

ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

BASE-CLIPPED SPEECH COMMUNICATIONS

Technical Report No. 57
Electronic Defense Group
Department of Electrical Engineering

By: J. L. Stewart

Approved by:


J. A. Boyd

Project 2262

TASK ORDER NO. EDG-8
CONTRACT NO. DA-36-039 sc-63203
SIGNAL CORPS, DEPARTMENT OF THE ARMY
DEPARTMENT OF ARMY PROJECT NO. 3-99-04-042
SIGNAL CORPS PROJECT NO. 194B

December, 1955

TABLE OF CONTENTS

	Page
List of Illustrations	iii
Abstract	iv
1. Introduction	1
2. Additional Modifications of Clipped Speech	4
3. Base Clipping	8
4. Microwave Applications	10
5. Minimum Pulse Rate	12
6. Practical Clipping Circuitry	13
7. Clipping Microphones	17
8. Speech Presentation	18
References	20

LIST OF ILLUSTRATIONS

		Page
Fig. 1	Standard Speech Clipping System	1
Fig. 2	Normal Speech and Clipped Speech	2
Fig. 3	Clipped Speech Operations	4
Fig. 4	The Integration Effect	5
Fig. 5	Technique of Speech Reconstruction	6
Fig. 6	Base-Clipped Waveforms	8
Fig. 7	Effect of Bias on Rate and Articulation	9
Fig. 8	Dependence of Articulation on Rate	9
Fig. 9	Typical Amplifier Characteristics	14
Fig. 10	Biased Rectification Circuit	15
Fig. 11	Multivibrator Clipping Circuit	16

ABSTRACT

The concepts of pre-emphasized and infinitely clipped speech are extended to suggest certain pulse communication systems. If the zero crossings of the speech waveform in both directions are retained, pulse rates during words is about 3000 per second. If only positive zero crossings are used, the rate is halved, yet the speech may be reconstructed satisfactorily to clipped speech. It is further possible to reduce the rate below 1000 by base clipping which does not transmit weak speech sounds and which effects a simple squelch circuit.

Speculations of scientific interest are made of the possibility of further reducing the pulse rate and for increasing the fidelity of clipped speech at the receiving end of the system.

Finally, practical base-clipping circuits are described, a system utilizing a triggered free-running multivibrator in particular. In addition, the construction of a microphone which directly realizes pulsed speech is discussed.

BASE-CLIPPED SPEECH COMMUNICATIONS

1. INTRODUCTION

It is well known that heavily clipped speech is highly intelligible, and, in fact, is quite satisfactory for normal communications channels.¹ Radio amateurs have used the technique for years in order to gain a power advantage.² As compared to an AM transmitter with linear speech modulation, an AM transmitter having the same average power but modulated with the square waveform of clipped speech may permit detection at signal-to-noise ratios of 10 to 12 db less.¹ Square-wave modulation in itself is interesting because a transmitter can be modulated at low level without sacrifice in efficiency because modulation can be essentially a keying operation.

A standard speech clipping arrangement is shown in Fig. 1. Normal speech is first passed through a linear circuit having a rising frequency charac-

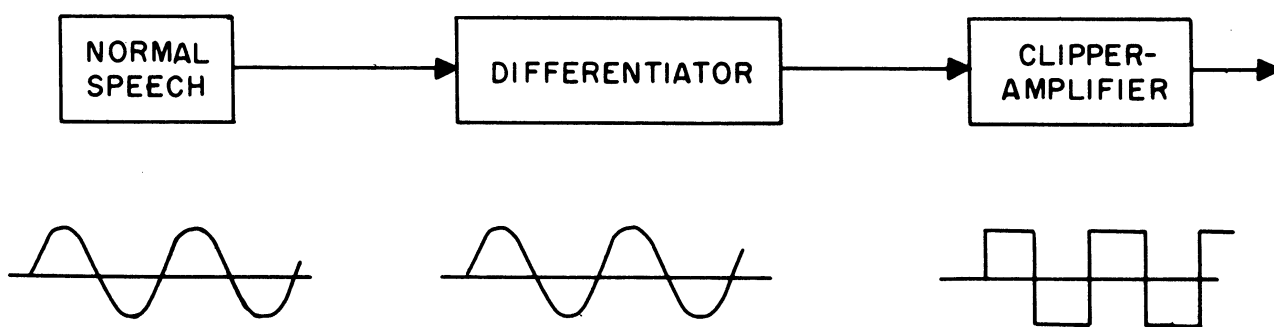


FIG 1. STANDARD SPEECH CLIPPING SYSTEM

teristic (at 6 db per octave), after which it is symmetrically clipped, amplified, clipped, and so on, until the waveform is essentially square. If the resulting waveform is presented to an observer and articulation score tests made, the artic-

ulation score is approximately 95 percent.¹ Articulation scores associated with undistorted speech run about 98 percent. Any system giving an articulation score above about 90 percent is quite satisfactory for normal speech communications because sentence structure is redundant.¹

Extreme clipping is a highly nonlinear process. When two signals are present in a waveform and one is weaker than the other, the clipping operation suppresses the weaker of the two (as in an FM receiver with limiters). Normal speech is composed of many frequency components. On the average, the strong components are at low frequencies and are associated with the vowel sounds. The high-frequency components are weaker and are normally associated with the consonant sounds. Clipping without pre-emphasis of the high-frequency components can therefore suppress much information, and has been found to lead to articulation scores on the order of only 40 percent.¹ Whether pre-emphasis circuits with slopes greater than 6 db per octave give even better articulation scores is not well known. There is some evidence that pre-emphasis with a circuit having a slope of 12 db per octave does not impair intelligibility.³

Let us consider the waveforms of speech and clipped speech as shown in Fig. 2. It is evident that of the complex waveform of normal speech, the only

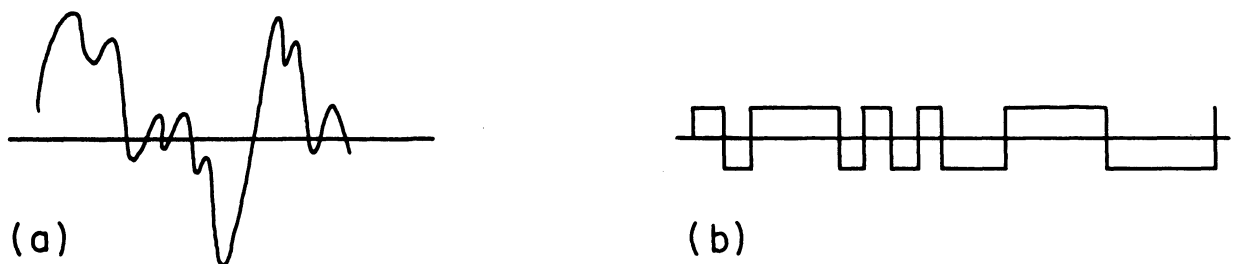


FIG. 2. NORMAL SPEECH AND CLIPPED SPEECH

characteristics retained are the zero crossings in both directions. During normal speech, there occur about 1500 square waves per second; therefore, in order

to account for zero crossings in both directions, 3000 events per second must be catalogued.¹ (The rate specified here is that during the time the words are spoken; it does not include the pause periods between words.)

It is possible to differentiate the clipped speech waveform and generate pulses (of any short width) at an average rate of 3000 per second for transmission. This can, for example, be done with very simple microwave equipment.

Ordinary clipped speech is quite intelligible when listened to through an amplifier with a flat low-pass characteristic; however, it is not very "natural" sounding. If an integrating network is placed in the speech amplifier preceding speaker or earphones (which is a circuit having a frequency characteristic sloping downward at 6 db per octave), the speech sounds more natural, yet no change in the articulation score results.¹ In terms of articulation scores, the frequency characteristics of the amplifier following the clipping stage are evidently quite uncritical.

There are two disadvantages to ordinary clipped speech which are of concern in communications systems. First is the noise that occurs between words; thermal noise, background noise, and hum are clipped the same as speech and subsequently transmitted (although hum is conveniently minimized by the presence of the pre-emphasis circuit preceding the clipper.) This makes a squelch circuit desirable. Experiments have shown that with a squelch circuit, the noise between words is removed, yet the articulation score is not affected.¹

In addition, the transmission of as many as 3000 pulses per second seems wasteful. Many speech sounds are so weak that they have little effect on the articulation score or naturalness. If these sounds are not transmitted, considerably fewer than 3000 pulses per second may be adequate.

2. ADDITIONAL MODIFICATIONS OF CLIPPED SPEECH

Obviously, it is an easy matter to transmit pulses corresponding only to zero crossings in one direction. This immediately halves the average required rate from 3000 to 1500. However, there may be some effect on naturalness, intelligibility, or both, unless proper measures are taken at the receiving end of the system.

Consider a sample of clipped speech and the pulses arising therefrom as shown in Fig. 3. The pulses of Fig. 3c occur at an average rate of 3000 per

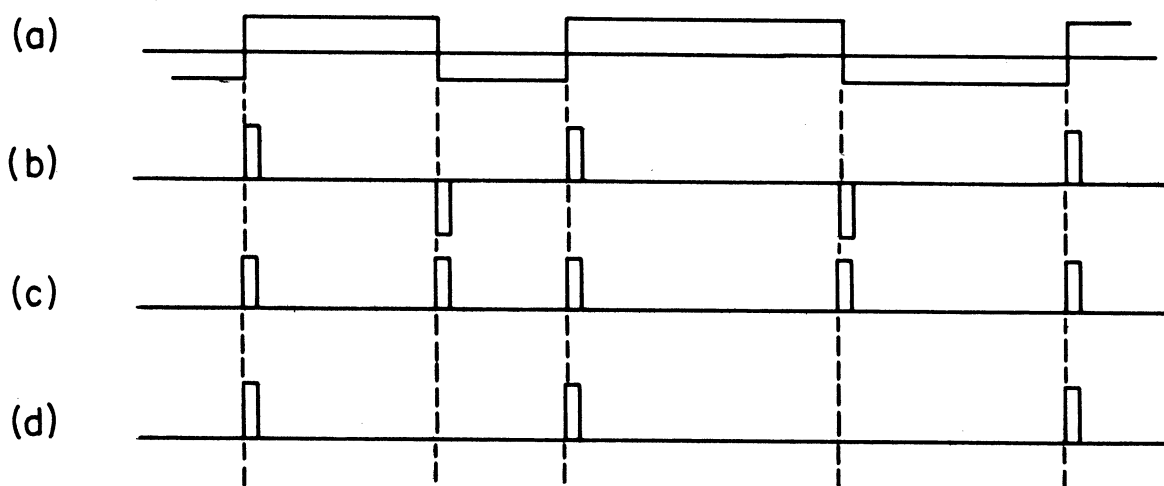


FIG. 3. CLIPPED SPEECH OPERATIONS

second. They can be converted back to the original clipped speech by causing them to operate a bistable trigger circuit. The effect of noise in the system will be to cause some false triggering, although the effect will be slight unless the noise is relatively large.

The waveform of Fig. 3d can be integrated directly and observed. However, some naturalness is lost; after all, less information is contained in Fig. 3d than in Fig. 3c because the positions of zero crossings in the negative direction are not indicated. Consider, for example, a speech sound with several

closely spaced pulses. Integration of this set of pulses yields a low-frequency component that was not in the original speech (see Fig. 4) which has a duration comparable to that of the speech sound. Normal clipped speech does not suffer

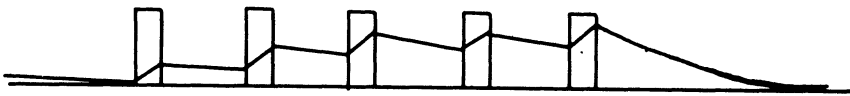


FIG. 4. THE INTEGRATION EFFECT

from this "integration effect" because the square waveform is symmetric.

Obviously, a simple bistable trigger operated from the waveform of Fig. 3d will halve all the speech frequency components. Whether this will have an effect on naturalness and/or the articulation score is not known.

The waveform of Fig. 3d can, however, be reconstructed into clipped speech. The result can never be identical to the original clipped speech, although the effect on naturalness and articulation appears to be negligible.³ It is surprising that a waveform as simple as that of clipped speech retains most of the characteristics of speech. It is even more surprising that even clipped speech is so redundant. Actually, only about 100 pulses per second are required to convey information at the same rate as normal speech (as with Morse Code). In terms of information alone, a rate of 1500 per second is not surprising. That

the 1500 rate retains naturalness and yet does not require complicated coding devices at the receiving end gives an indication as to how much information is required to give speech some naturalness and speaker recognition. Later, it will be shown that the average rate can be reduced even further--to something considerably less than 1000 per second.³

In order to reconstruct the waveform of Fig. 3d, pulse circuitry can be devised to put one full and symmetric square wave between each pair of pulses, with the period of the square wave determined by the adjacent pulse spacing. The original differentiated and clipped speech pulse waveform, 1500 pulse rate waveform, and reconstructed speech are shown in Fig. 5. The zero crossings in

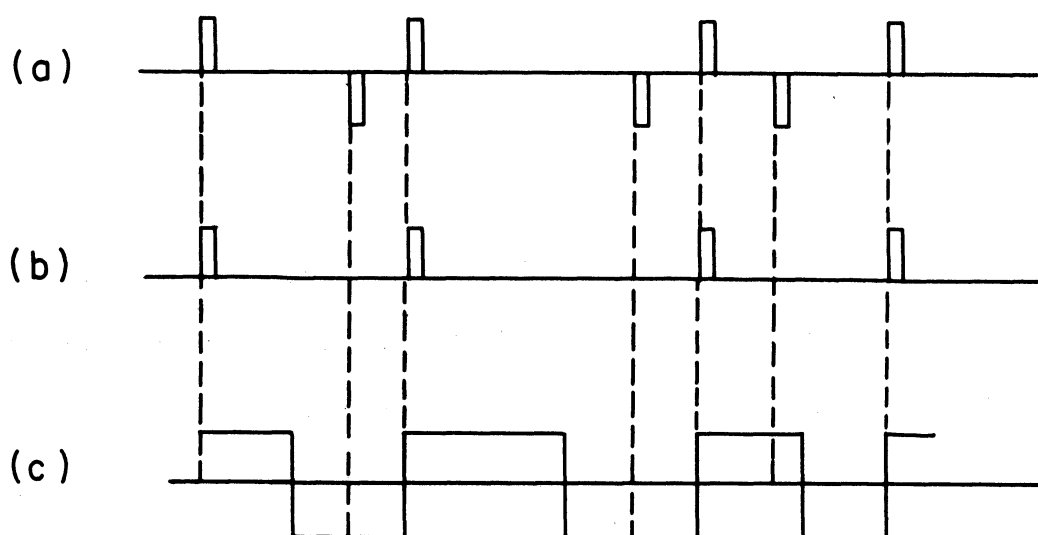


FIG. 5. TECHNIQUE OF SPEECH RECONSTRUCTION

the negative direction of the reconstructed speech are not the same as those of the original speech, although they are the same on the average. The imperfection of the reconstruction is apparently not significant in terms of articulation score as compared with ordinary clipped speech.

The reconstructed waveform of Fig. 5c can be approximated by causing

the pulses of Fig. 3d to trigger a monostable multivibrator with a pulse duration comparable to the average spacing between the pulses of Fig. 3c. However, some spacings are so short that such a multivibrator can not be retriggered as often as required, whereas other pulse spacings are so large that the output waveform still has the characteristics of short pulses and hence is subject to the integration effect. Reconstructed speech as in Fig. 5c is not subject to the integration effect shown in Fig. 4 because the waveform is symmetric on the average.

One final phenomenon that reduces the naturalness of clipped speech remains to be discussed. In undistorted speech, the sound builds up more or less gradually. With clipped speech all sounds begin abruptly. This causes, for example, organ music to sound reasonable after clipping, although jazz (i.e. jumpy music) sounds quite poor.³ (It is not to be inferred that clipped music is acceptable. The music example is merely a convenient and enlightening one.) When there is considerable noise between words, the abrupt beginning of a speech sound is not in itself noticeable because the average sound power reaching the observer does not greatly change. With squelch circuits, however, the abrupt beginning is noticeable, although the phenomenon does not appear to affect the articulation score.

The integration effect can be utilized as an "inverse" automatic volume control; some tube in the receiving system can be biased to give low gain, and the reception of several adjacent pulses, suitably integrated, can gradually increase the gain of this tube, thereby causing the sound level to build up more or less gradually. In order not to affect the articulation score, the integration time constant of the circuit must not exceed a few pulse periods. An abrupt word ending does not appear to be objectionable and hence the inverse AVC need not control the end pulses in a speech sound. (Even in Morse Code reception, a slow tone buildup is desirable. An abrupt tone ending is not objec-

tionable and, in fact, may be preferable to a slow decay.)

3. BASE CLIPPING

Consider the (pre-emphasized) speech waveform of Fig. 6a. If this waveform is rectified (Fig. 6b) and then clipped (Fig. 6c), the result is clipped speech as discussed before. The use of rectification prior to clipping introduces

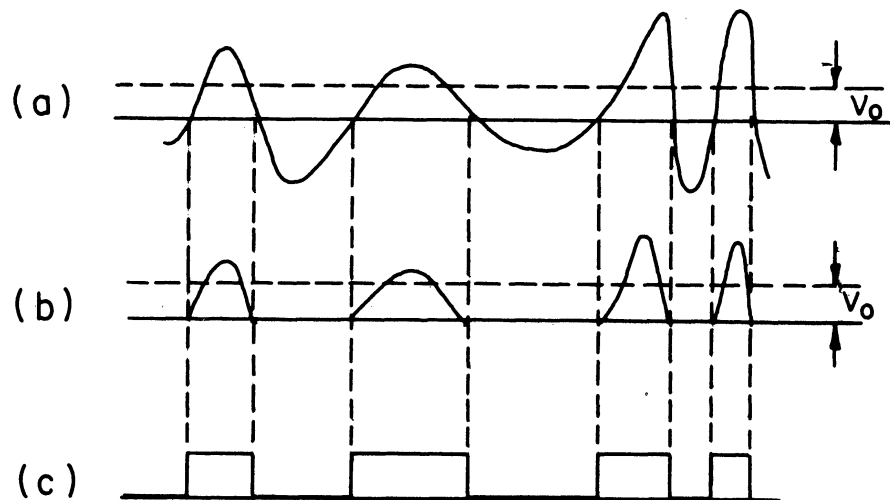


FIG. 6. BASE-CLIPPED WAVEFORMS

another concept, because rectification can involve a bias level. Accordingly, the rectifier can be biased so that only that portion of the speech that exceeds the bias level V_0 passes on to the clipper. For negligible V_0 , the result is clipped speech. For moderate V_0 , the noise between words (which is small) does not exceed V_0 , and hence a simple squelch circuit is realized; the speech during words remains the same as clipped speech. As the bias level is increased even further, some of the weak speech sounds are no longer strong enough to exceed the bias level and the average pulse rate during words is reduced below 1500 per second. However, because very weak sounds are probably not even heard by the observer, no less in intelligibility or naturalness results.

Let a graph be visualized giving the articulation score as a function of the rectifier bias setting, and also the average pulse rate (during words only) as a function of the same bias setting. (The dashed curve in Fig. 7 shows what might occur if the average rate includes the noise and hum pulses between words.) The result is Fig. 7 (which is strictly qualitative.) As V_0 is increased, the

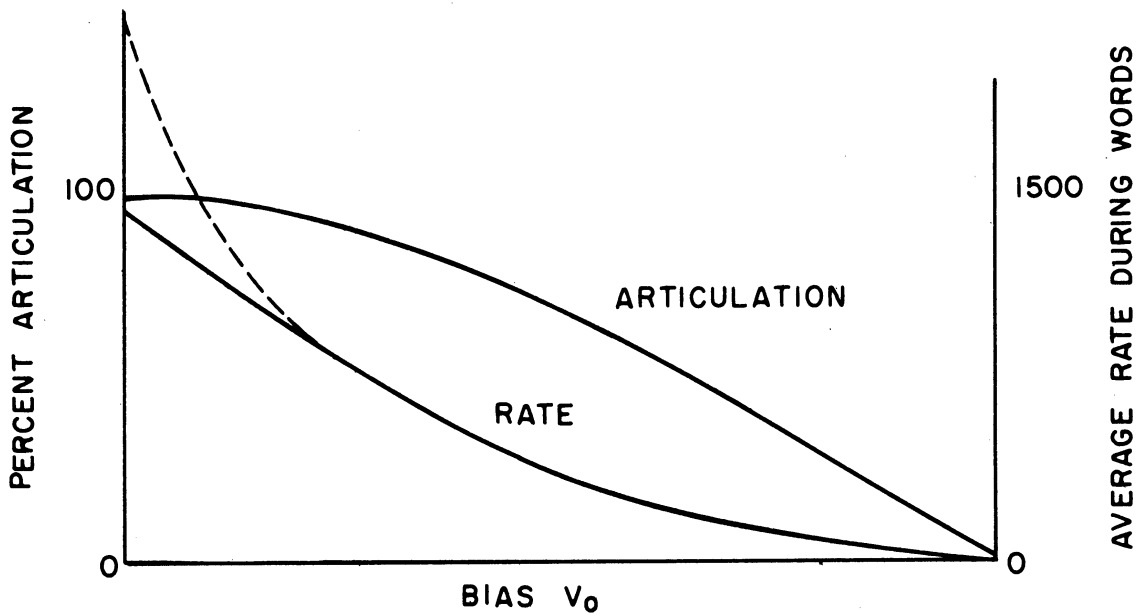


FIG. 7. EFFECT OF BIAS ON RATE AND ARTICULATION

removal of weak sounds reduces the rate appreciably, although it does not appear to have an effect on the articulation score. The shapes of the curves of Fig. 7 are probably intuitively obvious. They have been qualitatively confirmed.³ The curves in Fig. 7 can be combined as in Fig. 8 to give a qualitative curve of

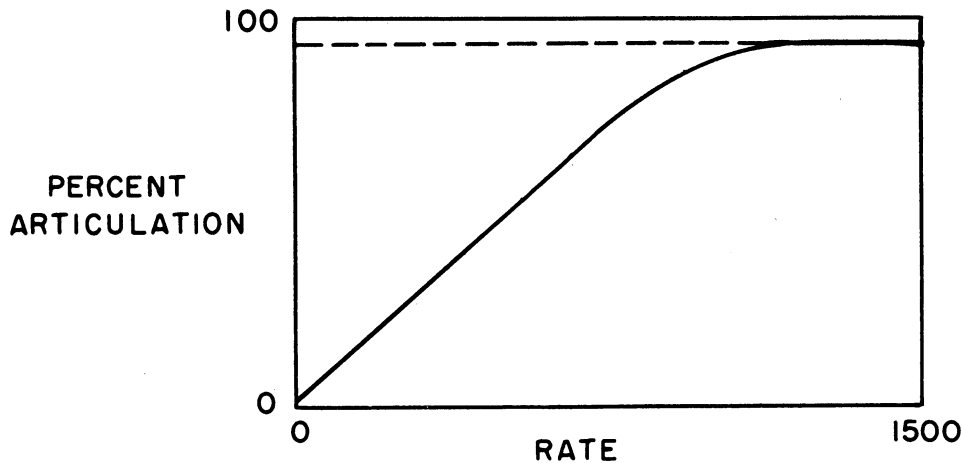


FIG. 8. DEPENDENCE OF ARTICULATION ON RATE

articulation score as a function of average pulse rate during words. The important concept to get from this figure is the appreciable reduction in rate without affecting either naturalness or articulation score.

Biased rectification is far superior to ordinary clipping methods. For zero bias, it gives the same result. For small bias, it gives squelch, which is difficult to mechanize when symmetric clipping stages are employed. For larger bias settings, it is capable of appreciably reducing the pulse rate during words (to something on the order of 700 or 800 per second) without serious consequences. Finally, the practical electronic circuitry is far simpler. (Practical clipping circuitry will be described later.)

It should be noted that with biased rectification, speech appears to require only 700 or 800 pulses per second during words in order to give high intelligibility, reasonable naturalness, and some speaker recognition. This rate can be contrasted to the 3000 per second rate heretofore thought needed for clipped speech (sending information regarding both positive and negative zero crossings) and the 20,000 or higher rate required by pulse-coded systems (although standard pulse-coded systems are high quality and suitable for music, etc.)

4. MICROWAVE APPLICATIONS

The low rate actually required for communications (not entertainment) systems makes simple microwave transmitter-receivers an attractive possibility. A small pulsed radar type transmitter can be pulsed at an average rate of below 1000 per second during words (and much lower than 1000 when averaged over many words.) A radar receiver with a bandwidth optimized for the pulse widths employed and simple "decoding" equipment at audio constitute the receiver. Bandwidths can be made large so that frequency stability is no problem. As long as

the bandwidth is adjusted to match the pulse width, such a microwave link has a sensitivity dependent only on average transmitted power and not on bandwidth. It is not difficult to show that system having 0.2 microsecond pulses and an IF bandwidth of perhaps 10 megacycles is just as sensitive as one using square-wave modulation of clipped speech with a bandwidth of perhaps 10 kilocycles (which would be very difficult to realize at microwave frequencies) for the same average power, antenna gain, and receiver noise figure.

Suppose a two-way microwave link (utilizing short pulses) is built so that when a transmitter pulses, the associated receiver is shut off (a T-R device). Then because the duty ratio is small, it is possible to carry on two-way simultaneous speech at the same radio frequency. This is not possible with conventional gear.

It can be added that a radar beacon makes an ideal repeater for pulsed speech as described.

Microwave communications systems are desirable, not only because of the radio spectrum space available, but because of the large antenna gain realizable with small antennas. However, such systems have not become common because clipped speech has not been fully exploited. With normal speech, it is necessary to amplitude or frequency modulate in the normal manner. Without quite complex gear, frequency stability problems make the optimization of bandwidth (10 kcs for AM) difficult at microwaves with the result that sensitivity suffers. Further, the magnetron and the disc-seal triode, which are the simplest, cheapest, and most efficient microwave power generators, work best as radio-frequency pulse generators, not as continuous devices. Existing pulsed systems are complicated and relatively inefficient for a communications system and are hardly suited for other than large-scale fixed installations.

Finally, a pulsed system is particularly suited to the incorporation of effective squelch devices (to combat receiver noise) without taking an excessive penalty in the form of reduced sensitivity.

5. MINIMUM PULSE RATE

The necessary pulse rate during words in order to transmit both positive and negative zero crossings of a speech waveform is about 3000 per second. With base clipping and reconstruction, the rate can be reduced to 1500. It can be reduced further if weak speech sounds are not transmitted. All this can be done without lowering articulation scores below about 95 percent.^{1,3}

It is evident that a study of base-clipped speech is of scientific interest in connection with the information content of speech. The question then arises as to the possibility of reducing the pulse rate considerably below 700 or 800 per second without sacrifice in intelligibility. In other words, experiments with clipped speech offer promise of knowledge of the "true" information rate that can be associated with speech without complete loss of information related to the voice characteristics of the particular speaker.

Evidently, only zero crossings of a speech waveform in one of the two directions need be transmitted; a reconstruction is possible without significant loss by generating a symmetric square waveform for each received pulse. It may be possible to go considerably beyond this by transmitting only every second, third, or fourth positive-going zero crossing and reconstructing the speech by generating two, three, or four symmetric square waveforms respectively for each received pulse.

If this can be done, the average transmitted pulse rate during words can be reduced considerably. No experiments of this nature have been performed.

Intuitively, it appears that the pulse rate can be reduced to perhaps 400 per second in this manner. No more than this can be said without experimentation. However, experiments with interrupted speech furnish a plausibility argument.⁴

It may also be possible to significantly increase the fidelity of clipped-speech transmission (using, perhaps, 1000 pulses per second) through slight repositioning of the pulses relative to the zero crossings. One method for increasing naturalness was mentioned, that is, inverse gain control at the receiving end of a system. There undoubtedly exist other strategies that might be evoked. For example, the biased clipping level might be made function of the undistorted speech waveform.

The characteristics of a clipped-speech signal are quite similar to a Morse Code signal. It would appear that concepts related to clipped speech furnish another tool which helps to correlate the psychophysical theory of an observer listening to a pure tone in noise and one listening to speech in noise.

Clipped speech allows the maximum efficiency of electronic devices to be realized. For this reason, it may have physiological applications. When in the form of short pulses, speech begins to look like neural impulses, which conjures up many interesting possibilities.

6. PRACTICAL CLIPPING CIRCUITRY

Pre-emphasis is easily obtained in a speech amplifier by means of R-C coupling network having a relatively small time constant. The circuit employed by the author consisted of a 12AX7 tube with a 100K plate load resistor, a 200 μ f coupling capacitor, and a 100K grid-leak resistor in the following stage. Because of the low value of resistance, the grid-leak resistor can be a gain-

controlling potentiometer without introducing problems from Miller and stray capacitance. If the pre-emphasis circuit appears after the speech signal has undergone considerable amplification, the hum problem is almost absent.

Pre-emphasis can also be obtained with a parallel-resonant circuit as a plate load; if the Q of the circuit is about unity, it becomes a very effective differentiator.

In order to minimize noise, the bandwidth of the speech system should be as small as is reasonable, perhaps about 5 kilocycles per second. Coupling networks other than the pre-emphasis circuit can have cutoff frequencies on the order of 200 cycles per second. A typical gain characteristic including pre-emphasis is shown in Fig. 9. The dashed curve in this figure is applicable when more pre-emphasis than that giving 6 db per octave is employed; then the speech

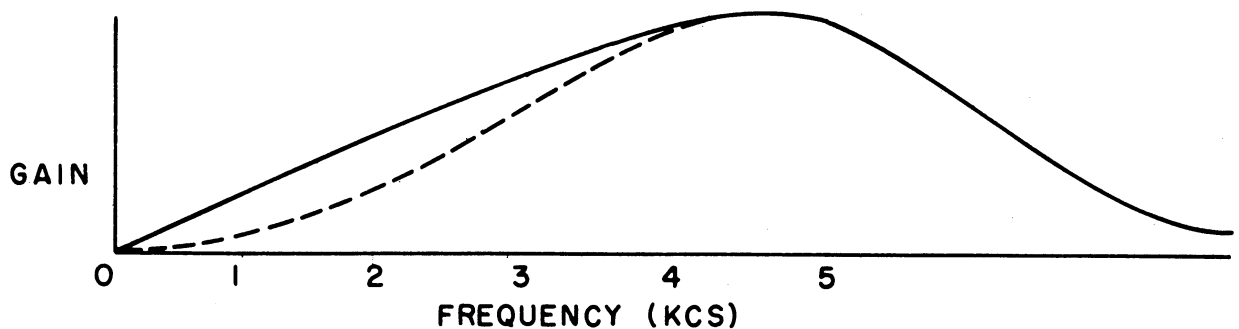


FIG. 9. TYPICAL AMPLIFIER CHARACTERISTIC

amplifier is more like a band-pass amplifier.

A circuit giving adjustable biased rectification is shown in Fig. 10. Tubes subsequent to the rectifier need handle only pulses and hence need not be linear. The first 6AL5 in Fig. 10 is the rectifier; the second 6AL5 is a clamp. The last stage employs grid clipping to help square the waveform. The main disadvantage of the circuit is that the clipping is not as sharp as might

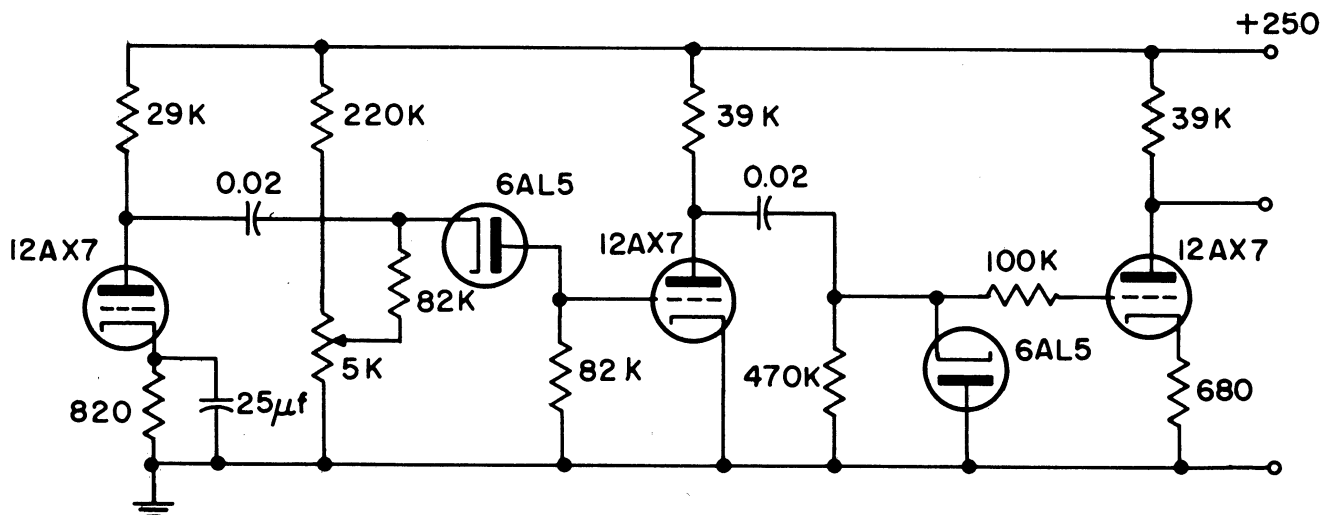


FIG. 10. BIASED RECTIFICATION CIRCUIT

be desired, particularly if circuits with any amplitude sensitivity are subsequently driven. Diode detectors require fairly large signals to give really sharp clipping, which leads to saturation problems in the speech amplifier where speech signals of widely differing magnitudes must be handled. (It should be pointed out no attempt was made to optimize the system of Fig. 10.)

What is felt to be the best answer to the problem is rather unconventional; the speech signal drives a cathode-coupled multivibrator which becomes free running when the signal exceeds a specified level. The result is base-clipped speech which is modulated on a relatively high-frequency square wave, which can subsequently be detected. In this manner the clipping is sharp, all rise times are the same, saturation in the speech amplifier is minimized, and mechanization is greatly simplified. (Trigger devices like the Eccles-Jordan circuit have a disadvantage in that they suffer from a hysteresis effect.) The circuit is shown in Fig. 11 (which does not show a single 6AU6 microphone pre-amplifier). The simplicity of Fig. 11 should be noted.

After pre-emphasis in Fig. 11, the signal is applied to the grid of a

12AX7 tube through a 1 megohm resistor. Grid clipping takes place so that the plate of the second tube in Fig. 11 swings only positive; in this manner, the

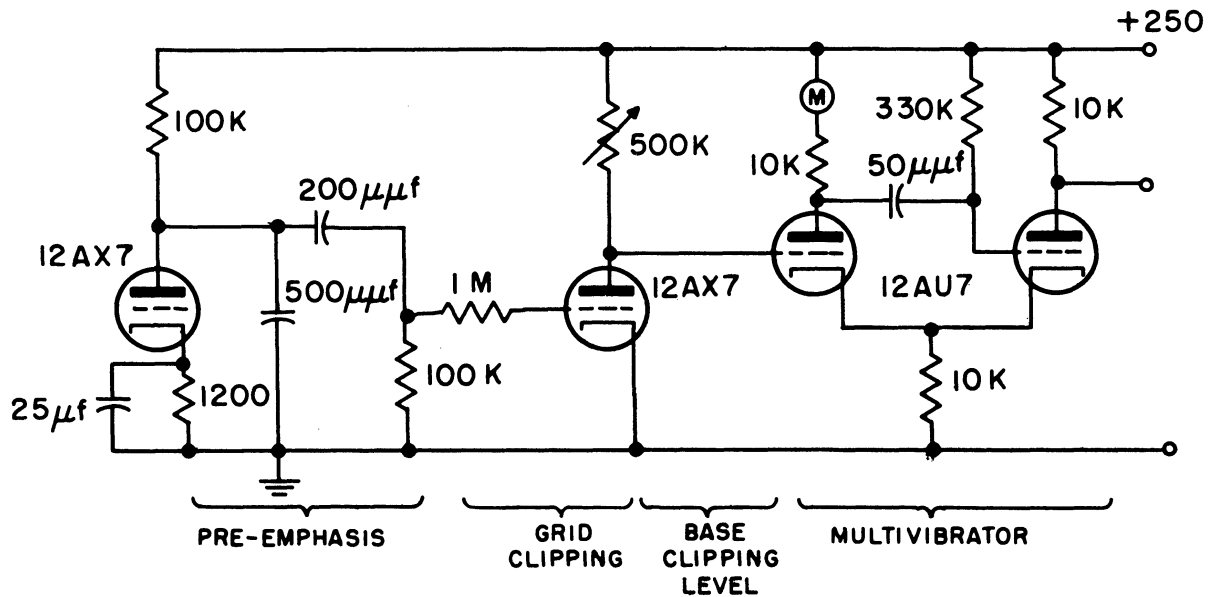


FIG. 11. MULTIVIBRATOR CLIPPING CIRCUIT

dynamic range of the part of the system that must handle speech prior to final base clipping is maximized. Without a signal, the potentiometer in the plate of the second tube is adjusted so that the multivibrator is not operating. Signals raise the plate voltage of the second stage causing the multivibrator to free run. The base-clipping point is controllable with the potentiometer. The meter in the plate circuit of the multivibrator (the normally off tube) makes a convenient and effective speech level meter.

The system of Fig. 11 is not intended as a final design, although it has only a single feature that might be improved; the free-running repetition rate of the multivibrator is not as high as might be desired. However, this poses only a minor circuit problem; for example, a triggered sine-wave oscillator might prove to be more suitable than a free-running multivibrator.

Mr. Reis was unfortunate. Had he, by accident, used a resonant diaphragm in order to realize pre-emphasis, he, rather than Bell, might have been called the father of the telephone.

8. SPEECH PRESENTATION

The method of presenting the base-clipped speech to the observer varied somewhat in the various experiments. The most direct method was that of applying the base-clipped speech directly to an audio amplifier driving either a speaker or earphones. An integrator could be used if desired. With the clipping bias at zero, ordinary clipped speech resulted. The bias adjustment permitted the behavior of the squelch mechanism to be observed as well as some characteristics when the weaker speech sounds were removed. For very large biases (few pulses), the speech becomes very "granular," although it is not entirely unintelligible.

Another presentation method was that of causing the differentiated base clipped speech (which gives a sequence of short pulses) to trigger a monostable cathode-coupled multivibrator. Only zero crossings in one direction would cause a trigger. The output pulse of the multivibrator, which was integrated and applied to speaker or phones, could be adjusted in width from about ten microseconds to several hundred microseconds. The system then begins to approximate a microwave communication link. With this equipment, the integration effect was noted. Also, approximate speech restoration could be realized. The intelligibility seemed to be about the same as that of clipped speech regardless of the pulse width. The integrator consisted of a one-, two-, or three-section R-C low-pass filter. The number of sections did not appear to be as important as the bandwidth (on the order of a hundred cycles). By proper adjustment of the R-C

filter, the voice pitch could be made the same as that of undistorted speech and hence added to naturalness and speaker recognition.

In all cases, listening with earphones seemed to give "better sounding" speech than with a speaker. This probably has something to do with the acoustic path and coupling of the sound, including room resonances, etc. There is no reason to assume that the acoustic properties of a speaker in a baffle and in a room are as suitable for clipped speech as for undistorted speech. In particular, it might be surmised that a speaker that transmits low-frequency sounds would not be too suitable because of the rumble resulting from widely spaced pulses. Earphones do not transmit frequencies much below two hundred cycles per second. It would therefore appear that the best speaker for clipped speech should be small (and hence cheap).

REFERENCES

1. J. C. R. Licklider and Irwin Pollack, "Effects of Differentiation, Integration, and Infinite Peak Clipping Upon the Intelligibility of Speech," Jour. Acous. Soc. of America, Vol. 20, No. 1, January, 1948.
2. Radio Amateur's Handbook, American Radio Relay League, West Hartford, Conn. (any recent edition, for example, 1952).
3. Many of the conclusions and statements in this report are based on qualitative observations from experiments performed by the author.
4. G. A. Miller and J. C. R. Licklider, "The Intelligibility of Interrupted Speech," Jour. Acous. Soc. of America, Vol. 22, No. 2, March, 1950.
5. Encyclopedia Americana, 1953 Edition, Americana Corp., Chicago, pp 374-378. (See "Telephone")
6. G. A. Miller, Language and Communication, McGraw-Hill Book Company, New York, 1951.
7. A. Chapanis, W. R. Garnes and C. T. Morgan, Applied Experimental Psychology, John Wiley and Sons, New York, 1949.