about 0.5 second long each) instead of just one word.

Other remotely relevant experimental and theoretical evidence is found in studies of reading and speech comprehension. Marks and Miller [7] have shown that sentences are easier to learn if they obey normal semantic and syntactic constraints. Slamecka [8] found that recognition of word strings was also facilitated by the same constraints. Studies of eye movements during reading have revealed a phenomenum known as "The-Skipping" [9]. Apparently the eye rarely fixates on the word "the". The information of where and what to skip is obtained from the semantic and syntactic knowledge already obtained and from visual information from saccadic eye movements (very rapid eye shifts).

These studies suggest that speech processing is done in chunks and that these chunks probably correspond to phrases in the sentence

structure. It seems plausible that if the eye can skip "the"s, then the ear hears "of the house" or "to the store" as a single linguistic unit.

By combining these ideas, a hypothesis about the comprehensibility of non-continuous sped-up speech can be stated. Expressed crudely as a simple set of rules the hypothesis is as follows:

(1) If the "on" time exceeds 2 seconds, the speech is comprehensible.

(2) If the "off" time is shorter than 1/2 second, the speech is comprehensible.

(3) Otherwise the speech is not comprehensible.

A more realistic version of the hypothesis would make the cut-off more continuous as shown in Fig. 5. The rationale for the "on" time cut-off of 2 seconds is to allow phrases of three to five words to remain intact and hence more easily decoded. The rational for the "off" time cut-off of 1/2 second is that sufficiently short breaks in the speech should be easy to "patch over". If there is an auditory short term memory of about one word in length (1/2 second) then it would allow piecing parts of words or phases back together.

If the phenomenum is continuous (and not binary as stated), then channel characteristics which satisfied both constraints ("on" time greater than 2 seconds, "off" time less than 1/2 second) would be more intelligible than a channel which marginally satisfied only one constraint. If the "on" time exceeds 10 seconds (about 20 words), then entire sentences are heard undistorted and presumably well comprehended regardless of the off time. Similarly, if the "off" time is imperceptably small, we might expect the speech to be perfectly comprehensible as long as the distortion did not get so rapid as to have frequency components in the speech frequency range.

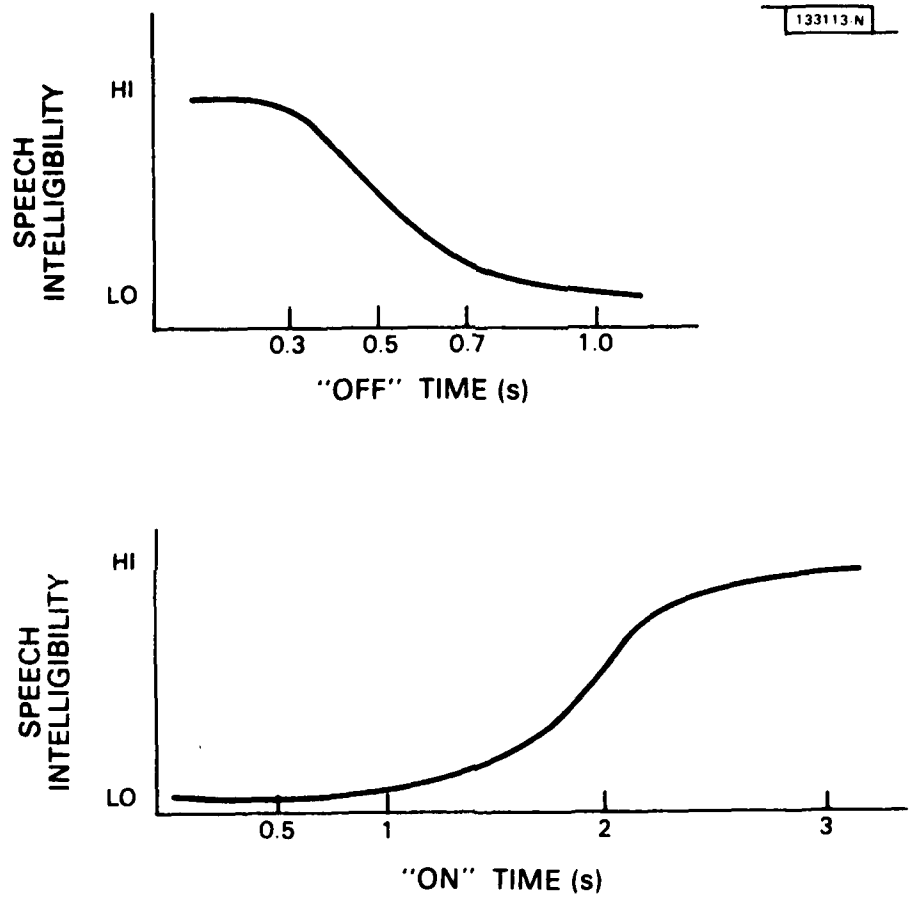


Fig. 5. Hypothesized speech intelligibility as a function of speech "on" time and "off" time.

The proposed model uses "on" and "off" times as the key variables in contrast to duty-cycle and period which were the two key variables in the Miller and Licklider [6] study. The two sets of variables are uniquely related; however, in choosing experimental values for the parameters it is essential to know what ranges are important.

Based on this model the prediction shown in Table I was made. This prediction was written in a notebook prior to conducting the formal quantitative tests, but after informal subjective observations had been made using a manually controlled switch to turn the channel on and off.

The model only gives a qualitative prediction of speech comprehensibiity because general quantitative models of normal speech comprehension are not available, a problem dealt with in the next section.

IV. SPEECH COMPREHENSION TESTING

Intelligibility tests such as the Diagnostic Rhyme Test (DRT) and digital speech quality tests such as the Diagnostic Acceptability Measure (DAM) [10] are not appropriate for testing the sped-up speech technique because the features measured by the tests are not appreciably affected by the interruption in, nor the speeding-up of, the speech. A variety of conversational tests have been developed [11] which typically have two subjects work on a task (such as picture identification) or play a simple game which requires their cooperation and communication. Performance of the communication system under test is assessed by impressions of the subjects during the tasks. These tests appear to provide realistic system assessment but do not provide a quantitative, objective measure which would aid in developing a theoretical model of system performance.

TABLE I

PREDICTED COMPREHENSIBILITY OF NON-CONTINUOUS SPED-UP SPEECH

		"on" time (seconds)	
		1/2	2
"off" time (seconds)	1/2	ok	excellent
	2	bad	ok

For these reasons a speech comprehension test was used in this study. A trained talker read, rather deliberately, paragraphs intended for a reading comprehension test at the eighth grade level [12]. The paragraphs were about 250 words in length and each one took about 2 minutes to read. Paragraphs at such a low level were chosen to insure that the content of the questions would not be challenging to any of the subjects. Even so, we anticipated that individuals might differ in their ability to comprehend spoken information and consequently each subject's response to the test conditions was normalized by their score on continuous normal-speed vocoded speech. A total of 15 paragraphs were recorded. There were 4 test conditions planned plus the normalizing condition so that each test condition was tested with three different paragraphs ((4+1)x3=15).

The subjects listened to a paragraph, then answered three or four written multiple choice questions which came with the test. In addition, the subjects rated the speech quality and comprehensibility of each paragraph. The questions are listed in Table II. The purpose of including the subjective questions was to have a back-up in case of interpretation problems with the objective measure and to compare the sensitivity of the two measures.

No claim is made that the comprehensibility test is adequate for more than making a first-order determination of the usefullness of the sped-up speech algorithm. Perhaps with some development effort the test could be made more useful; however, such an effort is well beyond the scope of

TABLE II

SUBJECTIVE QUESTIONS

A. How comprehensible did you find this last paragraph?

B. How easy was it to listen to?

C. How acceptable did you find the speech?

D. How intelligible did you find the individual words?

this project. The next section contains a description of the channel and vocoder simulation test setup.

V. SIMULATION AND EXPERIMENTAL FACILITY

The Lincoln Digital Signal Processor (LDSP) [13] is a high speed (about 50 n second cycle time) digital computer designed to do real time processing of digital speech signals. The LDSP is used by Lincoln Laboratory to develop and evaluate speech bandwidth compression algorithms. A 2400 b/s channel vocoder (similar to the BELGARD [14] algorithm) was chosen for this study because it required only one of the four available LDSP processors and appeared to be the easiest program to modify to include the HF channel simulation.

The speech speed-up was easy to implement because the number of time samples used to represent each frame of vocoder data is a parameter of the program. Since the sample rate is fixed, changing the number of samples per frame has the affect of changing the frame rate without changing the pitch of the speech. Speeding up the speech by a factor of 1.5 means changing the output frame rate from 50 to 75 frames per second.

The channel simulation could have been done using one of the following three approaches:

(1) Deterministic - channel turns on and off with specified duty cycle and period.

(2) Simple random process model - channel turns on and off following a Poisson, Markov, or similar statistical model.

(3) Physical simulation - channel turns on and off in accordance with a simulated signal-to-noise ratio determined from modeling HF multipath interference and HF modem characteristics.

The deterministic simulation would have provided the most information about speech perception processes but would have left unanswered how human perceptual processes respond to realistic variable conditions. On the other hand, the physical simulation has the opposite bias; it is much more realistic but offers little insight into the underlying perception processes. The simple random process model offers a good compromise and hence was used for the experiments reported here.

The random process model simulation of the channel had two modes. The first used a random process model to determine when the channel was on or off (corresponding to a non-fading or a fading situation). The second mode used a random process model to introduce random errors in the bit stream between the vocoder analysis and synthesis sections. The first mode produced an error-free noncontinuous sped-up speech signal while the second mode produced an error contaminated speech signal such as would be generated by a conventional narrowband speech system in the presence of fading and burst noise.

The simulation of the channel fading and burst noise was based on the following random process model:

(a) The fade or burst noise is defined as an event.

(b) The events occur with Poisson statistics.

(c) The duration of the event is a uniformly distributed random variable.

The channel model used the vocoder frame interval (20 ms) as a unit of measure. If the channel was not in a fade, the probability of starting a

fade at the beginning of the next frame was P_f . The mean number of frames between fades is $1/P_f + E(n)_\ell$ where $E(n)_\ell$ is the average number of frames in a fade and

$$E(n)_\ell = 1/2\ n_{max}$$

where n_{max} is the maximum number of frames in a fade.

The average on and off times (important in interpreting the cognitive model presented earlier) are

$$E(on) = 1/P_f$$

$$E(off) = (1/2)n_{max}$$

where the unit of time is the frame length.

The second random channel mode introduced bit errors into the data stream instead of turning the channel off during a fade or burst noise event. The errors during the burst were independent bit to bit. The probability of error was set to 15% although any BER above 10% made the portion of speech so corrupted totally unintelligible.

The results of feasibility experiments using these channel simulations had only a few surprises as discussed in the next section.

VI. TEST RESULTS

A conventional vocoder with burst errors occurring about 33% of the time produces incomprehensible speech. The statistics of the burst errors do not affect this result if the mean time between bursts is greater than about one second. Only informal, subjective responses were obtained in this mode because the best strategy is to turn the receiver off during the burst which yields a test configuration identical to that used by Miller

and Licklider [6]. If the system blocks out entire words, then recovery of all but the most redundant speech is not possible.

Informal subjective evaluation of the new techniques using sped-up speech transmitted during the channel "on" times showed that the speech was comprehensible under a wide variety of conditions. To make this result quantitative, an experiment was conducted to determine the limits of the technique. It was found that manually turning the channel on and off rather rapidly could sometimes give incomprehensible speech. The cognitive model of speech perception was formulated to understand this informal result and the following test of that model supports it.

Five conditions were tested:

1. perfect vocoder and channel
2. E(on) = 2 s E(off) = 1/2 s
3. E(on) = 2 s E(off) = 2 s
4. E(on) = 1/2 s E(off) = 1/2 s
5. E(on) = 1/2 s E(off) = 2 s

These conditions were identified in the section on the cognitive model as critical to verifying that model.

Fifteen paragraphs were prerecorded and three were used for each of the five test conditions. Nine volunteer subjects, adults between 19 and 55 years of age, were used. They varied in educational background from having completed high school to having completed college. There were four females and five males. Formal selection criteria were not used because the test procedure included normalizing each subjects' test score on the

fading channel with his or her test score on the perfect vocoder and channel (condition 1).

Results on three paragraphs are deleted from subsequent analyses because learning effects were suspected, and because it was discovered that in one paragraph the wrong test parameters had been used.

The mean and standard deviation of the scores averaged over the nine subjects is shown in Table III. Also shown are the cognitive model predictions. The mean of the perfect channel scores was 99.7% (it is not 100% because of round-off in the calculation). The standard deviation of the perfect channel scores was 33.4%. Two observations are immediate; the quantitative results are in complete accordance with the qualitative prediction of the model, and the standard deviation of the results, including the base-line condition, is very high. The test results are significant because the predicted "Excellent" condition (E(off) = 1/2, E(on) = 2) score is about one standard deviation above the predicted "OK" conditions (E(off) = 1/2 E(on) = 1/2, E(off) = 2, E(on) = 2) scores, while the predicted "Bad" condition (E(off) = 2, E(on) = 1/2) is about two standard deviations below the predicted "OK" condition score.

The high standard deviation of the base-line test scores (perfect channel) was not expected. It is possible that the variability of the other test conditions could be attributed to the test instrument itself rather than the inherent variability in the comprehensibility of the distorted speech. This high variability of the scores from a perfect channel could be due to several different factors:

TABLE III

EXPERIMENTAL RESULTS SHOWING THE MEAN AND STANDARD DEVIATION OF SPEECH COMPREHENSION TEST

(predicted results are shown in square brackets; standard deviation shown in parenthesis)

E(OFF) \ E(ON)	1/2	2
1/2	4 55 (49.3) [OK]	2 102 (38.8) [Excellent]
2	5 0 (0) [Bad]	3 68 (33.5) [OK]

(1) Listening to a passage could be much more difficult than reading a passage.

(2) The subjects were too unskilled in normal reading comprehension.

(3) The subjects were distracted by the distortion of the vocoded speech and would have done better had they had time to accomodate to it.

(4) The channel vocoder itself produces reduced quality speech and limits comprehension. It has DRT scores of about 85-90%.

Sorting these and possibly other issues should be done if the test of listening comprehension is to be used in a more sensitive discrimination of speech distortion. Fortunately, the test was perfectly adequate for the purposes of the present investigation.

It was previously noted that a speed-up factor of 1.5 did not detract from the comprehension of the speech. This claim is supported by the fact that the score for the "Excellent" condition is comparable to the reference (perfect channel) condition. Naturally, there could be some degradation, but a much better vocoder and a much better test of comprehension would be required to reveal it.

The last observation about the data is that the predicted "bad" condition (E(off) = 2 s, E(on) = 1/2 s) is really bad! Not one subject got a single question about the passage <u>correct</u>. The degree of this result was not expected. It points out a striking limitation of the human speech processing apparatus and reaffirms the value of an objective measure of comprehension, even one which produces results with a large standard deviation. The subjective test results are now presented and compared to the above objective results.

The mean and standard deviation of each question for the reference condition (condition #5, perfect channel) is shown in Table IV. The rating of the comprehensibility of the speech (question A) for the reference condition is high (4.72) but is not perfect. To take individual differences into account, the subjective responses to the four test conditions were normalized. Each subject's score was normalized by his or her score on the reference condition. The normalization procedure in this case was to add an amount to each subject's scores to make the reference score equal to five. Another way of expressing the procedure is to note that it is the algebraic difference between the score for the test condition and the score for the reference condition which is used in the analysis. This, or any other normalization procedure can cause interpretation difficulties. We observed uniformly low scores of condition 4 (the predicted "bad" condition). This condition received the bottom rating on questions A, B, and C for almost every paragraph and every subject. The normalization procedure will artificially raise some of these scores, and hence the clean-cut conclusion that condition 4 is really bad would be softened if only the normalized scores were inspected. None of the subjects answered any of the objective questions on the test paragraphs correctly; thus the subjective and objective test results are in complete agreement (and in agreement with the model).

The result of the normalization process is shown in Table V, together with the mean (μ) and standard deviation (σ) for each question and each test condition averaged over 18 or 27 paragraphs. The differences in the

TABLE IV

MEAN AND STANDARD DEVIATION FOR SUBJECTIVE RESPONSES TO REFERENCES CONDITION

Question	Mean μ	Standard Deviation σ
A	4.72	0.55
B	4.55	0.68
C	4.44	0.68
D	4.61	0.67

TABLE V

EXPERIMENTAL RESULTS SHOWING MEAN AND STANDARD DEVIATION ON SUBJECTIVE EVALUATION QUESTIONS

		E(on) 1/2 S	E(on) 2 S
E(off)	1/2 S	4 1.87 (0.62) [OK]	2 3.33 (1.16) [Excellent]
	2 S	5 1.46 (0.60) [Bad]	3 2.87 (0.82) [OK]

scores for questions A, B, and C which referred to the speech comprehensibility, ease of listening to, and acceptability are small. Apparently these questions tap equivalent subjective attitudes and hence only the comprehensibility (question A) scores will be used in the data analysis and interpretation that follows.

Question D asked about the intelligibility of individual words. Consequently, at least in principle, the question taps a different aspect of speech perception. The intelligibility of individual words is, however, almost identical in condition 1 and 2 because the "on" time is 2 seconds for both conditions. Only the "off" time is different for conditions 1 and 2. Since typical words are only 1/2 second long, most of them would be heard intact in both cases. But there is a difference in the subjects reported intelligibility of individual words. The author's own informal response agrees with the subject's response: it sounds as if word intelligibility degrades as the "off" time increases. Hence, question D, while provocative, is not used in the data analysis.

The subjective results are summarized in Table VI where the response to the comprehensibility question (A) is compared to the objective measure of comprehension (O) and the predicted result. The most important observation is that the subjective test rank orders the conditions in exactly the same way as the objective test and agrees with the prediction of the model. Hence, the subjective measure validates the model and the objective test procedure and the conclusion of the experiment is strengthened.

TABLE VI

COMPARISON OF RESULTS OF SUBJECTIVE QUESTIONS ABOUT VOCODER/CHANNEL COMPREHENSIBILITY WITH OBJECTIVE RESULTS

		E(ON) 1/2 s	E(ON) 2 s
E(OFF)	1/2 s	4 A: 1.83 O: 55 [OK]	2 A: 3.33 O: 102 [Excellent]
	2 s	5 A: 1.46 O: 0 [Bad]	3 A: 2.87 O: 68 [OK]

Note: Objective test scores are normalized to 100
Subjective test scores are normalized to 5

A is score on subjective question A

O is objective test score

excellent, ok, bad refer to the predicted performance

condition number shown in left top corner

The data analysis thus far has ben reported in terms of means and variances. In order to determine the statistical significance of these results a standard analysis of variance (ANOVA) was conducted, the results of which are shown in Table VII. With one exception the means which appeared to be significantly different are indeed so. The exception is the comparison of condition 4 and 5 where significance at the P=0.05 level was not quite attained.

There are, however, important differences in the subjective and objective test results. The first is that the predicted "Excellent" condition (E(on) = 2 s, E(off) = 1/2 s) has a subjective rating of only 3.3 which is well below 5.0 for the reference while its objective score is 102 which is essentially equal to the 100 for the reference. Apparently the subjects thought they were having some difficulty comprehending the speech, but in fact, were not. One might be tempted to conclude that the subjects don't know what they're talking about and that subjective tests were not valid.

Another interpretation is possible, however. Perhaps the subjects were responding to an aspect of speech comprehension not adequately accounted for in the objective measure. Perhaps some passages are more difficult to comprehend than others. If greater mental effort is required to process more difficult passages, then comprehension of those passages would degrade if some mental processing were allocated to deciphering distorted speech. The passages used in this study were taken from eighth grade reading tests and hence might only be difficult to comprehend if the speech were terribly distorted.

TABLE VII

PAIRWISE ANALYSIS OF VARIANCE (ANOVA)

	Objective			Subjective		
Condition Pair	F	Prob.	Significant	F	Prob.	Significant
2 - 3	9.69	.0033	yes	2.33	.1346	no
3 - 4	1.01	.3197	no	19.74	.0001	yes
4 - 5	36.21	.0000	yes	3.76	.0592	almost

Note:

Condition	E(on)	E(off)
2	2 S	1/2 S
3	2 S	2 S
4	1/2 S	1/2 S
5	1/2 S	2 S

The subjective data forces the conclusion that the predicted "Excellent" condition (E(on)=2 s,E(off)=1/2 s) may in fact only be "very good." Perhaps E(on) needs to be as long as 3 or 4 seconds to produce excellent results or E(off) needs to be as short as 0.3 to 0.4 seconds.

The second important difference between the objective and subjective test results is for the two predicted "OK" conditions (2: E(on)=2 s, E(off) =2 s; 3: E(on)=1/2 s, E(off)=1/2 s). The objective results for condition 2 and 3 are not statically significant (P=.319). While the subjective score for condition 2 is more than a standard deviation above the score for condition 3 (is statistically significant at the .0001 level). In fact, condition 3 is almost as bad subjectively as condition 4 (the very bad case). That the parameters of the model would need adjustment was to be expected, that the objective and subjective scores would be so different was not expected. It may be that the comprehension of speech is a complex combination of cognitive and perceptual processes and that the attempts to understand and quantify them must use more sophisticated constructs and instruments than were used in the study. Furthermore, introspection and self assessment of comprehension should be used carefully. Fortunately, the model and test methods used here were completely adequate to demonstrate the feasibility of using non-continuous, sped-up speech and to suggest practical limitations of the technique. The conclusion section of the report discusses these practical interpretations of the experimental results.

VII. SUMMARY AND CONCLUSIONS

The motivating problem for this study was to achieve reliable digital voice communications on HF channels. These channels suffer from random

fading and burst noise due to ionospheric propagation effects. The performance of vocoders operating with conventional HF modems degrades rapidly under these channel conditions. Complex and expensive state-of-the-art modems have been developed which greatly improve the HF communication link by using combinations of channel distortion compensation and error correction. Even these sophisticated techniques are vulnerable under long-lasting, low signal-to-noise conditions. An alternative approach was therefore pursued which might have application in various systems. The technique proposed was to transmit the digital speech only during the times the channel provided a low-bit error rate. To allow two-way interactive conversation, it was necessary to speed-up the transmission and the synthesis of the speech. The speed-up factor was chosen to be low enough (1.5) so that the speech would remain entirely intelligible. The speech would be broken up, however, coming in random bursts separated by silences. Although informal listening to speech transmitted in this manner seemed perfectly comprehensible in most cases, albeit somewhat disconcerting, it was found that rapid changing from an "on" to an "off" condition reduced comprehension. Thus the study of vocoded speech on HF channels led to a study of perception and cognition of non-continuous sped-up speech. If the human perceptual-cognitive system could tolerate this unusual distortion, then the utility of HF communication systems and other similar variable systems such as meteor-scatter links could be greatly increased.

Understanding the perception of speech is an active research area but one well beyond the scope and intent of the present study. Nevertheless,

it was essential to arrive at some elementary understanding of comprehension of spoken sentances and paragraphs in order to evaluate the proposed technique.

The required elementary understanding was achieved by constructing a simple cognitive model of speech comprehension of non-continuous speech and then testing the model with objective and subjective measures using a set of well chosen test conditions. Such testing is always burdensome and expensive because many subjects and trials must be used. Thus the test program scope was limited.

The constructed model was based on the ideas that long continuous intervals of speech and segments separated by short interruptions are comprehensible. If either criterion was solidly satisfied, then the speech comprehension should be excellent. If both criteria are not satisfied, then the speech should be imcomprehensible. In marginal cases, satisfaction of both criteria should produce more comprehensible speech than satisfying one of them. Given this structure, the model needs to have two parameters specified. How long is a long "on" interval, and how short is a short "off" interval. Of course a single answer may not exist; different individuals in different situations may vary. This complexity was ignored in this study, partly because the study is exploratory and partly because of the need for the speech system to be useful to most people most of the time. Arguing mostly heuristically, with reference to related studies of speech and reading comprehension, the hypothesis was put forward that a 2 second or longer "on" interval was sufficient to provide comprehensibility and that a 1/2 second or shorter "off" interval was

sufficient. The two parameters, each with two values, gave four test conditions which were augmented by a reference (perfect channel) condition giving five test conditions for the experiment. The test results, taken as a whole, confirm both the structure and parameter values of the model. Nine subjects, listening to two or three two-minute, reading-comprehension-test passages for each test condition, answered objective questions about the passage content and answered subjective questions about the speech comprehensibility. Although the data (18 or 27 data points for each test condition) had high standard deviation, there were significant differences in the test results for the various conditions (see Table VI for key results). In fact, the data proved to be sensitive enough to show interesting differences in the test results for the objective and subjective measures. These differences led to the conclusion that a much more complex perception model and a much more sensitive objective measure of comprehension would have to be developed before the subtleties could be unraveled. Fortunately, implications for practical HF communication systems can be drawn from the presented data and the first-order cognitive model.

Under quiet ionospheric conditions, a typical fade of an HF signal was previously defined to be about 1 to 4 seconds in duration and to occur at intervals of 10 to 40 seconds. The cognitive model has the rule that speech with "on" times exceeding 2 seconds will be comprehensible. The data suggested that the transition between comprehensibility and incomprehensibility might not be very sharp and may occur at 3 or even 4 seconds. In any case, the intervals between fades of the HF channel

comfortably exceed any of these critical time intervals. The model predicts that for such long "on" times, the "off" time is irrelevant. Hence it is concluded that the non-continuous sped-up speech will be completely comprehensible during quiet ionospheric conditions despite deep fades which would render conventional systems unusable.

Under disturbed ionospheric conditions, the typical fade of an HF signal was previously defined to be about 0.1 to 0.4 seconds in duration and to occur at intervals of 1 to 4 seconds. The cognitive model has the rule that speech with "off" times shorter than 1/2 second will be comprehensible. The data suggested, however, that perhaps this critical time interval was somewhat shorter (e.g., 0.3 to 0.4 seconds). In either case, the fade duration of the HF channel marginally satisfies the "off" time criteria. Based on this comparison of the channel and the perception model, it is concluded that the non-continuous sped-up speech will be comprehensible during disturbed ionospheric conditions despite frequent deep fades which would render conventional and perhaps even state-of-the-art digital speech/modem systems unusable.

For ionospheric conditions, intermediate between quiet and disturbed, channel useability depends on the ratio of fade duration to fade period (time between fades) and on the ratio of the comprehension model "on" time and "off" time criterion. The ratio used for the channel is 1:10 while the ratio for the model was 1:4 (1/2 seconds to 2 seconds). Under these conditions, one of the model's criteria must be satisfied. If however, the ratio for the channel drops to 1:5 and the perceptual model were revised to require "off" time shorter than 0.3 seconds or "on" time longer than 3

seconds, then the model ratio would be 1:10 and there would be channel conditions not satisfying the model criteria and hence less comprehensible speech would result. Even in this worst case, however, the speech would be partially comprehensible.

The conclusion is thus supported that the sped-up speech technique provides comprehensible digital voice communication for almost all HF propagation conditions. To clarify the "almost all" caveat will require extensive modeling and testing of both HF propagation phenomena and speech cognition and comprehension processes.

Recommendations for further research and development follow directly from this conclusion. To determine the utility and limitations of the non-continuous sped-up speech technique, a much more extensive experimental program needs to be conducted. This program would have a research phase and an equipment development and field test phase.

The research phase would be to obtain HF propagation data (not statistics) suitable for use with an LDSP implemented vocoder. This data should be obtained with several different HF modems, conventional and advanced design, and should be taken under a range of ionospheric conditions. It is critical that a relationship be established between the required availability - demands of the intended operational system (e.g., 80% versus 99.9% operability) and the expected state of the ionospheric disturbances (e.g., 1990-1991 will have unusually enhanced solar activity and hence produce very disturbed propagation conditions). In these experiments the measurements of comprehension were influenced by the intrinsic intelligibility loss through the narrowband vocoder. In order to

properly separate the effects of the vocoder from those of the channel behavior it might be better to use a higher quality speech digitizer such as 9.6 kb/s APC. Further research should also include theoretical and experimental studies of speech comprehension so that acceptability criteria can be established.

The equipment development and field test phase would be to develop a microprocessor based modem and vocoder which have the required buffers, speed-up capability, and channel feedback link. These units could then be field tested with existing HF communications links, either operational or experimental.

References

[1] Kenneth Davies, "Ionosperic Radio Propagation," National Bureau of Standards Monograph 80 (1965).

[2] W. M. Jewett and R. Cole, "Modulation and Coding Study for the Advanced Narrowband Digital Voice Terminal," NRL Memorandum Report No. 3811 (August 1978).

[3] D. D. McRae, R. S. LeFever, J. N. York, "HF Modem Feasibility Demonstration," Harris Corporation, RADC-TR-82-48 Final Report (March 1982), DTIC AD-A115433.

[4] V. Ellins, P. H. Anderson, W. Sandler, "HF Wideband Modem," GTE Sylvania, RADC-TR-82-85 Final Technical Report (April 1982), DTIC AD-A116859.

[5] E. Foulke and T. G. Sticht, "Review of Research on the Intelligibility and Comprehension of Accelerated Speech," Psychological Bulletin 1969 72, 50 (1969).

[6] G. A. Miller and J. C. R. Licklider, "The Intelligibility of Interrupted Speech," J. Acoust. Soc. Am. 22, 167 (1950).

[7] L. G. Marks and G. A. Miller, "The Role of Semantic and Syntactic Constraints in the Memorization of English Sentences," J. Verbal Learning and Verbal Behavior 3, 1 (1964).

[8] N. J. Slamecka, "Recognition of Word Strings as a Function of Linguistic Violations," J. of Experimental Psychology 79, 377 (1969).

[9] Kevin O'Regan, "Moment to Moment Control of Eye Saccades as a Function of Textual Parameters in Reading," in Processing of Visible Language, Vol. 1, P. A. Kolers, M. E. Wrolstad, and H. Bouma, Eds. (Plenum Press, NY, 1979).

[10] W. D. Voiers, "Diagnostic Evaluation of Speech Intelligibility," in Speech Intelligibility and Speaker Recognition, Mones E. Hawley, Ed. (Dowden, Hutchinson and Ross, Inc., Stroudsburg, PA, 1977).

[11] A. Schmidt-Nielsen and S. S. Everett, "A Conventional Test for Comparing Voice Systems Using Working Two-way Communication Links, IEEE Trans. Acoust., Speech, and Signal Processing ASSP-30, 853 (1982).

[12] M. Peters, J. Coleman, J. Shostar, D. Gunsher, Barron's How-to-Prepare for High School Entrance Examinations (Barron Educational Series, Inc., Great Neck, New York, 1961).

[13] The Lincoln Digital Signal Processor is an advanced version of the LDVT. See P. E. Blankenship et al., "The Lincoln Digital Voice Terminal System," Technical Note 1975-53, Lincoln Laboratory, M.I.T. (25 August 1975), DDC AD-A017569/5; or P. E. Blankenship, "LDVT: High Performance Minicomputer for Real-Time Speech Processing," EASCON '73 Record, p. 214a-214g.

[14] J. N. Holmes, "The JSRU Channel Vocoder," Proc. IEE 127, 53 (1980).

APPENDIX I

RAW DATA ON OBJECTIVE QUESTIONS FOR EACH OF 15 PARAGRAPHS AND 9 SUBJECTS AND 5 TEST CONDITIONS

Paragraph No.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Condition		2	4	1	3	2	5	4	1	2	3	4	5	1	3	5
Subject	1	3	1	2	3	3	0	1	2	4	4	0	0	3	3	0
	2	3	0	2	3	2	0	1	1	2	2	2	0	1	1	0
	3	2	1	2	2	1	0	2	3	3	1	2	0	2	1	0
	4	2	0	2	2	1	0	1	2	2	2	0	0	3	1	0
	5	3	0	1	2	2	0	0	0	4	2	1	0	4	1	0
	6	4	0	3	4	3	0	1	4	4	2	1	0	3	2	0
	7	3	1	2	2	2	0	3	2	4	0	0	0	4	1	0
	8	2	0	1	1	3	0	2	1	4	2	2	0	3	1	0
	9	3	0	1	2	2	0	3	4	2	1	0	0	5	2	0
Number of Questions Asked		4	4	3	5	4	4	4	5	4	5	3	4	4	3	3

APPENDIX II

RAW DATA ON SUBJECTIVE RESPONSES FOR EACH OF 15 PARAGRAPHS, 9 SUBJECTS, 4 QUESTIONS (A, B, C, D)

		SUBJECT NO.								
		2	1	3	4	4	6	7	8	9
	A	4	4	4	4	2	4	2	3	3
	B	3	4	3	4	2	4	1	2	2
(1)	C	3	4	4	2	2	3	3	2	2
	D	4	5	4	5	2	5	4	4	3
	A	1	2	1	1	1	1	1	1	1
	B	1	2	1	1	1	1	1	1	1
(3)	C	1	2	1	1	1	1	1	1	1
	D	1	3	1	1	2	3	2	2	4
3)	A	4	5	5	5	4	5	5	5	3
	B	4	4	5	5	4	4	5	5	3
(5)	C	4	5	5	4	4	4	5	5	3
	D	5	5	5	5	4	5	5	5	3
4)	A	3	4	4	3	2	3	2	3	2
	B	2	4	3	3	2	3	2	2	2
(2)	C	2	4	2	3	2	3	3	2	2
	D	3	4	3	4	2	4	4	3	4
5)	A	4	5	1	2	3	3	3	2	2
	B	3	5	2	1	3	3	3	2	2
(1)	C	3	3	2	2	3	3	3	2	2
	D	4	5	2	1	4	4	4	3	3
6)	A	1	1	1	1	1	1	1	1	1
	B	1	1	1	1	1	1	1	1	1
(4)	C	1	1	1	1	1	1	1	1	1
	D	1	2	1	1	2	3	1	1	1
7)	A	2	3	1	2	2	2	2	2	1
	B	1	3	2	1	1	2	2	2	2
(3)	C	1	3	2	1	1	2	1	2	1
	D	2	3	2	1	1	4	2	2	3
	A	3	5	4	4	3	5	4	2	3
	B	2	3	4	1	3	3	4	3	3
(5)	C	1	3	4	2	3	3	4	2	3
	D	4	5	4	5	2	5	5	2	3
9)	A	4	4	3	3	4	4	4	4	2
	B	2	5	3	1	4	3	3	4	2
(1)	C	2	4	3	3	4	3	2	4	2
	D	4	5	4	5	4	4	3	4	4

		SUBJECT								
		2	1	3	4	5	6	7	8	9
10)	A	2	2	1	2	4	3	1	3	2
	B	2	3	2	1	2	2	1	3	1
(2)	C	2	2	2	2	2	2	1	3	2
	D	2	3	2	3	2	4	1	4	3
11)	A	1	1	2	1	1	1	1	2	1
	B	1	1	2	1	1	1	1	2	1
(3)	C	1	1	2	1	1	1	1	1	1
	D	1	2	2	1	1	2	2	1	1
12)	A	1	2	1	3	1	2	1	1	1
	B	1	1	1	1	1	1	1	1	1
(4)	C	1	1	1	2	1	1	2	1	1
	D	2	2	3	2	1	2	2	2	3
13)	A	4	5	5	5	5	5	5	5	5
	B	3	5	5	5	5	5	5	5	5
(5)	C	3	5	5	5	4	5	5	5	4
	D	3	5	5	5	4	5	5	5	4
14)	A	2	3	3	3	2	3	2	3	3
	B	2	4	3	1	2	3	3	2	3
(2)	C	2	3	2	2	2	2	3	3	3
	D	2	4	3	4	2	3	4	3	3
15)	A	2	1	1	1	1	1	1	1	1
	B	1	2	1	1	1	1	1	1	1
(4)	C	1	1	1	1	1	1	1	1	1
	D	1	1	2	4	1	1	1	1	1

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ESD-TR-83-230	2. GOVT ACCESSION NO. A138650	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Evaluation of the Comprehension of Non-Continuous, Sped-Up Vocoded Speech: A Strategy for Coping with Fading HF Channels		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER Technical Report 671
7. AUTHOR*(s)* John T. Lynch, Bernard Gold, Joseph Tierney and Carol J. Bowers		8. CONTRACT OR GRANT NUMBER*(s)* F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173-0073		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 33401F Project No. 7820
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Sytems Command, USAF Andrews AFB Washington, DC 20331		12. REPORT DATE 5 January 1984 13. NUMBER OF PAGES 56
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Electronic Systems Division Hanscom AFB, MA 01731		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

digital voice | HF communications | speech comprehension testing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A novel technique for coping with fading and burst noise on HF channels used for digital voice communication has been developed and evaluated. The technique transmits digital voice only during the high signal-to-noise ratio time intervals, i.e., channel "on" times, and speeds up the speech when necessary in order to avoid delays which would hinder conversation. The technique was evaluated using a model of the human speech comprehension process, which was tested using a spoken version of a reading comprehension test. The test involved fifteen spoken, two-minute paragraphs processed by a real-time channel vocoder simulation which had been modified to also simulate the on/off characteristics of a fading HF channel. Using a speed-up factor of 1.5, the percentage of correct test responses verified the comprehension model. If the average "on" time is longer than about two seconds or the average channel "off" time is shorter than about one-half second, then the speech is comprehensible. Since these conditions are met for most disturbed and undisturbed ionospheric conditions, it is concluded that the sped-up speech technique is appropriate for HF digital voice communication systems.

DD FORM 1 Jan 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

END

DATE
FILMED

4 - 84

DTIC