JOINT REPORT

NAVAL UNDERWATER SOUND LABORATORY
NAVAL SUBMARINE MEDICAL CENTER

Report Number
574

SPEECH INTELLIGIBILITY IN A STATIONARY MULTIPATH CHANNEL

by

Murray B. Sachs, Gary J. O'Brien, Russell L. Sergeant

and

Joseph S. Russotti

Bureau of Medicine and Surgery, Navy Department
Research Work Unit MF12.524.004-9011D.03
and
NavShips Systems Command Task 5-23-18-08622
(USL-9-A-611-03-00)

Released by:

Gerald J. Duffner, CAPT MC USN
COMMANDING OFFICER
Naval Submarine Medical Center

21 March 1969

SPEECH INTELLIGIBILITY IN A STATIONARY MULTIPATH CHANNEL

by

Murray B. Sachs and Gary J. O'Brien
Navy Underwater Sound Laboratory
New London, Connecticut

and

Russell L. Sergeant and Joseph S. Russotti
Naval Submarine Medical Center
Groton, Connecticut

SUBMARINE MEDICAL RESEARCH LABORATORY
U. S. NAVAL SUBMARINE MEDICAL CENTER REPORT NO. 574

Bureau of Medicine and Surgery, Navy Department
Research Work Unit MF12.524.004-9011D.03

This is a Joint Report by personnel of the Navy Underwater Sound Laboratory and the Submarine Medical Research Laboratory. It was approved for publication on 21 March 1969 and has been designated as SubMedResLab Report No. 574, with Distribution Statement No. 1 assigned, indicating that it has been approved for public release and sale; its distribution is to be unlimited.

The NavShips Systems Command Task No. is 5-23-18-08622 (USL 9-A-611-03-00).

ABSTRACT

The reception of speech transmitted through an acoustic channel such as the ocean is limited by multipath "time-smearing." The purpose of this study was to obtain a quantitative measure of the effects of such time-smearing on speech intelligibility. A linear, time-invariant channel was used as a model of the ocean. The impulse response of this channel was a sample of band-limited Gaussian noise. Using Fast Fourier Transform techniques, words of the Modified Rhyme Test were convolved with, or smeared in the time domain, by this channel impulse response. The intelligibility of these "smeared" words was measured as a function of the impulse response duration, T. Intelligibility decreased monotonically to about 75 percent as T increased to 200 milliseconds. Further increase in T did not significantly lower intelligibility. Distortions in time evaluated herein did not impose serious limitations to the reception of short words. However, a detailed analysis of consonantal errors revealed that sounds occurring in the middle of a word are much harder to hear correctly than are sounds at the beginning or ending of an utterance. We conclude that time-smearing will more seriously interfere with the intelligibility of connected or conversational speech.

SPEECH INTELLIGIBILITY IN A STATIONARY MULTIPATH CHANNEL

I. INTRODUCTION

Acoustic signals undergo multipath distortions in transmission through the ocean. The present study concerns the effect of these distortions upon a speech signal.

A source of acoustic energy located near the surface of the ocean, and a receiver also near the surface but some distance removed are depicted in Figure 1. Sound energy travels from source to receiver over one or more of a variety of possible propagation paths. Since the paths differ in their length, and thus, in the time required for energy to travel over them, a single burst of energy from the source arrives at the receiver as a sequence of two or more distinct bursts. These different paths are called gross paths and the energy arriving over them are called gross arrivals.

For a particular gross arrival over any one of the paths, the signal may be further distorted. Since the ocean's bottom is not smooth., energy is reflected by many different facets within that area ensonified by the source. Thus, a single burst of energy, after traversing any gross path, will arrive at the receiver as many overlapping bursts. A short pulse is consequently "smeared out" in time.

Figure 2 shows the envelope of a received signal for a pulse transmitted from source to receiver in a situation similar to that described in Figure 1. Note the three gross arrivals delayed successively by about ½ sec. These arrivals correspond to a direct, a single bottom, and a double bottom bounce-path. The transmitted signal was a single 16-msec pulse. Note also that over the two bottom bounce-paths the pulse was smeared out to durations greater than 100 msec.

Two problems arise when transmitting speech over a multipath channel such as the one described in Figure 1: first, every word arrives at the receiver via all three paths to produce three repetitions of each word, separated from each other by about ½ sec. Echoes such as these make continuous speech quite unintelligible. Second, speech received even over a single path is smeared out in time. In fact, in the absence of gross echoes, speech so smeared may or may not be fairly intelligible depending on the amount of smearing.

II. MAJOR PURPOSE

This study investigated the effect upon speech intelligibility of time smearing in one path only. The maximum effect on intelligibility was found to be a reduction to 75%. Interpretation of some of the preliminary results led to a sub-experiment, that of investigating the effect of consonantal position within a word upon a phoneme's intelligibility in smeared speech.

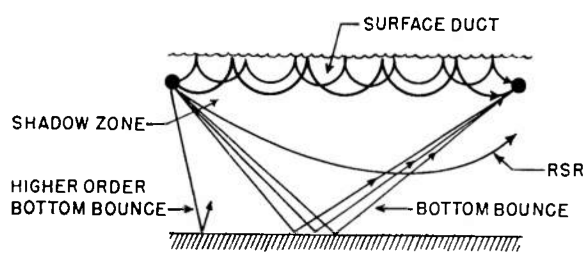


Fig. 1. Dominant undersea acoustic paths.

III. PROCEDURES

Because experiments at sea are costly, the ocean's acoustic paths were simulated on a computer as a linear time-invariant channel. As shown in Figure 3, such a channel is completely characterized by its impulse response since, for any transmitted signal s(t), the received signal y(t) is the convolution of s(t), with the impulse response h(t). Also, the Fourier transform of the received signal y(f), is the product of the Fourier transforms of the impulse response and the transmitted signal. Stockham[1] has shown that with "Fast Fourier Transform (FFT) Techniques," it is often faster to compute the Fourier transforms X(f) and H(F), multiply them together, and find y(t) as the inverse transform of this product than it is to carry out the convolutions in the time domain. Using a digital computer we have simulated the channel impulse response h(t) and used an FFT convolution technique to convolve our simulated channel impulse response with speech for use in testing the intelligibility of speech.

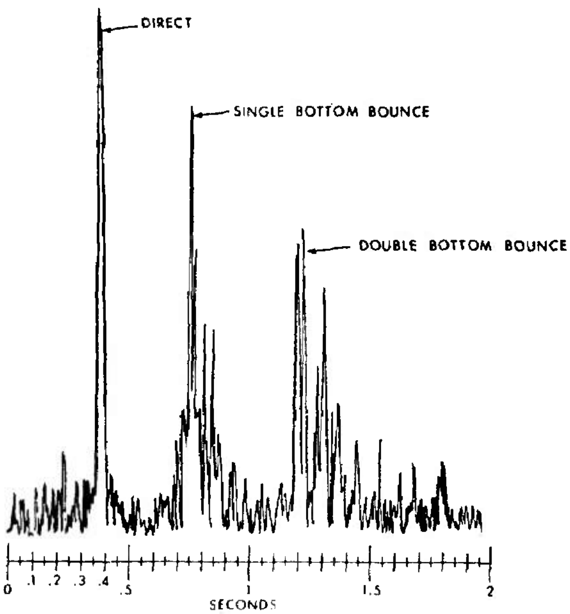
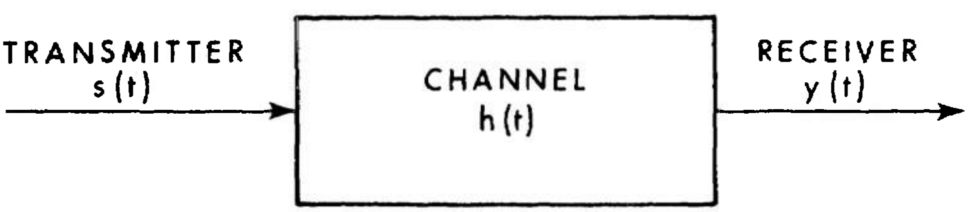


Fig. 2. Envelope of an underwater signal that has three different paths of arrival.

IV. RESULTS AND DISCUSSION

A. **Description of the Channel Impulse Response.** A typical channel impulse response used in the simulation is shown in Figure 4. Remember, that our model represents only a single gross ocean path. The response h(t) is a sample function from Gaussian noise with a flat spectrum limited to the range 0 to 5 kHz. Measurements using panels of listeners were made of the intelligibility of speech passed through a channel with such an impulse response as a function of the duration of the noise sample. In Figure 4(a) we show a sample h(t) which has a duration of 95 msec. In Figure 4(b) the logarithm of the magnitude squared of the Fourier transform of h(t) is plotted. In this, as in all cases, the acoustic spectrum of the sample function has a noise-like appearance.

The squared magnitude of the Fourier transform of a typical word ("sold") is shown in Figure 5(a). The same transform multiplied by the Fourier Transform of the 95-msec impulse response is shown in Figure 5(b). The general shapes of the two spectra are clearly similar, and the effect of multiplying by the channel transfer function, H(f), is simply to add noise in the frequency domain to the spectrum of the word. Thus, the long time magnitude spectrum of a word is not significantly distorted by passage through this single gross multipath channel, even though the phase spectrum is completely randomized.



$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, s(t-\tau)\, d\tau$$

$$Y(2\pi f) = H(2\pi f)\, S(2\pi f)$$

Fig. 3. Model of linear time-invariant channel used to simulate an acoustic path in the ocean.

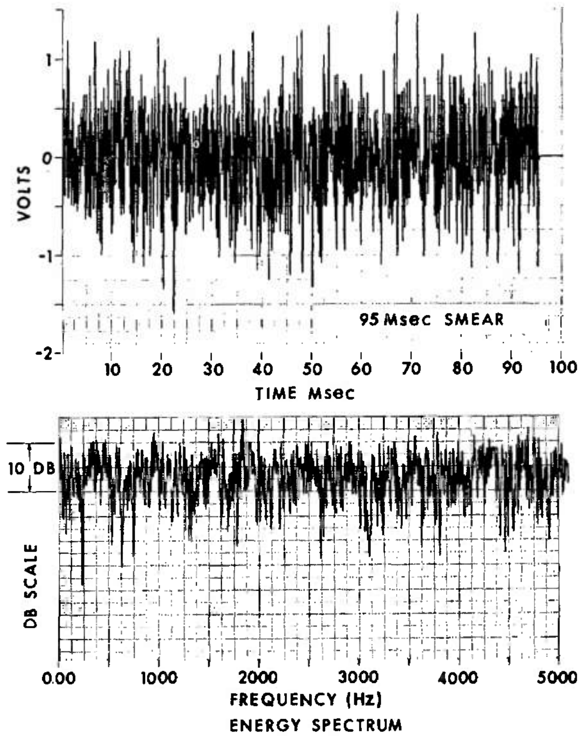


Fig. 4. Typical channel impulse response, h(t), used to simulate an acoustic path in the ocean. In the upper figure an h(t) with a duration of 95 milliseconds is plotted. The lower graph is the logarithm of the magnitude squared of the Fourier transform of h(t).

The effect of time smearing on speech can be examined in time-frequency plots as in Figure 6. Spectrographic recordings were made of the word "late", both unsmeared and convolved with 95, 285, and 570 msec impulse responses. The most noticeable effect of the process of convolution as shown in Figure 6 is the lengthening of the word by the duration of the impulse response. The time-structure associated with the speaker's pitch periodicity is also smeared out in the lower three spectograms. Notice, however, that some important features of the spectograph, although stretched in time, are maintained. For example, the initial transition from the "l" to the "a" sounds is present even for the 570 msec case. However, certain other characteristics are destroyed by the longer smears. For Figure 7 we increased the gain of our spectographic recording to bring out the stop consonant "t". The "t" appears in the unsmeared case as a burst of energy clearly separated from the "a" by about 100 msec. When smeared 47.5 msec, there is a shorter but still distinct separation between the "a" and "t". However, when the impulse response duration is as great as 285 msec, the "a" becomes smeared into the "t", and no distinct separation appears. The 570 msec condition is even worse. Thus, we expect some phonemes, like vowels and semi-vowels, to survive even the longest smears we have used, while other phonemes, like stop consonants, which depend upon bursts of energy for recognition, may be destroyed by longer smears.

B. **Intelligibility of Speech Convolved with Channel Impulse Response.** Words of the Modified Rhyme Test[2] were convolved with impulse responses whose durations ranged from 0 to 665 msec. Then the material was presented to various groups of normal-hearing Naval enlisted men in a group audio-testing room which contains 50 matched monaural headsets. The men had no previous experience as experimental listeners.

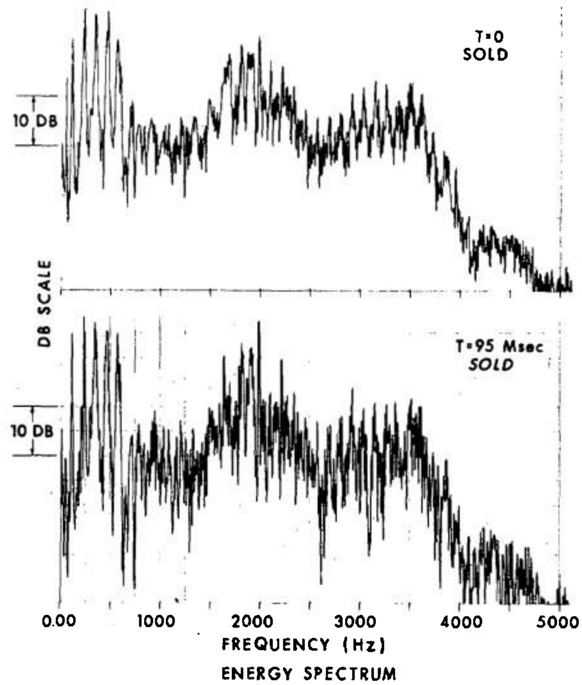


Fig. 5. Graphic representation of the word "sold" depicted as a channel impulse response of the squared magnitude of the Fourier transform.

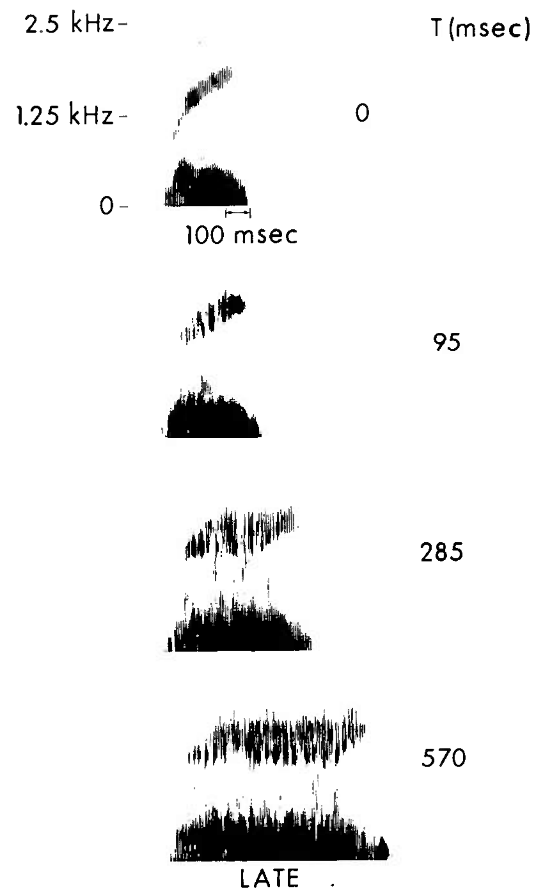


Fig. 6. **Spectrographic recordings of the word "late". The duration of the impulse response in milliseconds is shown to the right of each spectrogram.**

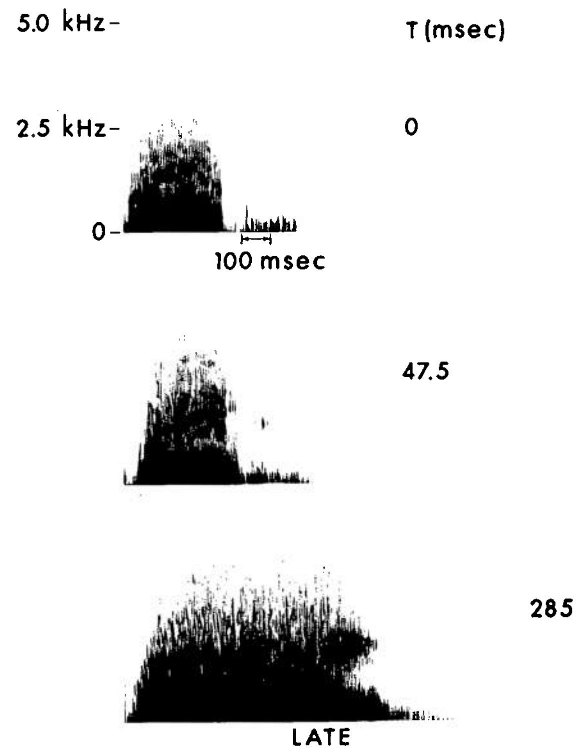


Fig. 7. **Spectrographic recordings of the word "late". The duration of the impulse response in milliseconds is shown to the right of each spectrogram.**

They were presented five lists of 50 words each, one list undistorted, and four additional lists distorted by the channel simulation.

Results in Figure 8 show that the intelligibility of speech decreased up to a point with increasing impulse response duration (T). The curve approaches 75 percent intelligibility asymptotically at 200 msec.

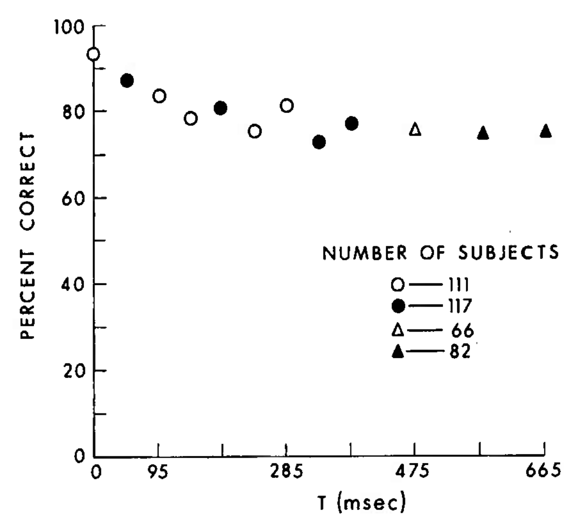


Fig. 8. **Intelligibility of speech (Percent Correct Responses) as a function of impulse response duration (T).**

The results shown here are consistent with spectrograms of the speech shown in Figures 6 and 7. Recall that some sounds which depend upon sharp bursts of energy, such as stops, will be smeared over by the convolution process. Other sounds, such as nasal consonants or semi-vowels, which have smoother transitions between themselves, might be less affected by the process of convolution. We would thus expect the intelligibility curve to approach some non-zero asymptotic value at smear lengths long enough to destroy certain sounds while only prolonging the duration of others. Preliminary analysis of our data supports this interpretation of results. For example, with a smear length of 380 msec, less than 50 percent of the stops are correctly identified, whereas more than 95 percent of the semi-vowels and nasals are heard correctly.

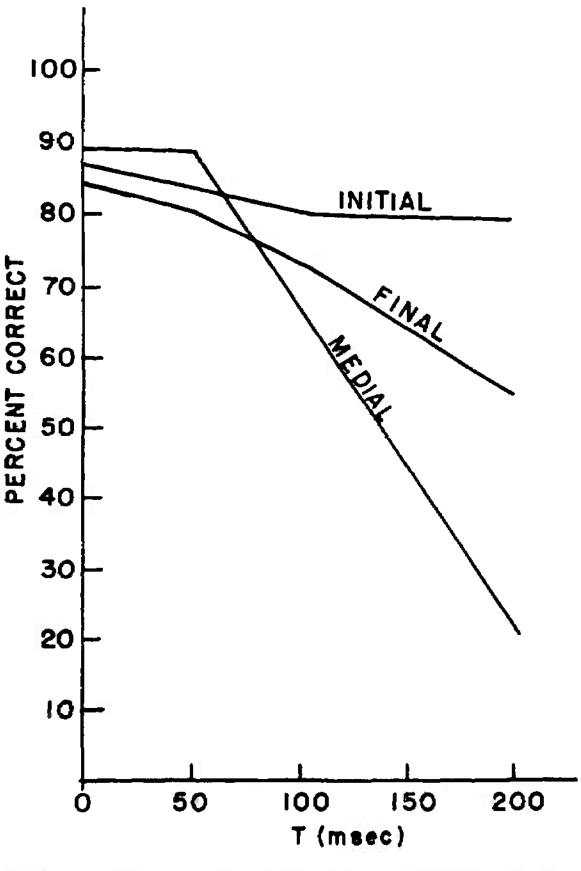


Fig. 9. Intelligibility of speech (Percent Correct Responses) as a function of impulse response duration (T). Parameters indicate the position of consonant sounds within test words.

C. **Experimental Evaluation of Consonant's Position in Words.** The Modified Rhyme Test evaluates intelligibility of consonants at the beginning and end of monosyllabic words. Interpretation of our results suggests, however, that consonants in the medial position, as they are typically found in conversational speech, would be more affected by time smearing than either initial or final consonants, since the medial consonant would be distorted by energy displaced in time both from the preceding and the following vowels. Consequently, we predicted the intelligibility curve for consonants in the middle of words to reach a non-zero asymptote at a much lower percentage correct point than consonants in the initial and final position. To test this hypothesis, we convolved nonsense words with the simulated channel response and presented these words to panels of listeners. The words had the form "consonant-ah", "ah-consonant-ah" and ah-consonant". In Figure 9 the mean percent correct responses for initial, medial, and final consonants are plotted as a function of the durations of smear. As expected, the medial consonants are much less intelligible for the longer smears than are the initial or final consonants. We conclude that the intelligibility of connected time-smeared speech will be worse than that obtained for single monosyllabic words.

V. SUMMARY

A single gross ocean transmission path was modeled as a linear, time-invariant channel whose impulse response is a sample of Gaussian noise. The intelligibility of speech passed through this simulated channel was measured as a function of the impulse response duration or "smear length." Intelligibility decreased to about 75 percent at a smear length of 200 msec for tests containing monosyllabic words. Increasing the duration of smear beyond 200 msec did not significantly lower intelligibility further. We conclude that distortions of time-smearing imposed by transmission of monosyllabic words through the ocean, as simulated in the present study, do not impose serious limitations on acoustic transmission of speech. However, since consonants in a medial position in a syllable are more affected than when in an initial or a terminal position, these time-smearing distortions will more seriously interfere with the intelligibility of connected or conversational speech.

REFERENCES

1. Stockham, T. G., High Speed Convolution and Correlating. 1966 Joint Computer Conference, **AFIPS Proceedings**, Spring Meeting, Vol. 28, pp. 229-233.
2. House, A. S., Williams, G., Hecker, M. H. L., and Kryter, K. D., Psychoacoustic Speech Tests: A modified Rhyme Test. **Technical Documentary Report No. ESD-TDR-63-403**, 1963.

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) Naval Submarine Medical Center, Submarine Medical Research Laboratory and Navy Underwater Sound Laboratory	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

SPEECH INTELLIGIBILITY IN A STATIONARY MULTIPATH CHANNEL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim Report

5. AUTHOR(S) (First name, middle initial, last name)

Murray B. Sachs, Gary J. O'Brien, Russell L. Sergeant and Joseph S. Russotti

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
21 March 1969	6	2

8a. CONTRACT OR GRANT NO. b. PROJECT NO. MF12.524.004-9011D.03 c. NavShips Systems Command Task 5-23-18-08622 d. (USL-9-A-611-03-00)	9a. ORIGINATOR'S REPORT NUMBER(S) SMRL 574 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Submarine Medical Center Box 600 Naval Submarine Base Groton, Connecticut 06340

13. ABSTRACT

The reception of speech transmitted through an acoustic channel such as the ocean is limited by multipath "time-smearing." The purpose of this study was to obtain a quantitative measure of the effects of such time-smearing on speech intelligibility. A linear, time-invariant channel was used as a model of the ocean. The impulse response of this channel was a sample of band-limited Gaussian noise. Using Fast Fourier Transform techniques, words of the Modified Rhyme Test were convolved with, or smeared in the time domain, by this channel impulse response. The intelligibility of these "smeared" words was measured as a function of the impulse response duration, T. Intelligibility decreased monotonically to about 75 percent as T increased to 200 milleseconds. Further increase in T did not significantly lower intelligibility. Distortions in time evaluated herein did not impose serious limitations to the reception of short words. However, a detailed analysis of consonantal errors revealed that sounds occurring in the middle of a word are much harder to hear correctly than are sounds at the beginning or ending of an utterance. We conclude that time-smearing will more seriously interfere with the intelligibility of connected or conversational speech.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification

3ND PPSO 13152

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech intelligibility						
Underwater communication						
Speech distortion underwater						