Speech-signal processing and applications to single-sideband
by Norman A Shyne

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in Electrical Engineering
Montana State University

Abstract:
The results of signal processing as applied to SSB systems is described in this paper.

In the preliminary phases of the research, certain procedures, designs, and definitions are established: 1. An articulation test procedure 2. The definition of a measure of signal-to-noise ratio useful for peak-limited systems 3. General equipment and experiment designs as related to SSB signal processing A review of processing as applied to audio signals is presented and the equivalence of intelligibility scores for fixed- and variable-volume listening conditions is cited. Linear processes of power-spectrum equalization and differentiation are shown to provide minor intelligibility increases for λ defined at audio. For λ defined at narrow-band and in conjunction with finite degrees of clipping at audio, DSB, and SSB these linear operations generally give rise to intelligibility degradation. Intelligibility considerations for an audio AVC unit are presented; the results indicate that, under certain conditions, the AVC unit provided nearly the same intelligibility increases as 12 db of audio clipping.

The three most fundamental (and accurate) intelligibility estimates were made for: audio, DSB, and SSB clipping with λ defined at narrow-band. Of these, SSB clipping provided the most significant increase (for a given signal peak value) and at the same time more fully preserved signal quality. It is shown, for example, that for a given noise level, 60% intelligibility can be maintained by using an SSB signal which has been clipped 36 db and has a peak value 12 db below that of an unprocessed signal. The interdependent and iterative combinations of the three clipping operations yielded advancements in intelligibility estimates; however, in no case did they exceed those derived from an equivalent amount of SSB clipping.

The final consideration of this investigation was a reduced-X signal condition and its effect upon the intelligibility estimates for audio and SSB clipping. For both audio and SSB, the reduced-X signal yielded lower intelligibility scores. It appears that a parametric relationship of the form C max= λ+δ,&dealta;≈ db exists for maximum clip- ping; greater amounts of clipping decrease intelligibility scores for a given amount of noise.

SPEECH-SIGNAL PROCESSING AND APPLICATIONS TO SINGLE-SIDEBAND

by

N. A. Shyne

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Byron J. Bennett
Head, Major Department

Donald K. Weaver
Chairman, Examining Committee

Leon H. Johnson
Dean, Graduate Division

MONTANA STATE COLLEGE
Bozeman, Montana

March, 1962

ACKNOWLEDGEMENTS

I wish to express my appreciation to Professor Donald K. Weaver, Jr. for his thoughtful guidance and encouragement in this research. Thanks are also extended to Mr. John F. Bowker for his part in the realization of the experimental design, compilation of test materials, and suggested improvements of the manuscript.

The research reported in this paper has been sponsored by the Electronics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under Contract AF19(604)7365.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

The results of signal processing as applied to SSB systems is described in this paper.

In the preliminary phases of the research, certain procedures, designs, and definitions are established:

1. An articulation test procedure
2. The definition of a measure of signal-to-noise ratio useful for peak-limited systems
3. General equipment and experiment designs as related to SSB signal processing

A review of processing as applied to audio signals is presented and the equivalence of intelligibility scores for fixed- and variable-volume listening conditions is cited. Linear processes of power-spectrum equalization and differentiation are shown to provide minor intelligibility increases for λ defined at audio. For λ defined at narrow-band and in conjunction with finite degrees of clipping at audio, DSB, and SSB these linear operations generally give rise to intelligibility degradation. Intelligibility considerations for an audio AVC unit are presented; the results indicate that, under certain conditions, the AVC unit provided nearly the same intelligibility increases as 12 db of audio clipping.

The three most fundamental (and accurate) intelligibility estimates were made for: audio, DSB, and SSB clipping with λ defined at narrow-band. Of these, SSB clipping provided the most significant increase (for a given signal peak value) and at the same time more fully preserved signal quality. It is shown, for example, that for a given noise level, 60% intelligibility can be maintained by using an SSB signal which has been clipped 36 db and has a peak value 12 db below that of an unprocessed signal. The interdependent and iterative combinations of the three clipping operations yielded advancements in intelligibility estimates; however, in no case did they exceed those derived from an equivalent amount of SSB clipping.

The final consideration of this investigation was a reduced-λ signal condition and its effect upon the intelligibility estimates for audio and SSB clipping. For both audio and SSB, the reduced-λ signal yielded lower intelligibility scores. It appears that a parametric relationship of the form $C_{max} = \lambda + \delta$, $\delta \approx 8$ db exists for maximum clipping; greater amounts of clipping decrease intelligibility scores for a given amount of noise.

SPEECH-SIGNAL PROCESSING AND APPLICATIONS TO SINGLE-SIDEBAND

1. INTRODUCTION

1.1. History and Background

The process of communication is fundamental to the existence of our society as we know it today. As used in the general sense, the term communication is broadly inclusive; humans communicate with one another in myriad ways. In the case of human-to-human communications, the spoken word is probably the most important vehicle of information, although the communicative process may be carried on by appealing to any of the senses.

For many years researchers have placed a firm concentration on auditory speech signals; and, as a result, have gained an immeasurable amount of information concerning the speech production and speech detection mechanisms. Initially, such information was of primary concern to psychologists and linguists. However, with the advent of rapidly expanding telephone networks in the 1920's, engineers became aware of the problems of transmitting aural signals and turned to previous knowledge. For example, in 1922 Stewart proposed "An Electrical Analog of the Vocal Organs"; Crandall and MacKenzie (1922) analyzed the energy distribution of speech. Later, Steinberg (1929) and Knudsen (1929) considered the effects of echoes, phase distortion, and reverberation on the intelligibility of speech. Martin (1930) initiated bandwidth studies; Dudley's "The Vocoder" (1939) embodied fundamental concepts which are in wide use in present bandwidth-reduction schemes.

World War II provided a great impetus to research investigations concerned with speech intelligibility. Many significant results were obtained during this period. Of particular interest to the war effort was, of course, the improvement of intelligibility of speech in the presence of noise for military applications. Some contributors during this era were Egan et al (1943), Egan and Wiener (1946), Smith (1946), French and Steinberg (1947), Pollack (1948), and Licklider and Pollack

(1948). These investigations confirmed that a speech signal and the characteristics of the human ear are such as to tolerate remarkable amounts of alteration or processing of the signal without giving undue loss of intelligibility.

Since 1950, the acquisition of knowledge concerning speech production, processing, transmission, and reception has been directed in five general areas:

1) Signal processing related to communication systems
2) Bandwidth conservation
3) Oral machine control and machine translation of languages
4) Equipments
5) Linguistics, speech, psychology

The first four areas have been of interest to engineers and scientists in the general field of communication. Research in the second, third, and fourth areas has been typified by advancements in time and frequency domain vocoders, digital-computer simulation of speech (David, Mathews, MacDonald, 1958), auditory recognition (David, 1958), and complex systems for military and civilian communications networks. The first area, of particular interest to the author, has acquired additions such as speech peak clipping (Martin, 1951; Kahn, 1957; Wathen-Dunn and Lipke, 1958); the evaluation of processing devices and more sophisticated system evaluation techniques.

1.2. Signal Processing Related to Communication Systems

Communication systems range in complexity from a simple conversation between two humans standing face-to-face to the most highly complex electronic devices covering a span of thousands of miles. Even with this latitude of complexity, a speech communications system is generally construed to be composed of a transmitter, a receiver, and an inter-connecting channel. System requirements then determine the complexity of each of the system components.

Due to its high ratio of peak-to-average values, a speech signal is not generally well-suited for transmission over a given communication system. The question arises then: "Can a certain form of processing make a speech signal more suitable for transmission?" Fortunately, researchers have recognized this problem and have found that speech signals possess certain elements which are not essential to the psychoacoustic process of speech recognition. However, no one has been able to show precisely the totality of what elements are essential or why. On a piecewise basis it has been shown that certain bands of frequencies in the spectrum of a speech signal are not required for recognition; it has also been shown that a severe restriction of peak excursions of a signal will not seriously impair its intelligibility. So without real definition, researchers have undertaken the task of mapping auditory speech signals to a domain where they are manageable to the point of (hopefully) preserving only the information-bearing characteristics. Circuit and system components have been designed to preserve these characteristics, with constraints applied by the ultimate receptor of the signal.

In the communicative process, the receptor of a speech signal is a human being (barring a few possibilities such as voice-controlled machines). If a signal is to be useful, it must be intelligible; signal intelligibility is a relative measure of the effectiveness of the transmission of the signal with a given set of system parameters. Unfortunately, there exists no instantaneously-reading meter which displays intelligibility as a function of signal-to-noise ratio, speaker clarity and enunciation, system bandwidth, signal envelope preservation, or the listener's ability to hear. The only available meter, if it may be called such, for the determination of single-word articulation scores is a test using a team of listeners. Certain limitations and perturbations of speech signals are observed by a team of normal, well-trained listeners. From this point, one possibility is for the listeners to write down what they hear; these written results compared to a standard key will define articulation (or intelligibility) score.

Recent conditions of overcrowding the spectrum space have prompted numerous investigations of bandwidth compression techniques. Perhaps due to the concern regarding this spectrum crowding, certain aspects of signal processing as applied to analog speech signals in communication systems have been slighted or perhaps even overlooked. The application of more sophisticated combinations of speech processing techniques does not seem to have been generally accepted or widely recognized as a valid possibility of improving system performance. On the contrary, most present military specifications for voice communications equipment specify rather rigid tolerances on linear response and a uniform speech pass-band, though some include provisions for a nominal amount of peak clipping. This situation is somewhat surprising in view of the possible improvements in system performance which might be obtained at nominal costs through the application of more sophisticated processing techniques.

1.3. A statement of the Research Problems

Signal processing is a mapping operation. Presumably in an actual system this mapping must be performed by devices which retain at least part of the information-bearing characteristics of the signals. For the case of single-sideband (SSB) communication systems, there has been no concerted effort to determine: what degree of processing may be performed, or what effect processing in one part of the system has upon signal characteristics in another part.

As a result of definition and physical restrictions, signal processing may be categorized into the following types:

I. Linear
II. Instantaneous non-linear
III. Slow operations on averages
IV. Fast operations

A process of $P(V)$ is linear if for every pair of variables V_1 and V_2 and for each pair of constants k_1 and k_2 the following relation holds:

$$P(k_1V_1 + k_2V_2) = k_1 P(V_1) + k_2 P(V_2) \tag{1-1}$$

Examples of linear processing are: spectrum tilts, frequency translation, and equalization of the power spectrum. Type II processing requires that there exist an output signal which is uniquely defined for each input signal and is independent of the past history of the input (no memory); peak clipping is an example of such processing. Type III effects changes in signal characteristics as dictated in some unique way by the mean or average of the signal; an example is automatic gain control. Type IV processing is contingent upon fast operations which are dependent upon the harmonic content of the signal. Note that each of these processes may be considered an analog signal operation: the input and output signals of these processing devices are describable by equations of similar form on a one-to-one basis. The vocoder, by this definition, is not an analog operation.

For purposes of this investigation, the better of two processes will be taken as that which yields the highest intelligibility score for given signal bandwidth and noise corruption. Further criteria such as "quality" and "naturalness" of the processed signal as presented to the receptor will not be of primary concern.

Obviously, it was impossible to investigate all combinations of processes applicable to all systems. This study was delimited to an investigation of various linear and non-linear operations on analog speech signals in a representative, peak-limited, SSB system; the intrinsic effect of these operations upon the psycho-acoustic process of speech recognition was of primary concern. Only combinations of Types I, II, and III processing were considered.

In order to study signal processing as applied to SSB systems, it is necessary to define certain signal and system models to which fundamental constraints may be applied. To effect this end, basic definitions, derivations, assumptions, and theoretical considerations are outlined in Chapter 2. Most of this information is not new, but it forms a logical basis for the experimental design presented in Chapter 3. The results of the study are given in Chapter 4, and a discussion based upon these results is presented in Chapters 5 and 6. In general, circuit diagrams and descriptions, lengthy derivations, and subsidiary data are placed in the appendixes following the literature consulted.

2. DEFINITIONS, DERIVATIONS, AND THEORETICAL CONSIDERATIONS

Signals and systems may generally be classified into major groups, with relatively precise group definitions. In order to clarify divers discussions of signal processing as applied to SSB systems that follow, certain classifications, definitions and assumptions are necessary.

2.1. Properties of Speech Signals

As initially derived from the human voice-production mechanism, speech is a continuous, single-valued function of time having a dc component.* A speech signal is limited in frequency range and considered to be a broadband** signal. The peak-factor (defined as the ratio of the peak to rms values) of a speech signal is about 15 to 20 db. In general, the peak excursions of the signals are unsymmetrical (even

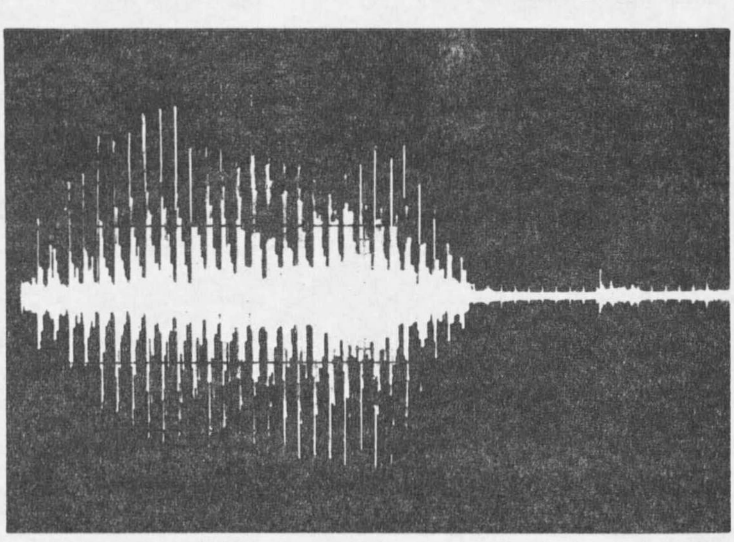

Figure 2-1. An Oscillogram of the word "add" band-limited from 300 to 3000 cps.

* The "breath" bias due to exhaling while speaking normally.
**Defined in Appendix B.

if the dc component is neglected) and a sample of finite duration contains a finite power. These properties become evident after consideration of the "time-domain picture" of the word "add" as shown in Figure 2-1. It is convenient for the communications engineer to consider an analog speech signal as an electrical waveform, representative of the sound-pressure waveform, with a frequency characteristic restricted to the baseband (typically 300 to 3000 cps). Obviously, the analog signal no longer contains the dc component.

2.2. Speech Transmission Systems

The process of modulation is an inherent part of signal transmission. A modulation may be effected in a variety of ways which yield a variety of signal types. Some familiar types of modulation are:

1. Amplitude
2. Double-sideband, suppressed carrier (DSB or D)
3. Single-sideband (SSB or S)
4. Phase
5. Frequency
6. Pulse Code

With due regard to the attendant problems of other methods of modulation, the delimitation to consider only an SSB system was made for the purposes of this investigation; however, some of the results obtained may be applicable to other methods.

The ideal process of generating an SSB signal is a linear frequency-translation to the domain of a narrow-band signal. In practice, translation may be effected by one of the following methods:

1. Filter
2. Phase
3. Quadrature Function (QF)

The first two methods have been recognized and used extensively; Weaver (1958) has pointed out advantages of the QF method. The decision to use the filter method of SSB generation for this study was made after