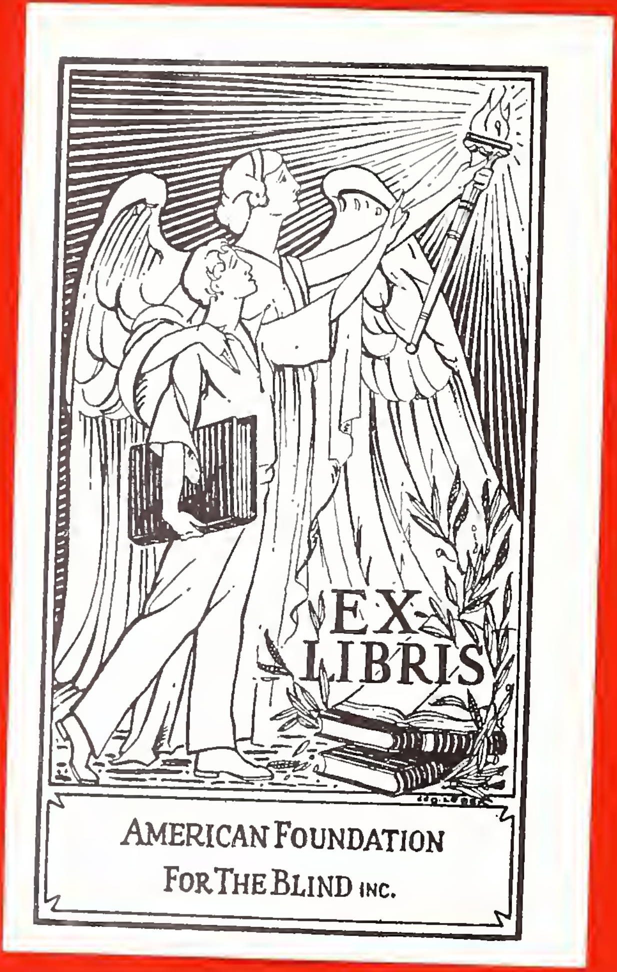
RESEARCH BULLETIN

NUMBER 9 APRIL 1965



The RESEARCH BULLETIN
is a publication of
The American Foundation for the Blind
15 West 16th Street
New York 11, New York

Leslie L. Clark, Editor
International Research
Information Service

This publication was supported, in part, by a research and demonstration grant, number RD-1407-S, from the Vocational Rehabilitation Administration, Department of Health, Education, and Welfare, Washington, D.C. 20201.

PREFATORY NOTE

The *Research Bulletin* of the American Foundation for the Blind is intended to be a means of publication for some scientific papers which, for a variety of reasons, may not reach the members of the research community to whom they may prove most useful or helpful. Among these papers one may include theses and dissertations of students, reports from research projects which the Foundation has initiated or contracted for, and reports from other sources which, we feel, merit wider dissemination. Only a few of these find their way even into journals which do not circulate widely; others may never be published because of their length or because of lack of interest in their subject matter.

The *Research Bulletin* thus contains both papers written especially for us and papers previously published elsewhere. The principal focus may be psychological, sociological, technological, or demographic. The primary criterion for selection is that the subject matter should be of interest to researchers seeking information relevant to some aspect or problem of visual impairment; papers must also meet generally accepted standards of research competence.

Since these are the only standards for selection, the papers published here do not necessarily reflect the opinion of the Trustees and staff of the American Foundation for the Blind.

The editorial responsibility for the contents of the *Bulletin* rests with the International Research Information Service (IRIS) of the American Foundation for the Blind, an information dissemination program resulting from the cooperative sponsorship of the Foundation and certain scientific and service organizations in other countries. In the United States financial assistance is provided by the Vocational Rehabilitation Administration of the United States Department of Health, Education, and Welfare, and by certain private foundations.

Since our aim is to maximize the usefulness of this publication to the research community, we solicit materials from every scientific field, and we will welcome reactions to published articles.

M. Robert Barnett
Executive Director
American Foundation
for the Blind

CONTENTS

SIMPLE READING MACHINES FOR THE BLIND*

M.P. Beddoes
University of British Columbia
Van Couver, Canada

INTRODUCTION

Research into reading machines for the blind is currently attracting much attention. Research workers, seriously involved with the problem, include many already famous men and interdisciplinary co-operation is extended between psychologists, electrical and mechanical engineers, medical surgeons and neurosurgeons, and physicists. A fairly substantial measure of success will be needed to maintain this widespread interest at such a high level. As one eminent engineer recently remarked: "I shall spend four years in this field; but I'm waiting to see whether or not my time will prove to have been wasted." The workers in this field fall into two classes: the doubters and the others, with the doubters, at present, the more numerous.

Simple machines exist at the present with roots back to 1914. One machine, the Optophone, has enjoyed a qualified (2) success in England and machines made on the same principle are being experimented with in America (the Battelle Optophone). The results of experiments on both sides of the Atlantic indicate that the Optophone's upper speed is probably 60 words a minute. This rate can be achieved only after a prolonged training period and the subject must be extremely gifted. The most successful manipulator of the Optophone is Miss Jamieson. She is a prodigy in this field, and she has worked with this machine for most of her life. She remarked to the author last summer: "I use the Optophone mainly to read back my typewritten letters: for this it is invaluable. I generally read five pages a day from a novel; more than this tires me."

The last sentence is significant. Sighted people can, with comparative ease, read a 200 page novel in a day. The contrast with Miss Jamieson's performance indicates the order of magnitude still separating the blind from the sighted in the matter of reading.

---

* Reprinted from The Engineering Journal, Vol. 46, No. 5 (May 1963), pp. 50-52.

A comparison between experiments done by the author at the University of British Columbia with some experiments done by Clowes (3) at the National Physical Laboratories, England, shows that a substantial increase in speed is to be hoped for from the Optophone if its present sound code is changed to a multidimensional code called Tonal Morse. This is the main thesis of the paper. Tonal Morse originated with the author (2) and its performance in this context is the subject of continuing work using a machine similar to that shown in Figure 2. The stage has been reached where some intelligent predictions can be made as to the performance of Tonal Morse with an Optophone print reader.

The paper also describes results of recent experiments done at the University of British Columbia (UBC) with a code called Spelled Speech Code. This Code originated with Metfessel (8) of the University of California: the work at UBC produced Spelled Speech by a method different from Metfessel's. Unfortunately, Spelled Speech requires a very complicated operation to be made on the print information (by a 'letter recognizer') and a machine using this code will be more complex and expensive by an order of magnitude than the Optophone. The experiments with Spelled Speech are quoted mainly because they reinforce the promise of Tonal Morse operating with the Optophone.

DISCUSSION OF READING MACHINES

Existing reading machines are classified into types: a) Direct Translation; b) Letter Recognition.

A third class has been proposed 'shape recognizers,' which lies in complexity between these two. A schematic showing the genesis of reading machines is given in Figure 1. All machines are basically the same in their method of obtaining print information. They all scan along a line of print a letter (or portion of a letter) at a time.

The simplest machine, the Optophone (2) (invented 1914 in the UK) and Argyle's Reading Machine (6) (invented 1952, Vancouver), are direct translation machines: they scan a narrow vertical slit which moves along the line of print; a series of very simple decisions are made by the machine and a sound is produced. The effort of reading is very great.

A more complicated machine, suggested by Mauch in 1958 (5) works with a shape recognizer. In this machine, the print information is processed by the machine so that various shapes, e.g. straight lines, circles, closed loops etc., are recognized. Each shape then triggers an appropriate sound, and the statistics of the appearance of the shapes is matched to the code so that noises are produced which are 'speech-like.'

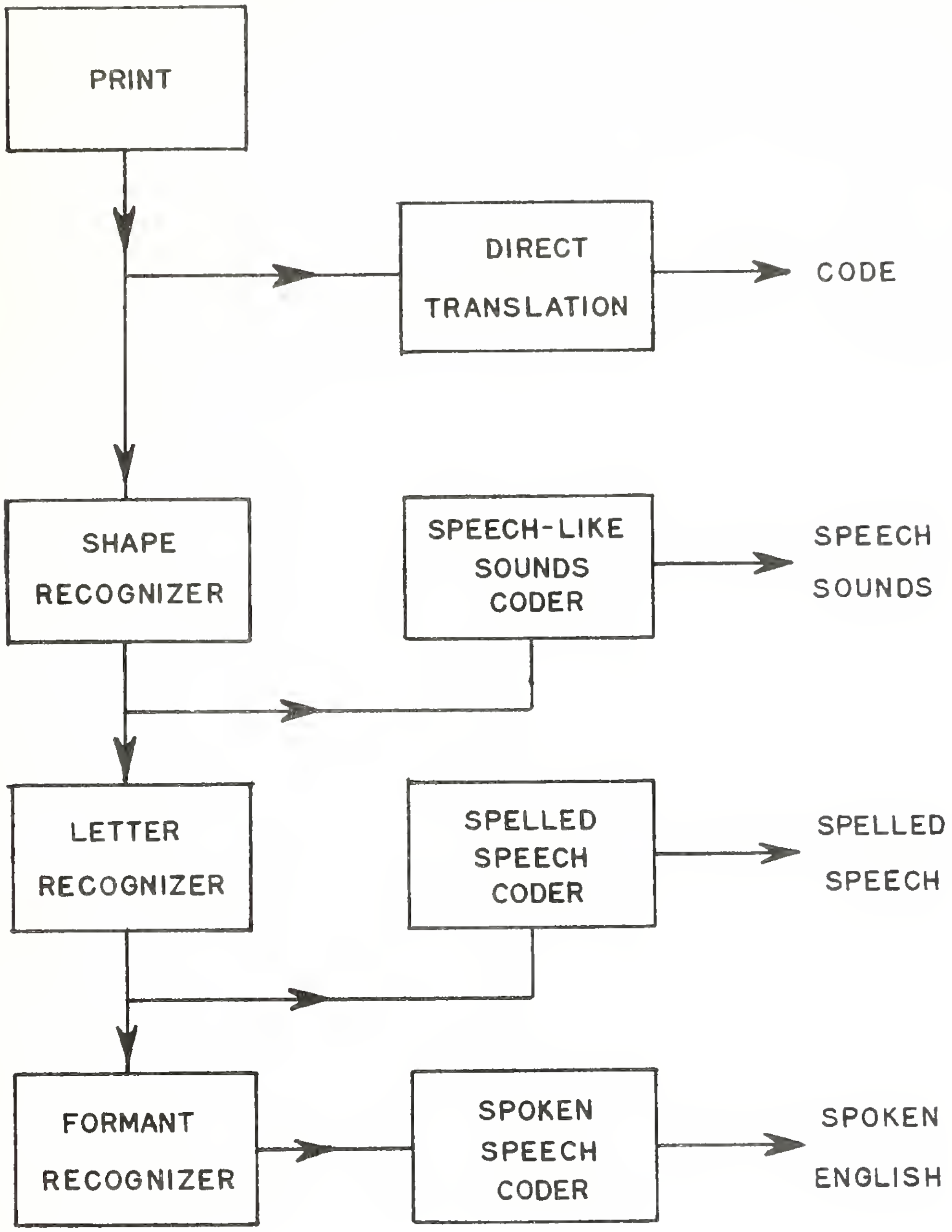


Figure 1. A Classification of Reading Machines
(After Dr. F.S. Cooper Haskins Laboratories, N.Y.)

A more sophisticated operation on the print information (a letter recognizer) produces letter information (4, 7, 10, 11). A suitable code would be Morse. Early work by the writer in 1959 (2) resulted in Tonal Morse which was claimed to be more suitable for blind reading machines. But subsequent work by the writer and independent results of Metfessel of the University of Southern California have shown that artificially produced Spelled Speech, for this particular usage, is a more suitable code. Briefly, Spelled Speech is a means of conveying print information by spelling it aloud a letter at a time as in the elementary grades of school. For example, the message "the cat who sat..." would sound as follows: "Tee Aitch Ee: Cee Aye Tee: Double-you Aitch Oh: ...." Each alphabet sound can be time compressed either by using a variable speed tape recorder or by other means. Such artificially time compressed Spelled Speech is easy to comprehend, after negligible practice, even at fast rates, and the instrumentation of a coder is quite simple.

Given a letter recognizer, it is possible to obtain the spoken output. Work reaching towards this goal is progressing mainly at the Haskins Laboratories, N.Y. under Cooper, et al. Instrumentation corresponding to the boxes designated in Figure 1 as "Formant Recognizer" and "Spoken Speech" is very considerable and the cost would justify this approach for a large library only.

The machines shown in Figure 1 and explained briefly above are those on which most work is being done. A notable feature is the emphasis on aural communication. Some pilot work is being done exploring tactile and electrical channels (notably at Massachusetts Institute of Technology and the National Physical Laboratories, UK).

EXPERIMENTAL STUDIES OF SIMPLE MACHINES

Experimental studies comparing Argyle's reader, the Optophone and a tactile reader are reported by Clowes, et al. (3) working at the National Physical Laboratories, UK. Figures from their report will be contrasted with some results taken using a Tonal Morse machine at the UBC.

The Argyle reader (Figure 2) was used in a simulated form and the apparatus is shown schematically in Figure 3. The character to be read is imaged on a rectangular screen, *s*, for monitoring purposes: the light from the screen is scanned by a rotating disk, *D*, which contains (in the case of Argyle's original machine) eight holes equispaced at a constant radius. One hole at a time falls in the field of the letter and if the letter is stationary with respect to the pick-up device, the path traced out by the scanning hole is the arc of a circle shown in aperture *s*. High intensity variations in the object plane produce corresponding changes of signal from the monitoring photocell *P*. Let-

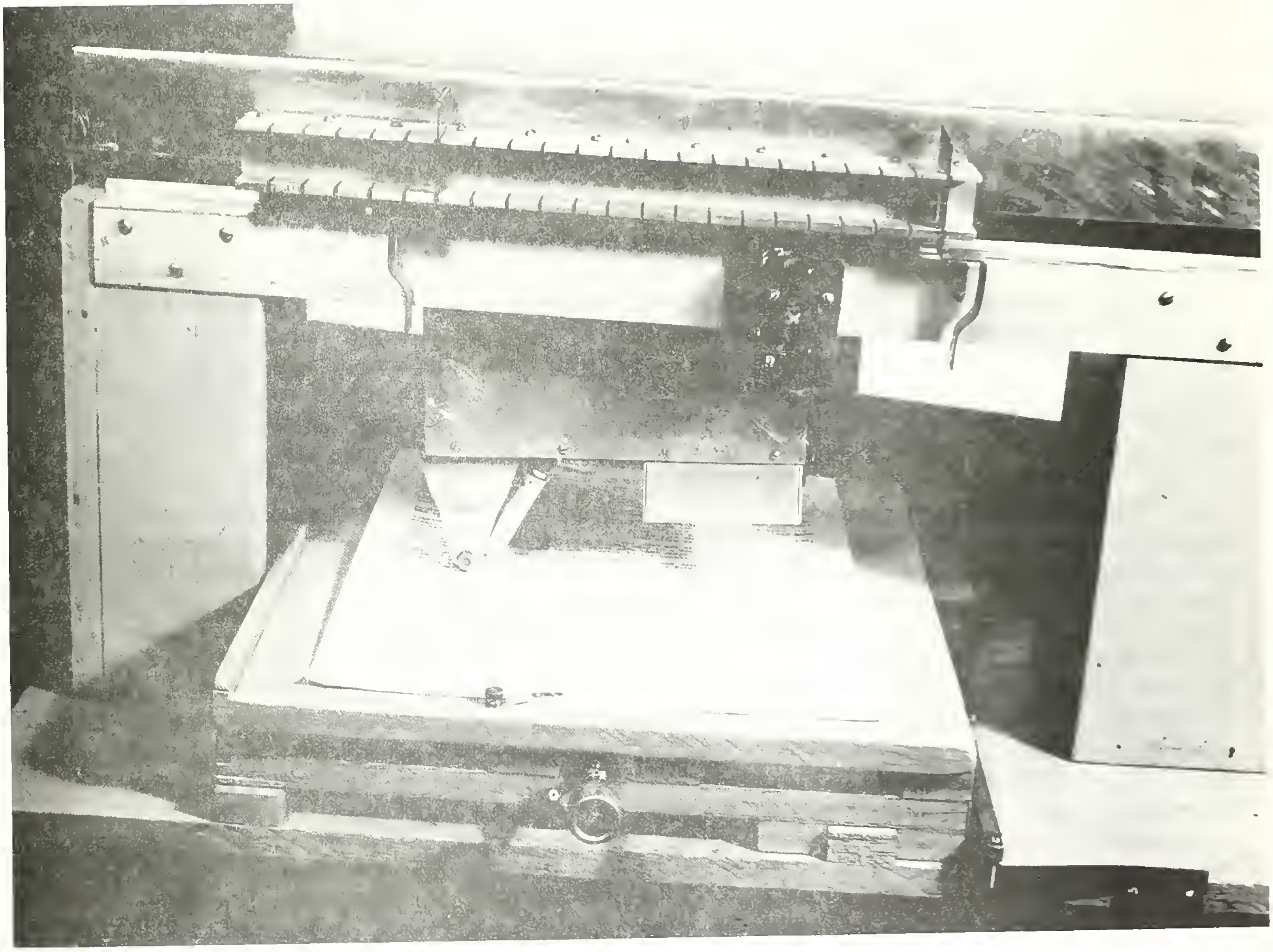


Figure 2. The Argyle Reader.

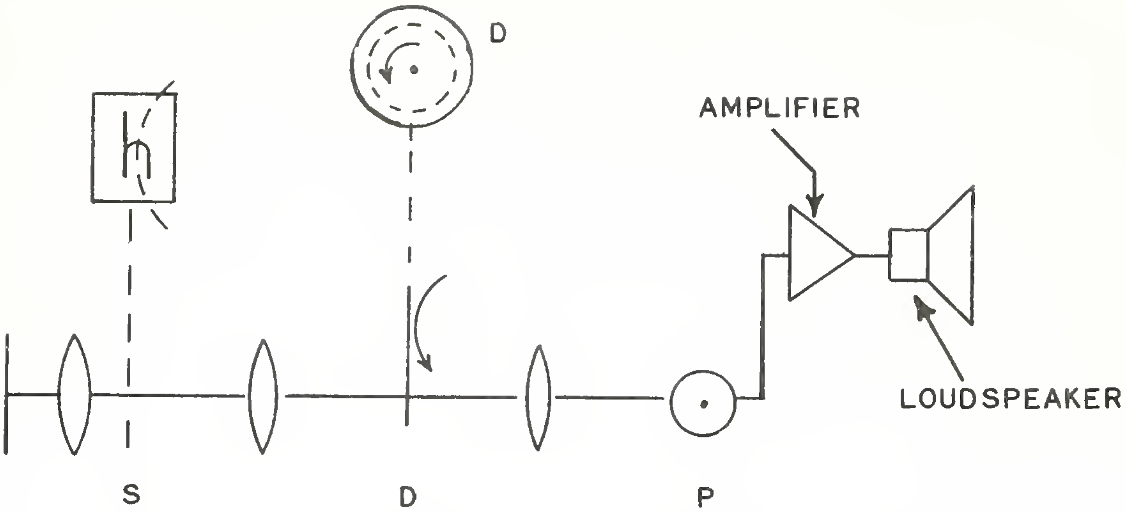


Figure 3. Simulator of the Argyle Machine. (Not to Scale)

ters are identified by their characteristic spectrum.

The Optophone is described elsewhere but, briefly, the feature of using a slit which moves across the letter is common to both Argyle's reader and to the Optophone. The latter avoids the scanning mechanism by using a number of photocells mounted along a vertical viewing slit, and the output from each cell varies the amplitude of a square wave of characteristic frequency.

Tactile presentation consisted of embossed letters on a heavy paper. The letter occupied about one square inch, and a succession of letters were mounted on a paper tape which was driven past the subject's finger tips.

The results of a short learning session with a small number of subjects (a total of 14 took part) with the three devices is shown in Figure 4. It will be noted that the speed is quite slow, corresponding to 18 to 25 words a minute. (These rates are based on the standard five letters to the word. They cannot be used to predict the learning rates in an actual reading situation because the tests were very short. They are used here to compare the performances of different codes.)

Learning sequences using a multidimensional code called Tonal Morse were obtained at UBC using 37 sighted subjects and 6 blind ones. The results are plotted in Figure 4. Tonal Morse consists of sounds from two sources: a variable-bandwidth noise generator, and a variable-pitch, variable-waveform tone generator. It has

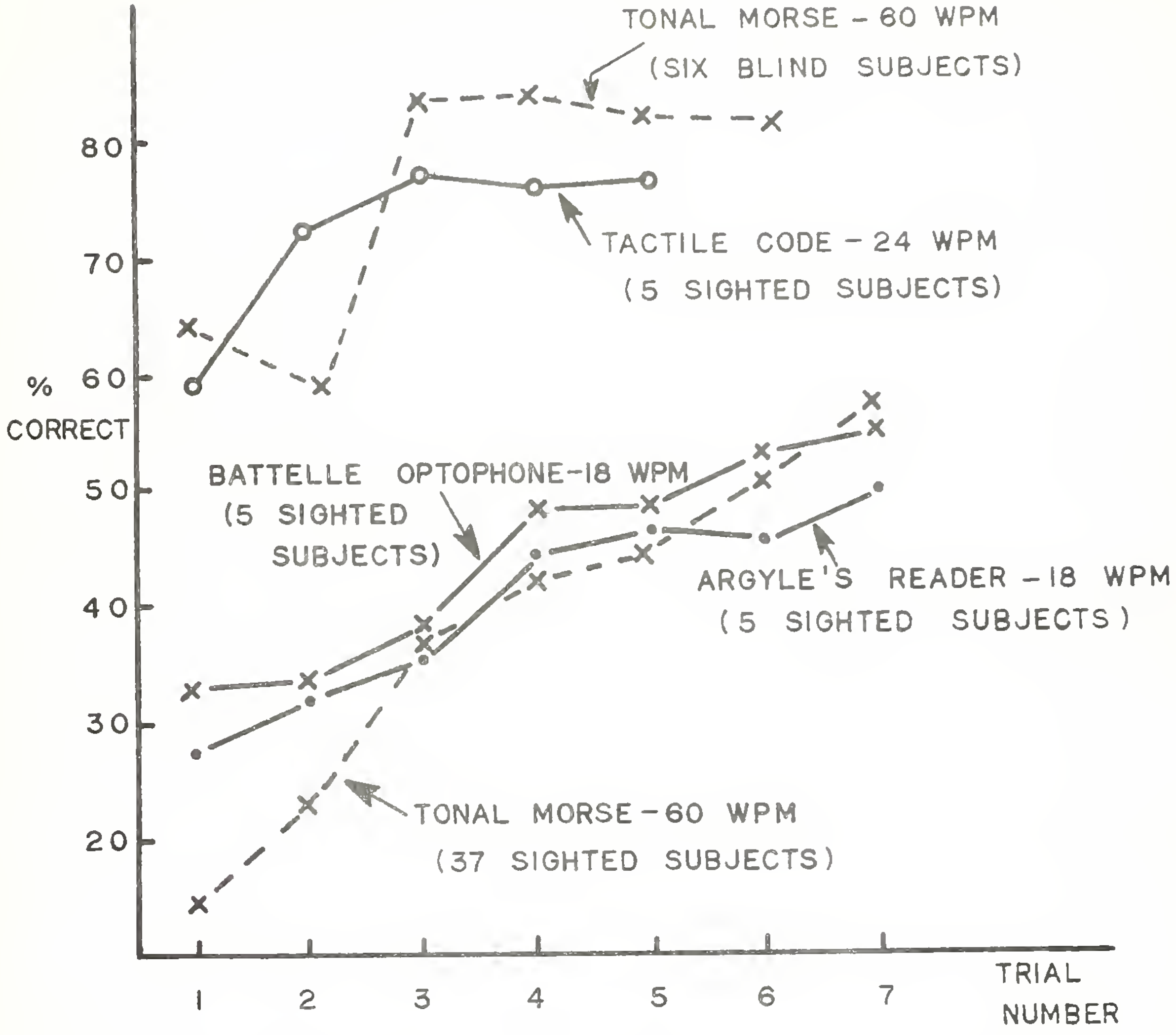


Figure 4. Comparison between Codes for Simple Machines.

been shown (2) that permutations of the above variables allow 63 distinct sounds to be produced which are well differentiated from one another, and each letter of the alphabet is allotted one characteristic code combination.

From Figure 4, it is seen that Tonal Morse with the sighted group gave much the same learning pattern as did the (sighted) subjects with the Battelle Optophone and Argyle's reader; but the reading rate exceeds the others by a factor of three. The performance using blind subjects was best of all, but this will not be used in the comparison between the various codes.

DISCUSSION OF RESULTS

The graphs shown in Figure 4 were obtained from two groups of people and it may, on this account, be unfair to use them to compare the codes. But, taking them on their face value, what do they indicate?

Obviously Tonal Morse can be learned just as well per trial, as the present Optophone or Argyle's machine and at three times the speed.

The tests with Tonal Morse assumed that a single sound was used to represent each letter: a letter recognizer would be required. The print reader in the Optophone, on the other hand, produces three characteristic signals for each print letter (in time sequence); and if the Tonal Morse generator were controlled by such a print reader, the rate of sounds to be decoded would be trebled and it might appear that the three to one speed advantage of Tonal Morse would be lost.

There is evidence that the present Optophone produces sounds (or clues) faster than the subject's decision rate. His slow performance is probably due to excessive decision periods needed to decode information from sounds which are very nearly alike. Some examples illustrate these points.

Corresponding to the letter 'h,' the Optophone produces three sounds: a full chord followed by a single note taken from the chord, followed by a chord made from the middle and lower notes. At first the subject will probably make a decision based on the chards which are loud and obvious. On their evidence, the letter could be an 'h' or perhaps a 'b.' Probably no further decoding is needed if contextual information can be used and only one decision needs to be made. But is this one decision easy to make?

Probably not. Early work with Tonal Braille (12) shows poor discrimination between two chords consisting of an open fifth and the major triad: this is supported by work by B. White of the Lincoln Labs., (private communication). One decision is needed for the letter 'h.' Of course, it could be argued, contrariwise, that this decision is one of a hierarchy of decisions such as the following: is this a chord I am listening to? is the top note of the chord present? is the next one present? etc. In fact in order to render the problem into a number of YES-NO decisions, a vast number of decisions would have to be made. In speech perception in particular, this sort of approach cannot be used by the human operator. He has slow reaction times of 0.2 to 0.5 seconds but, as Miller (9) notes, "the time required to decide between two alternatives is effectively the same as that required for 30 alternatives." Miller states also that one natural decision unit for speech is probably two to three words at a time. Thus in one form of aural communication, there is manifested a short term memory. This means that the speed limitation of decoding need not be set primarily by the number of sounds presented per second but by how easy these sounds are to decode after a number of them have been stored. A parallel between human decoding of aural signals and decoding long sequences of distorted signals can be drawn. Both require a memory. The larger this memory is the more effective and accurate the decoding becomes.

Corresponding to the Optophone's top speed of 60 words a minute, an average of 15 distinct sounds, chords and single notes, are presented to the subject each second. It is impossible to suppose that he can make a decision on each of these, and some discarding of information or, more likely, some short term storage is used to group the information into packets requiring a more leisurely decision rate for their decoding.

The great difficulty with the Optophone code is that some of the clues are nearly the same (e.g., the chords). This sort of difficulty can be avoided completely with Tonal Morse and the speed of this should be, consequently, much higher.

SPELLED SPEECH SOUND CODE

If a cheap letter recognizer can be produced, and there is strong evidence to the contrary (7), then Spelled Speech operating on a letter-by-letter basis performs in a much better fashion than any of the codes discussed in the previous section.

Basically, Spelled Speech is the stuff which used to be taught in schools and on this account it is a very well known letter code. Metfessel (8) produced a form of artificial Spelled Speech by first tape recording an alphabet, then by judiciously eliminating parts of the sounds, he managed to collapse the time scale by a factor of two or more. He obtained comfortable decoding rates up to 90 words a minute. The author (1) has achieved the same sort of performance using a variable-speed tape recorder, and he obtained top speeds of 120 words a minute. Subjects can learn to use Spelled Speech in a very short time possibly because of their previous schooling.

The work with Spelled Speech demonstrates an important point: that it is physically possible to assimilate at a very high rate printed information which is presented only one letter at a time. Thus, to the Times Square problem "How much of a sentence must be displayed at one time in order to read with ease?" the answer is one letter.

If the human operator is capable of storing at least one letter's information, when using the simple machine, then surely the performance with the well-designed multidimensional code can be made to approach that of Spelled Speech.

CONCLUSIONS

To sum up: in the very simple machine, use is made of various human capacities, e.g., short term storage, and ability to discriminate between a large number of alternatives, and if this ability can be exploited fully, then reading rates approaching those of Spelled Speech should be possible. The present simple machines are difficult to read because their sound output codes produce very nearly similar sounds. The human mechanism balks at the hurdle of decoding these sounds at high reading rates. It is believed that the substitution of a well-designed multidimensional code such as Tonal Morse will remove the hurdle.

Experiments with Spelled Speech indicate that letters presented one at a time can be decoded very rapidly. The results confirm the promise of the simple machine. To be practicable, Spelled Speech requires a letter recognizer and this will add an order of magnitude to its cost and complexity.

ACKNOWLEDGMENTS

The author thanks Dean Myers for suggesting that this paper should be written. He also acknowledges with pleasure the assistance of Edward Orme in the work with reading machines presently being undertaken at the University of British Columbia.

The work is financially supported by the National Research Council of Canada and by Mr. and Mrs. P.A. Woodward's Foundation.

REFERENCES

1. Beddoes, M.P., "Possible Uses of a Printed Braille Reader with Spelled Speech Output," in L.L. Clark (ed) Proceedings of the International Congress on Technology and Blindness, Vol. I. New York: American Foundation for the Blind, 1963, pp. 325-341.

2. Beddoes, M.P., E. Belyea, and W.C. Gibson, "A Reading Machine for the Blind," Nature, Vol. 190 (June 3, 1961), pp. 874-875.

3. Clowes, M.B., K. Ellis, J.R. Parks, and R. Rengger. Notes on Reading Machines. UK: National Physical Laboratory, February, 1961.

4. Clowes, M.B., and J.R. Parks, "A New Technique in Automatic Character Recognition," Computer J. Vol. 4, No. 2 (1961), pp. 121-128.

5. Emanuel, A.F., and H.A. Mauch. The Development of a Reading Machine for the Blind: Summary Reports. Dayton, Ohio: Mauch Laboratories, Inc., June 1959-1962.

6. Freiberger, H., and E.F. Murphy, "Reading Machines for the Blind," IRE Trans. Professional Group on Human Factors in Electronics, Vol. HFE-2 (March 1961), p. 18.

7. Kazmierczak, H., "Stage of Development of Automatic Character Recognition and Complex Reading Machines for the Blind in Europe," in L.L. Clark (ed) Proceedings of the International Congress on Technology and Blindness, Vol. I. New York: American Foundation for the Blind, 1963, pp. 261-278.

8. Metfessel, M.F. "Experimental Studies of Human Factors in Perception and Learning of Spelled Speech," in L.L. Clark (ed) Proceedings of the International Congress on Technology and Blindness, Vol. I. New York: American Foundation for the Blind, 1963, pp. 305-308.

9. Miller, G.A. "Decision Units in the Perception of Speech," IRE Trans. Inform. Theory, Vol. IT8 (February 1962), pp. 81-83.

10. Schiffman, M.M., "Sensory Aids Research," MIT Quart. Progress Report. Cambridge, Massachusetts: Massachusetts Institute of Technology, October 1962, pp. 251-258.

11. Shepard, D.H. Apparatus for Reading. Patent: 2663, 758, December 1953.

12. Zahl, P.A. Blindness. Princeton, New Jersey: Princeton University Press, 1950.

PERCEPTION OF APPARENT MOVEMENT FROM CUTANEOUS ELECTRICAL STIMULATION*

Robert H. Gibson
University of Pittsburgh
Pittsburgh, Pennsylvania

INTRODUCTION

Last year at these meetings, I reported conditions of pain-free electrical stimulation of the touch system. Brief (0.5 msec) pulses of direct current, when combined in short trains, at low pulse and train repetition rates, and delivered through sufficiently large electrodes, reliably arouse painless touch (1, 2).

However, since effective multiple stimulation of the touch system on some body surfaces (e.g., the back or chest) requires rather widely spaced electrodes, it might appear that the graded continuity and the potential complexity of electrically aroused cutaneous experience would seriously be limited to less than the spatial discriminatory power of the skin. The spacing is due partly to the necessarily large electrodes, and partly to the necessity to avoid spatial summation of otherwise subthreshold pain stimulation. This problem might be circumvented, given the following facts.

1) Electric touch stimuli, individually suprathreshold, when led *simultaneously* through two or more widely separated body sites, may arouse only a single "phantom" touch, at a position between the sites that varies with the relative stimulus intensities.

2) *Successive* electric stimulation of two or more widely separated sites, under certain conditions will bring reports of "apparent" movement of the "phantom" touch from one site to the other. This paper will be concerned only with variables relevant to apparent motion.

APPARENT MOVEMENT

The purpose of the present experiments was to determine conditions for the optimal arousal of cutaneous apparent movement with elec-

---

* This paper was read at Eastern Psychological Association meetings, April 11-13, 1963. The research was supported in part by grant NB-02022-05 from the National Institute of Neurological Diseases and Blindness to Carnegie Institute of Technology.

tric stimuli, and quantitatively to relate the "goodness" of the apparent movement to stimulus variables. Tactual apparent movement aroused by simple tactile, or by vibrating stimuli (3), has been reported by other investigators. Electrical stimulation has not been used, however, despite its measurement and control advantages. In fact, there has been little agreement of the stimulus properties optimal for arousal of cutaneous apparent movement. And there has been insufficient quantitative description of the properties of the stimulus dimensions involved.

Apparent movement in the present sense may be considered theoretically as merely a special case of movement perception. A pencil point dragged 10 inches per second over the skin surface sequentially excites receptors and groups of laterally connected receptors. No one knows what the upper speed limit is for such movement at which the perception of motion fails; we know little about the lower limit. It would be useful to know the transformations across stimulus dimensions that leave invariant a vivid perception of motion (where there is no motion in the stimulus). Such knowledge should thereby increase our understanding of the mechanisms by which motion on the skin is perceived. Also, the extent to which apparent motion on the skin and, for example, in the eye correspond or respond differently to changes over the same physical dimensions should reflect fundamental properties that characterize and distinguish them as spatial receptor systems.

Five observers were paid hourly to serve in the experiments. They were intelligent, careful people with low initial responsiveness to electrical stimulation of the skin. All were upperclass undergraduate engineering students accustomed to working with numbers, and familiar with imposing looking electrical apparatus.

The effects of three variables on the "amplitude" of perceived movement were determined in two experiments. These included:

1) distance on the body between two electrode sites

2) stimulus duration (pulse train length), and

3) time between stimulus onsets.

Direct estimation scaling procedures were used. O's were instructed to judge directly the "amplitude" (i.e., the impressiveness, or "goodness") of perceived movement from a given stimulation, and to report their estimate as a ratio to the amplitude of movement in the standard stimulus, called "10." Prior to the first session O's were familiarized with several stimuli, and given samples of stimuli that resulted both in impressive movement and in little or no movement. The standard was presented several times at the beginning of each session, and presented and identified after every five judgments.

Stimuli to be judged were delivered through electrodes arranged vertically from the shoulder down the right side of the back. Stimulation (with trains of pulses with 5 msec separations) was through two 15 mm diameter active electrodes, with a common large indifferent electrode on the right footpad. The "standard" movement was obtained from stimulation of the right dorsal forearc (with the same arrangement of two active electrodes, one indifferent electrode, as on the back).

In the first experiment, four interelectrode distances (2 inches to 16 inches), four stimulus durations, and three interstimulus intervals were employed. The results of this experiment indicated that the interstimulus interval, measured as onset time differences, had little effect, a surprising result contrary to effects reported from the use of vibrotactile stimuli (3). Thus, in the second experiment, seven interstimulus intervals extended and covered the appropriate range, the range and number of stimulus durations was increased, and two interelectrode distances were selected which had been found in the first experiment to represent the extremes of the effect.

All stimuli used at all loci were adjusted by each O to be equal in apparent intensity to single pulse stimulation at the shoulder site. Single pulse stimuli were set triple absolute threshold at that locus, a "moderately loud" stimulus.

In each of three sessions, all possible combinations of the three variables were presented randomly to each O, with the restriction that all combinations appeared in each third of the session.

Means of each O's median estimates were plotted on log-log coordinates separately as a function of each variable.

Figure 1 shows the effect of the linear distance between two sites stimulated sequentially on the vividness of the perceived movement. Beyond 4 inches, as the distance between electrode sites increased, the amplitude (or impressiveness) of the movement decreased. The data in this range are fitted by a power function (i.e., are considered linear in these logarithmic coordinates) with a slope considerably more gentle than minus one. From 4 to 16 inches doubling the distance between electrodes decreases the amplitude of movement by a proportion that is less than the physical decrease. Decreasing the distance to less than 4 inches, which is roughly electrical two-point threshold with these stimuli, essentially left the impressiveness of the apparent movement little changed.

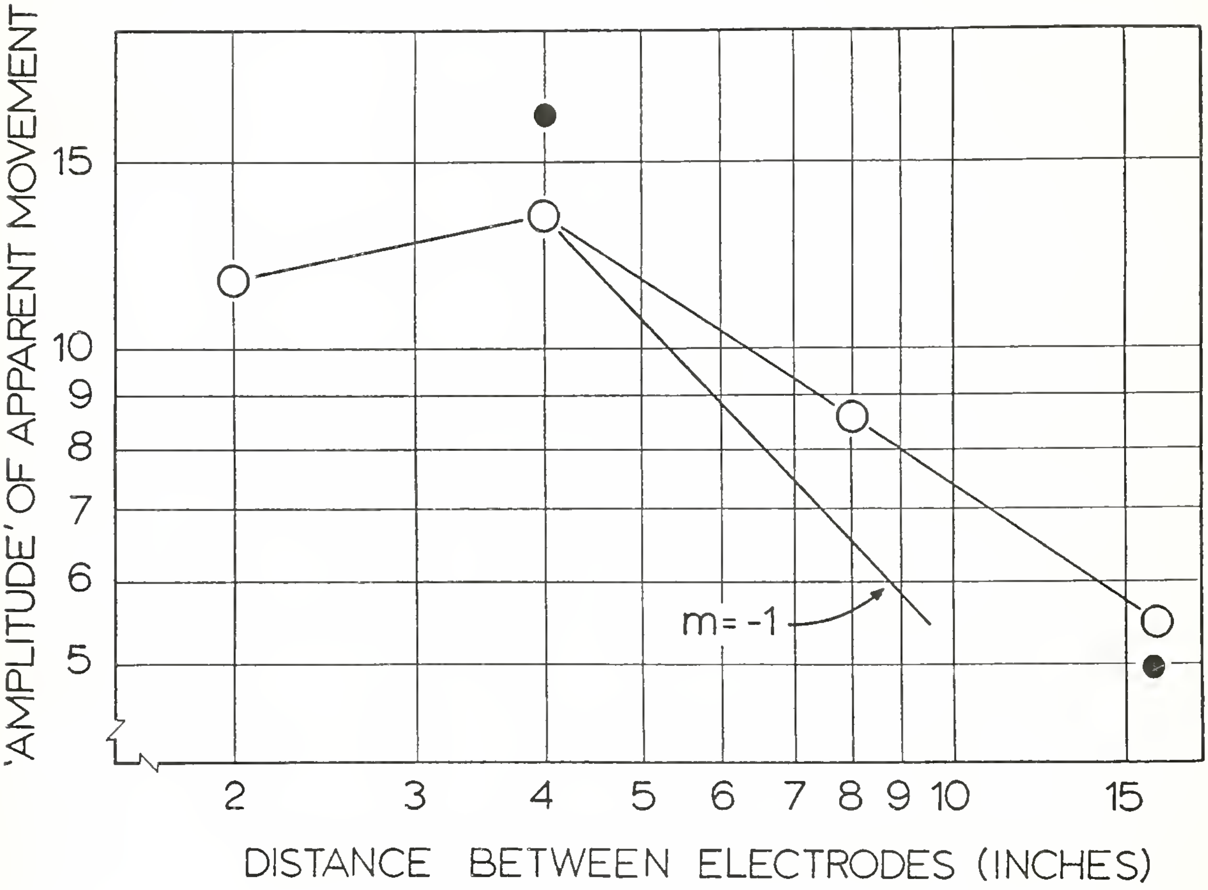


Figure 1. Direct estimates of apparent motion amplitude as a function of distance on the back between two electrodes. Open circles are means of five observers' median judgments, Expt. 1. Closed circles are values from Expt. 2, same O's summed across a greater range of stimulus duration and interstimulus interval. The line with a slope of -1 is for ease of visual comparison.

It is interesting to note that movement is reported at all distances less than that at which two simultaneously stimulated points are felt as a single point. The change is locus, even the direction of movement of a pencil point dragged only 1/8 inch or so along the back of the arm, is easily perceived, although this distance between points is not resolved when there is no temporal interval between them. For coding purposes, for example, the maximum usefulness of a given region is potentially greater than is indicated merely by stating two-point resolving power measures, presumably provided conditions are good for apparent motion perception.

Figure 2 shows the effect of time between stimulus onsets on judgments of the impressiveness of the apparent movement.

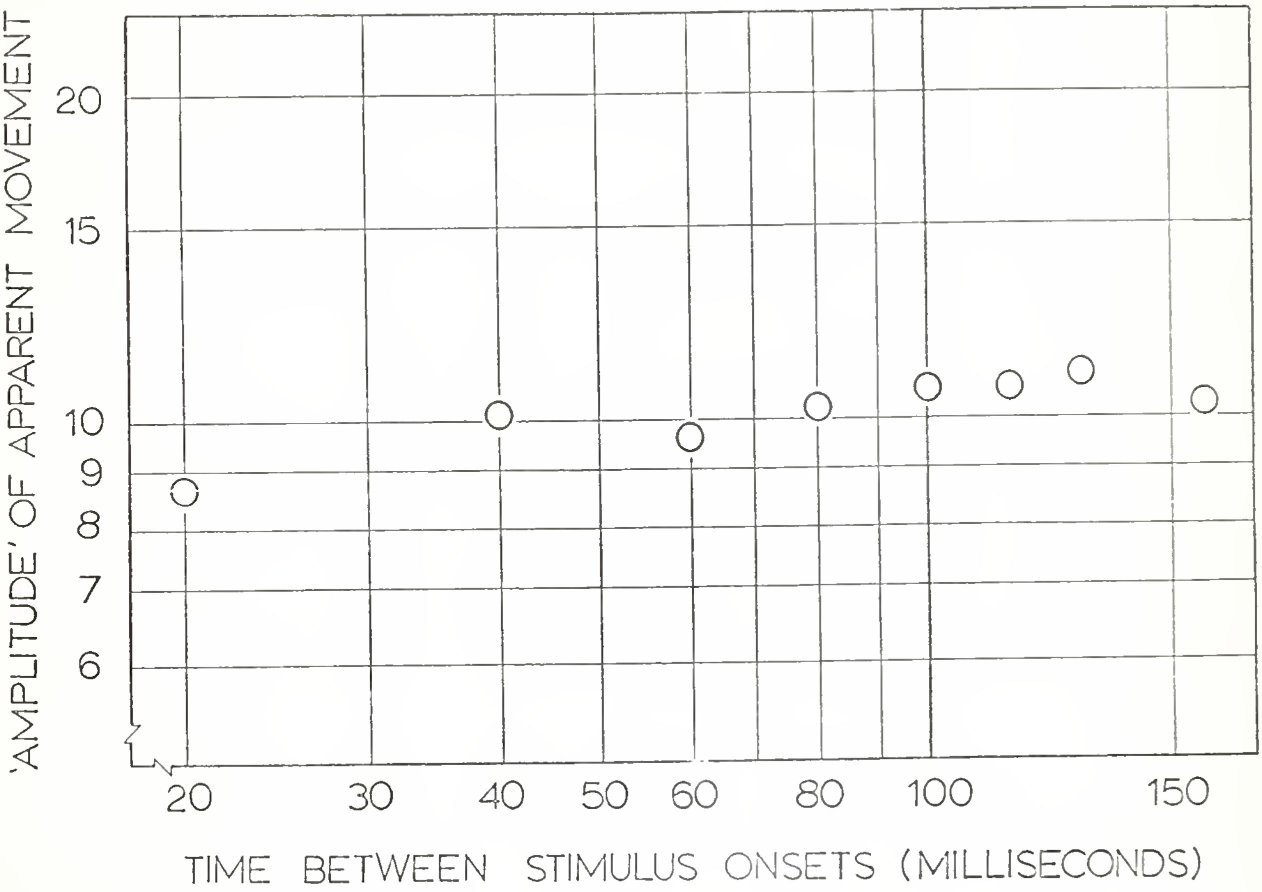


Figure 2. Direct estimates of apparent motion amplitude as a function of time between stimulus onsets. Points are means of five observers' median judgments.

Within the range of times used, the time between stimuli was obviously not a variable with a simple major effect. This is a result that is reliable with these procedures, contrary to Sumby's results with vibrotactile stimuli (3), and not in line with one of Korte's laws for visual apparent movement. (Discussion of this will appear in another paper.)

Figure 3 shows how the amplitude of apparent movement varies with stimulus (train) duration. Beyond roughly 20 msec, increasing the stimulus duration increases the amplitude of the movement. A portion of these data also are fitted nicely by a power function.

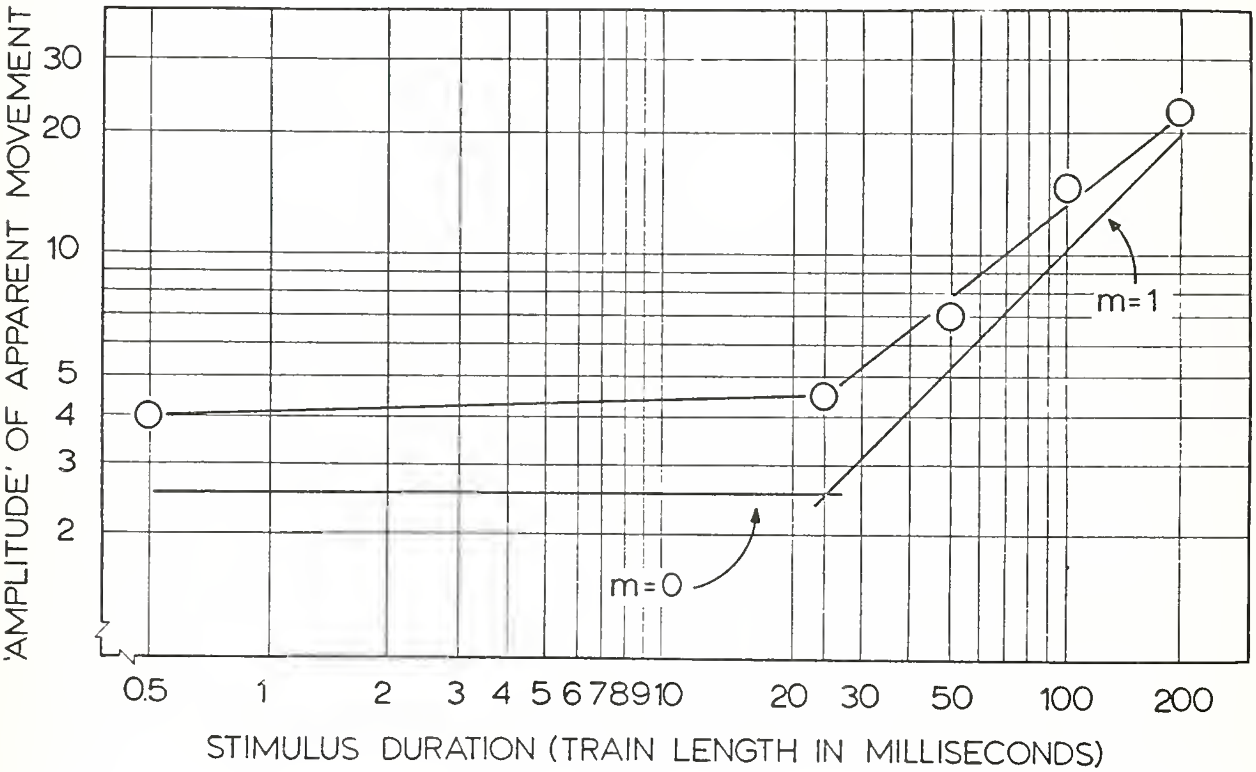


Figure 3. Direct estimates of apparent motion amplitude as a function of stimulus duration (measured as train length). Points are means of five observers' median judgments. Lines with slopes of 0 and 1 are for ease of visual comparison.

Decreasing the train length from 10 to 0.5 msec (a single pulse), when the stimuli are equally intense, does not affect the amplitude of the movement. This is as it should be, given that 10 msec represents the limit of the critical, integrating interval for the skin; successive events separate by less than about 10 msec are perceived as single events.

Since each of the five O's received every combination of delay, distance, and duration, an analysis of variance appropriate for such a repeated measures design was performed.* The significance of each treatment and interaction was tested against its interaction with subjects.

The principal aspects of the analysis are summarized as follows. Distance, duration, and their interaction accounted for the major portion of the variance. The interaction between duration and delay was significant, although its contribution to the variance was small. The distance *x* subject interaction was moderate; the interaction of duration *x* subject was greater. The distance *x* duration was significant. (At 4-inch interelectrode separation, varying duration from one extreme to the other, 0.5 msec to 200 msec, resulted in a range of apparent motion magnitude that was twice that from the same variation in duration at 16 inches separation. Duration, that is, has greater effect at smaller distances.) Individual differences were indicated also in a moderate distance *x* duration *x* subject interaction.

Figure 4 shows the relationships among these three variables. It is a schematic composite made from each of the individual functions essentially as fitted to the data points in the previous graphs. (For simplicity, a horizontal line represents interstimulus interval.) It is clear that the rate of *increase* in movement amplitude with stimulus duration increase is nearly the same as the rate of amplitude *decrease* with increasing distance. (The angles from horizontal at the break in the functions are nearly the same.) Double the distance between electrodes on the body and double the stimulus duration, within wide variation in interstimulus interval, and the impressiveness of the movement should remain essentially unchanged. (Such matching experiments have not yet been done.)

This procedure is useful for matching one variable against another. In this instance it is possible to determine, from these curves, the trading relations between time (stimulus duration) and distance on the body surface.

---

* The analysis was kindly performed for me by K. Kotovaky, Carnegie Institute of Technology.

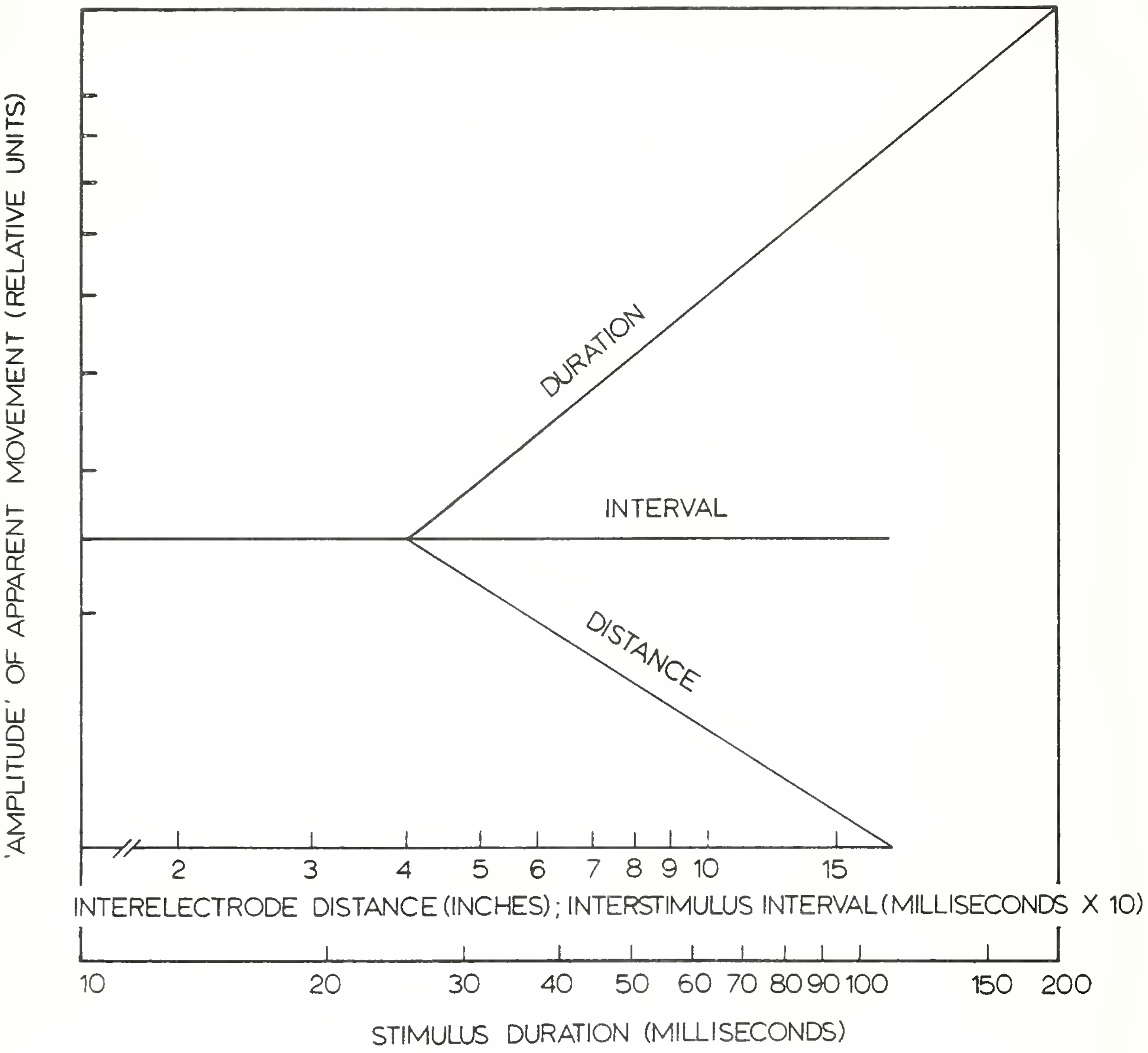


Figure 4. Composite of functions fitted to previous three graphs. Horizontal line is used for simplicity to represent effect of interstimulus interval. Ordinate is in log relative units to enable the curves to be put in registration at low stimulus values. All axes are to the same scale.

Thus, conditions for optimal apparent movement include using 4-inch electrode separation, long stimulus durations, and any moderate interstimulus interval greater than zero. It is also clear that if it were desirable to avoid perceptual interaction among separate stimulations of different body regions, it would be desirable to use distances of 16 inches or greater, coupled with brief duration stimuli, regardless of moderate changes in interstimulus intervals. Beyond a minimum, the time between onsets of overlapping trains of electric pulses, unlike with vibrotactile stimuli, seems to have little influence on the impressiveness of cutaneous apparent movement. Since these times can thus be made small without sacrificing the illusion of movement, upper limits for speed of pattern transmission are thereby materially raised.

In one sense, the recommendation is for (some of) the richness of visual detail (as in object detection), or possibly auditory speech, to be replaced by trains of frequency modulated cutaneous clicks suitably spaced around the body. We know little of the nature or possible extent of cutaneous imagery, nor how drab such might prove. The stimuli needed to find out have not previously been available. We have long known that the body surface can provide at least gross patterning. Now, finely drawn patterns seem feasible using tactual apparent movement as a major element. The primary learning problems might even just be partly circumvented by the fact that a younster has a major investment after only a few years of life in having learned anatomical distinctions on his own body. He knows where his elbow is, and the back of his knee, and has names for these regions.

REFERENCES

1. Gibson, R.H. Conditions of Painless Electrical Stimulation of Touch. Paper delivered at Eastern Psychological Association, April 26-28, 1962.

2. Gibson, R.H., "Requirements for the Use of Electrical Stimulation of the Skin," in L.L. Clark (ed) Proceedings of the International Congress on Technology and Blindness, Vol. II. New York: American Foundation for the Blind, 1963, pp. 183-207.

3. Sumby, W.H., "An Experimental Study of Vibrotactile Apparent Motion," Research Bulletin, No. 9 (April 1965), pp. 71-101.

SYNTHESIS OF ORIGINAL VOCAL PITCH IN ACCELERATED PLAYBACK SPEECH*

Jay Harold Ball
Bolt, Beranek, and Newman, Inc.
Cambridge, Massachusetts

INTRODUCTION

Object of the Thesis

Modern recording media, such as magnetic tape, discs, wire and film, have afforded us with an off-line means for increasing the flow rate of unilateral spoken information. Although playing back a recording of speech at a rate higher than that at which it was recorded cannot reduce the redundancy of speech, it does present us with the originally recorded information in a shorter time. We can, by so doing, circumvent some of the 70 percent temporal redundancy of speech (15).

In a unilateral communication system comprised, say, of a high-speed tape playback and a listener it is the inability of the latter to handle the high data output rate of the former which limits the performance of the ensemble. This thesis project was undertaken in an attempt to determine the factors underlying the loss of intelligibility in accelerated playback speech, and to determine whether or not a relatively simple electronic system could be constructed to aid the listener.

Motivation for the Thesis

The motivation for this research stemmed originally from a request by the Sensory Aids for the Blind group of the Research Laboratory of Electronics (Massachusetts Institute of Technology). Blind students who use tape recorders for note taking are hampered by the fact that, since they record virtually all of a given lecture, they must listen to the playback of a great deal of unimportant material surrounding the "core" information. They are unable, in other words, to take notes in the truest sense of the word, and the time required to listen to a day's recordings is comparable to the number of class hours spent in making them. To compensate, these students try to play back their tapes at a higher than nor-

---

* This publication is based on a thesis submitted in partial fulfillment of a Master of Science degree at the Massachusetts Institute of Technology, Department of Electrical Engineering.

mal rate. With practice, they can learn to understand speech recordings played back at up to 1.6 times the recording speed. At playback/record speed ratios greater than 1.6 or so, the intelligibility of the recordings drops so sharply that even a practiced listener cannot make use of them. This limit has been verified by the author and by Klumpp and Webster (23).

Information Assimilation Rate Limits

Let us compare the abilities of an observer to assimilate linguistic information via the visual and auditory modalities. An average reader can assimilate printed prosaic speech at a rate of about 300 words per minute, and a trained reader can handle upwards of 1000 words per minute. Since conversational speech normally occurs at a rate of 150 to 200 words per minute (disregarding pauses), there is no reason to believe that the auditory modality is saturated. Indeed, the blind students referred to above have been able to handle 250 to 330 words per minute. Why, then, does this seem to approximate the comprehension limit of a listener?

In point of fact, it does not. Fairbanks and Kodman (15) have demonstrated that, under certain conditions, speech can be comprehended at rates of 750 words per minute or more. The technique is described below.

CAUSES FOR LOSS OF INTELLIGIBILITY OF ACCELERATED PLAYBACK SPEECH

Introduction

There are three possible causes for the drop in intelligibility with increased playback speed. The first of these is that the duration of any given word or phoneme may become too short for correct perception. The second is that all of the frequency components are multiplied by a common factor, and may thereby be shifted too far from our normal speed frame of reference. Third, the frequency response of the playback equipment may grossly distort the signal because of mismatches between record and playback equalizations. While such improper playback equalization may in practice account for some of the loss of intelligibility, the effort required to elimimate this factor is trivial.

The problem now resolves itself into evaluating the effect upon intelligibility of shortened phoneme duration (and accelerated inter- and intraphoneme transitions, of course) and formant frequency shift.

Spectral Distortion

The steady state frequencies of the first and second formants

(F_1 and F_2) determine, to a large extent, what vowel is perceived (39). Reference to plots of F_2 versus F_1 (39, 42) show, for example, that doubling both formant frequencies would result in the perception of an entirely different vowel. It has also been demonstrated that the locus* of F_2 determines for a given vowel the perceived preceding stop or nasal consonant (11). Since the F_2 locus proper (rather than the ratio of the F_2 locus to the steady state value of F_2) is the critical factor in the correct perception of stops and nasals, one would expect rapid deterioration in the intelligibility of these sounds with increasing playback speed.

Temporal Distortion

Two effects of frequency shift upon the perception of speech when there is no temporal distortion are considered in the examples cited above. If the converse situation is considered - namely, time compression in the absence of frequency distortion - some very encouraging results are found. Through the use of sampling techniques, Fairbanks, Everitt and Jaeger (14) and others (27, 45) have managed to time compress speech with minimal spectrum distortion. In one study (14) of phonetically balanced (PB) words, compression ratios of up to 5 resulted in word intelligibilities of up to 95 percent. (Egan (13) has shown that a score of 84 percent on PB words is equivalent to a score of approximately 100 percent on sentences.) Denes (12) reported circumstances of nonspectral characteristics, such as duration, serving as the basis for phoneme recognition.

Ideal Method for Resolution of Spectral and Temporal Effects

The ideal method for determining the relative importance of spectrum shift and time compression would involve the use of a spectrogram playback, such as the ones described by Borst (5), Cooper (8), and Vilbig (49), which can generate the sound represented by a hand painted sound spectrogram. Such a device could scan a given pattern at any speed without changing the formant frequencies generated. It could also be swept at normal speed over a hand painted spectrogram whose formant lines were placed higher than normal on the frequency scale. Since a spectrogram reader was not available, recourse was taken to making intelligibility tests on lists of PB words which had been read at slower than normal rates and recorded. The details of these investigations will be found under "Intelligibility Tests" and "Results of Intelligibility Tests."

---

* The place on the frequency scale at which a transition begins, or to which it may be assumed to "point."

FREQUENCY DIVISION AND TIME COMPRESSION

Existing and Applicable Techniques

Since it is desirable to improve the intelligibility of speech recordings being played back at a higher than normal rate, it might well be assumed that restoration of the original formant frequencies would be advantageous. In other words, what is desired is a device or technique for performing frequency division of the time compressed speech. Frequency division has been accomplished in a number of ways but none, unfortunately, are sufficiently simple or economical for our blind student. Nevertheless, a few of the systems which do perform frequency division are examined below to see if any of the techniques they employ are applicable toward a simple, economical frequency divider.

Tape Sampling Time Compressors

The first system of interest is the multiple rotating head tape playback (14, 27, 45). This instrument transports the tape at higher than normal speed past a drum, on whose periphery are mounted several equally spaced tape playback heads. The drum rotates in the direction of tape travel, but with a peripheral velocity lower than the linear velocity of the tape. The net result is that, when a playback head is in contact with the tape, the *relative* velocity between head and tape is equal to the original recording speed. During the time any given head is "active," therefore, it is reproducing a sample of the tape at normal speed and therefore reproduces the originally recorded frequencies.

As the drum turns, the "active" head eventually loses contact with the tape. Since the tape is wrapped around the drum to some small extent, however, the next playback head makes contact almost immediately. Some of the tape has been skipped, though - namely, the portion between the end of the section which the first head reproduced and the beginning of the sample reproduced by the second head. In fact, if the ratio of the tape linear velocity to the original recording speed is, say, four to one, then a maximum of one-fourth of the total length of the tape can be scanned by the heads. The degree of tape wrap around the drum is adjusted so that as the signal from one head is fading out, the signal from the next is fading in. The net result is that the sequential, noncontiguous samples extracted by each head are, in effect, stretched and abutted.

This type of time compression (frequency division) system performs best when the sample reproduced by each head is of the order of 10 msec long - in other words, when the sample is short compared to the duration of the average phonetic element.

When one realizes the precision machining involved in the manufacture of a single playback head, it is clear why the cost of a multiple, rotating head assembly would be prohibitive for any student, blind or sighted.

String Divider

A second type of frequency divider is the tuned string system reported by Vilbig (50). In this device, an ensemble of taut steel strings, each of which is tuned to a different resonant frequency, is excited by the signal to be frequency divided. The strings, which act as mechanical bandpass filters, are slightly damped to broaden the passbands. Each is driven at a harmonic of its fundamental mode, and the drive and pickup mechanisms are arranged so that the output of any string is a subharmonic of its input.

This system does not fill our requirements either. It requires some rather complex driving and pickup mechanisms for each string, and if any string is excited by two (or more) frequencies within its passband, it will respond only to the one with the larger amplitude. The number of strings required to handle the bandwidth encompassed by speech signals, therefore, would result in a device prohibitive in both size and cost.

Below is an examination of some other systems which, although not expressly designed as adjuncts to a time compression system, do perform wideband frequency division.

Single-Side-Band Modulator

One of these is the system reported by Marcou and Daguet (34). It was designed to compress the speech frequency band by a large (10-100) factor in order to make the most efficient use of the limited range of frequencies available for radio broadcasting.

The speech signal was defined in terms of an analytic signal (21) $s(t)$, resolvable into the product of a time varying amplitude $a(t)$ and the cosine of a time varying angle $\cos[\phi(t)]$. By single-side-band suppressed carrier (SSBSC) modulating a high frequency carrier $\cos\,[\Omega t]$ with the speech signal $s(t)$ and limiting the resultant waveform, a constant amplitude signal $\cos\,[\Omega t + \phi(t)]$ was obtained. By using this signal to drive an "n-fold dividing circuit" (presumably a locked oscillator), a new signal $\cos\,[\frac{\Omega t + \phi(t)}{n}]$ was obtained. The new signal was also of constant amplitude, and had a frequency deviation $1/n$ times as large as that of the limited SSBSC signal from which it had been derived. This narrow band (small deviation) signal was the one actually transmitted.

By means of a harmonic generating "n-fold multiplier" at the receiver, the transmitted signal was converted back to the original limited SSBSC wave, with attendant deviation expansion. Demodulation (synchronous detection) resulted in a constant amplitude signal corresponding to $\cos[\phi(t)]$.

According to Marcou and Daguet:

> "If $\cos \phi(t)$ drives the loudspeaker, the output gives essentially the same aural sensation as the original signal $s(t)$. The intelligibility is completely conserved and the voice quality essentially unimpaired."

Presumably, we might frequency divide the limited SSBSC signal $\cos[\Omega t + \phi(t)]$ by 2 instead of some larger number, and demod- the resulting $\cos[\frac{\Omega t + \phi(t)}{2}]$ with $\cos[\frac{\Omega t}{2}]$. The result, according to the mathematics of Marcou and Daguet, would be a constant amplitude signal $\cos[\frac{\phi(t)}{2}]$. This is a signal in which the instantaneous frequency is one-half that of the original speech signal.

The prospect of so neatly accomplishing the desired objective is a tempting one indeed, but Kryter's (26) opinion of the quality of the Marcou-Daguet system did not jibe at all with their evaluation. Now it may be that this system, operated as described above, would result in highly intelligible frequency divided speech. Further work on the problem of frequency division should certainly give this technique closer scrutiny, but the disagreement as to the intelligibility of speech processed by this system, as well as the complexity of the equipment, motivated the author to turn his attention toward other techniques.

The VoBanC

A second bandwidth compression device is the Bell Laboratories' VoBanC, reported by Bogert (4). In this system, the speech is heterodyned up in frequency, and the lower sidebands are divided into three parts by means of crystal filters. Following each filter is a regenerative modulator (balanced modulator, bandpass filter, and feedback loop). The maximum energy component in the output of each regenerative modulator is at one-half the frequency of the largest amplitude component of the input, although the components adjacent to the maximum ones remain separated by the same amount as before.

The modulator outputs are bandpass filtered and demodulated with a frequency one-half that of the original modulating frequency. The demodulated signals are combined and the resultant summation signal is transmitted, with appropriate restoration techniques being applied at the receiver.

Although this system, like that of Marcou and Daguet, was designed primarily to reduce the required transmission channel bandwidth, it could be adapted to meet our frequency divider needs. It has the distinct advantage of having a fairly large (30-35 db) overall input-output dynamic range.

Further investigations of the possible application of the VoBanC to our frequency division problem were not carried out because of the expense and complexity of the system and - to be honest - because it was discovered late in the research program. The VoBanC should, nevertheless, be scrutinized far more closely in any further study of this problem.

Formant Trackers

There is a group of systems (7, 16, 18, 19, 28, 36, 48) which do not perform frequency division but which could be adapted to the purpose. These are the formant tracking devices whose primary goal is reduction of transmission channel bandwidth. The philosophy underlying all formant tracking systems is that the rate of change of frequency of any formant is small compared to the formant frequency itself. An ensemble of voltages is generated in which the instantaneous value of each voltage is proportional to the frequency of the formant being tracked. These voltages vary at a maximum rate of about 8 cycles per second.

The speech signal is analyzed for periodicity, and if it is detected the machine sends out a signal indicating the presence of a vowel. A different signal is generated for consonants. By transmitting this consonant-or-vowel signal and the formant tracking voltages instead of the original speech signal a considerable saving is realized in the required transmission channel bandwidth.

At the receiver, the consonant-or-vowel signal triggers a noise generator, while the formant tracking signals command either a set of voltage controlled oscillators or a set of voltage controlled bandpass filters fed with white noise.

A formant tracking system, as described above, could be used for frequency division purposes by simply adjusting each oscillator or filter at the receiver to generate or pass, for a given tracking voltage, a tone whose frequency was one-half (or any other ratio) that of the one which produced the tracking signal.

The problem again is one of complexity and cost, since formant trackers usually fill several six-foot relay racks! The technique is worthy, nevertheless, of a trial application to our frequency division problem.

There are two more possible techniques by which division of speech frequencies might be accomplished. To appreciate why the author actually undertook their construction, the reader must first have some familiarity with some of the properties of speech, especially as they relate to its perception.

INFINITELY CLIPPED SPEECH

Introduction

Infinitely clipped speech is speech that has been reduced by repeated peak clipping to a rectangular wave whose axis-crossings bear a one-to-one correspondence to those of the original speech signal. Numerous studies have been conducted to determine which parameters of a speech signal are characteristic thereof. The problem has been attacked from a phonetic standpoint (1, 22), but more often the approach has been one of signal analysis and synthesis.

Potter, Kopp and Green (41), Peterson and Barney (39), and others (42) described speech per se by means of spectrographic analysis, with special emphasis upon the location of the vowel formants and the transitions thereof from the preceding or to the following consonants. This approach did not attempt to relate the measured parameters to aural perception. The work of Licklider (29, 30, 31, 32, 33), Kryter (24, 25), and others (17, 37, 40, 46), however, was strongly oriented towards tying the listener into the communication chain. They were concerned with the effects of limited bandwidth, sampling, nonlinear processing, noise, etc., upon the intelligibility of speech.

In one comprehensive study, Licklider (32) showed that even speech which has been infinitely clipped - and therefore contains none of the amplitude information of the original signal - is moderately intelligible. The inference drawn was, of course, that the axis-crossing rate is a sufficient cue for the correct perception of most speech sounds. Another study (29), in which the effects of center clipping were investigated, proved that axis-crossing information is necessary as well as sufficient.

Licklider also showed that differentiation of a speech signal prior to clipping markedly improved its intelligibility. The signal resulting from these two operations was a rectangular wave whose axis-crossings coincided with the zero-slope points of the original speech signal.

Axis-Crossing Density and Formant Frequencies

It has been postulated (6, 10, 44) that the density of axis-crossings of a speech wave roughly corresponds to the frequency of the first formant (F_1), and that the density of axis-crossings of a differentiated speech wave is approximately equal to the frequency of F_2. It is easily shown that if the amplitude of F_2 is small compared to that of F_1, then an infinitely clipped speech wave will retain only first formant and fundamental pitch information. Infinitely clipped predifferentiated speech, on the other hand, will accentuate the second formant at some expense to the first. The improvement in intelligibility of clipped speech with predifferentiation is not unexpected, therefore, in light of the importance of the second formant.

Noise

There is always noise present in the playback of a recorded speech signal. In most cases, the signal/noise ratio is sufficiently high so that the intervals between phrases and words sound relatively quiet. With an infinite clipper, however, the intervals between words are just as full of sound as are the periods occupied by the words. This "rectangular noise" disappears as soon as the speech signal appears, though, because the input to the clipper consists of the sum of the speech signal and the ambient noise.

A much more disturbing noise is that generated by the clipping process per se. Licklider has shown that there is no clear-cut relationship between impairment of intelligibility and the severity of distortion due to clipping expressed in terms of measurements with sinusoidal test signals.

Spectral Distortion

If the input to an infinite clipper is a pure tone, we know that the distortion products will occur at odd harmonic multiples of the input frequency. We can eliminate these distortion products by low-pass filtering, thereby "resurrecting" the original pure tone. If, on the other hand, the input signal is the sum of two nonharmonically related sine waves, then we can never eliminate all of the distortion products resulting from infinite clipping. This is clear if it is realized that the sum of two nonharmonically related sine waves will have a longer period than either of its two components, and that some of its odd harmonics are almost certain to fall into the passband defined by the component frequencies.

Prediction of the spectrum of an infinitely clipped speech wave is extremely difficult because both the relative amplitudes and the frequencies of all the formants must be known (38). The problem is further complicated by the fact that the formant frequencies, although described in previous paragraphs as though they were pure tones, are in fact only the centroids of energy concentrations in the frequency domain. Since a given formant "frequency" varies, the "steady state" waveform of a vowel is only approximately periodic.

FREQUENCY DIVISION OF INFINITELY CLIPPED SPEECH

The Binary Frequency Divider

What would happen if we acknowledged the existence only of every second axis-crossing in an infinitely clipped speech signal? In other words, what would be the sound of a bistable multivibrator which triggered, say, only on the positive going axis-crossings of a speech input signal?

Figure 1 shows the output signals which would be obtained from such a binary frequency divider for three different bivariate input signals. In Figure 1(a) the input is a rectangular wave which spends an equal amount of time in each successive state. The output signal derived from this square wave input is also a square wave, but one whose period is twice that of the input. The fundamental frequency of the output and the frequency of all of its harmonics are one-half those of the input signal. The relative amplitudes and phases of the harmonics have been retained. For a square wave input signal, therefore, we can truly perform frequency division in the time domain by this technique.

For the nonsquare, periodic input signal shown in Figure 1(b), the period of the output signal is twice that of the input, so we have at least managed to shift the fundamental frequency down one octave. As far as the rest of the input and output spectra are concerned, however, the results are somewhat questionable. There is clearly less high frequency information in the output than in the input, but what has been accomplished is more like halving the upper bound of the input spectrum than dividing all its component frequencies by 2.

In Figure 1(c) we find that not even the fundamental frequency has been changed. Again, we have in effect halved the upper-bound frequency of the input spectrum (though in truth there are components in both spectra in any arbitrarily high frequency region), but we can say little else about what we have succeeded in doing.

On second thought, that last statement is not quite true. It can be seen that if the number of pulses per cycle (or the number of positive going or negative going axis-crossings per cycle) is odd, we do divide at least the fundamental frequency by 2. In either case, what has been done is to eliminate all terms of one sign (whether + or - is arbitrary) in the Fourier transform of the input function, and reverse the sign of half of the terms which remain. All attempts at description of this operation in Fourier notation have so far proved unsuccessful.

Now we can draw the waveform which would result if we could,

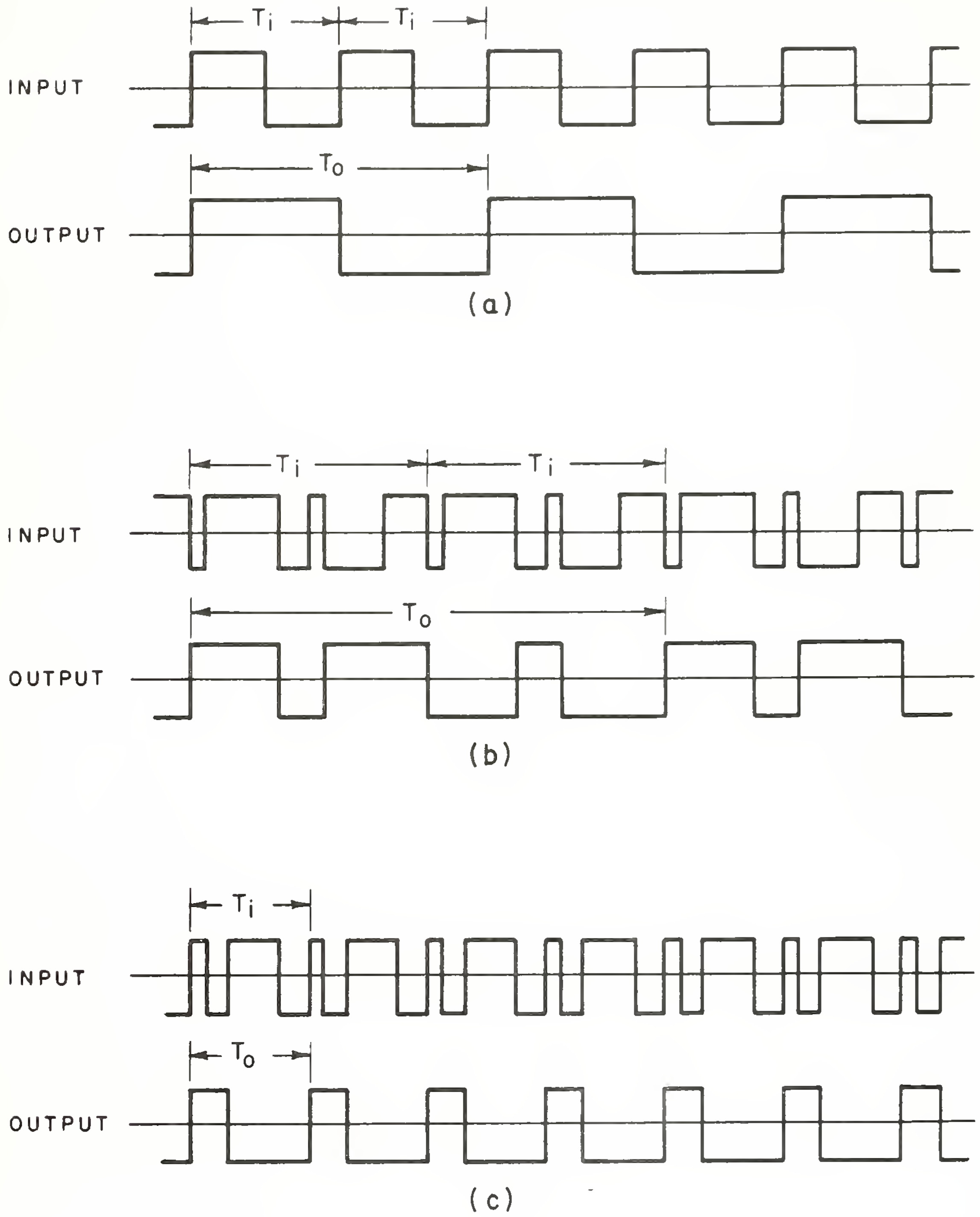


Figure 1. Binary Frequency Division by Recognition of Alternate Input Axis-Crossings.

by some magic, in fact divide by 2 all of the frequencies present in a periodic bivariate wave. A typical example is the set of waveforms shown in Figure 2. The desired output signal was "generated" by doubling the successive interaxis-crossing intervals

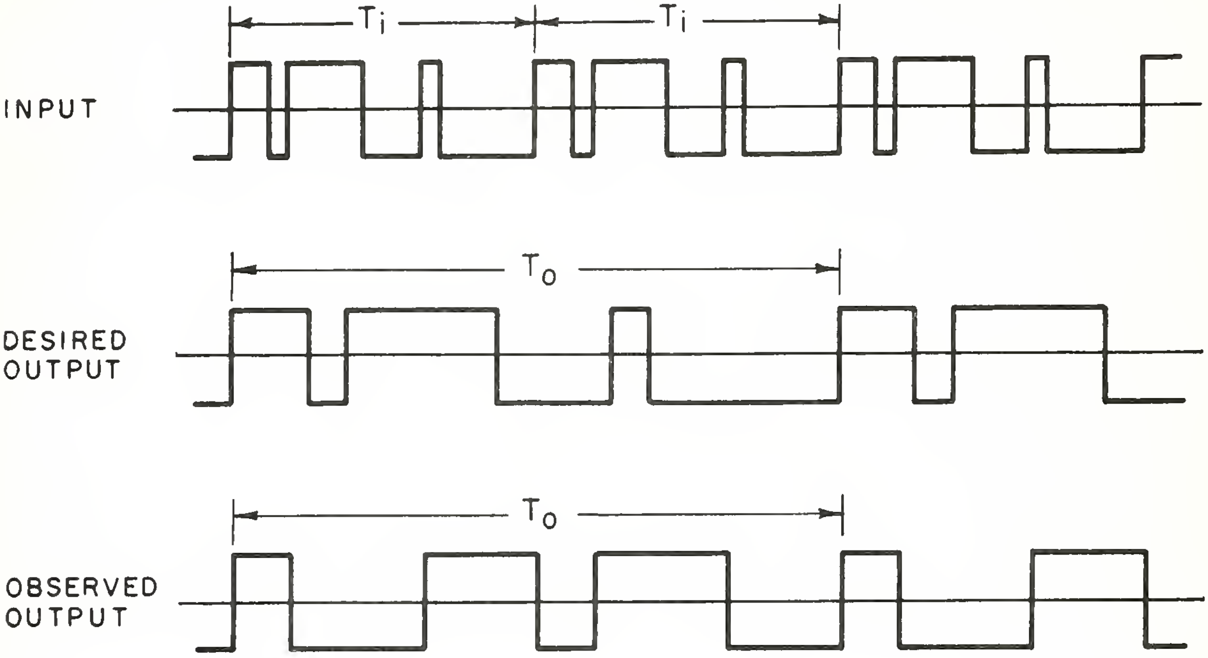


Figure 2. Desired and Observed Outputs of Binary Frequency Divider.

of the input. *Note that transitions occur in the desired frequency divided output signal at times when there are none in the input wave.* This is the heart of the problem of using axis-crossing information only for the generation of a frequency divided output signal, since a transition in the output can only occur when there is one in the input signal.

Another fault of binary frequency division is that, by the very nature of the technique of disregarding alternate axis-crossings, we are "throwing away" half of the information contained in the input. By the same token we would, in the case of division by 3, be discarding two-thirds of the axis-crossing information.

The "Phase-Angle" Frequency Divider

There is another technique by which we may approximate frequency division of a rectangular wave *without* disregarding any axis-crossings. The principle, suggested by Mason (35) is as follows:

A rectangular wave has, by definition, two stable levels. We may consider each change in level to correspond to a 180-degree change in phase angle. If we have a device whose output can assume three stable levels instead of two, each change of output level can be looked upon as a change in phase angle of 90 degrees. If successive transitions of a bivariate wave are used to control successive changes in state of the tri-stable device and the control signal has an odd number of positive going or negative going axis-crossings per period, then the output period will be twice that of the input. For the same input signal conditions, the output of a device with a succession of four stable output levels would have three times the basic period of the input.

Figure 3 illustrates the principle of "phase-angle" frequency division, by a factor of 2, for the same input signals as in Figure 1. The reader should note the strong resemblance of the tri-stable state waveform to the general form of a center clipped signal. Licklider has shown (29) that only small amounts of center clipping are required to completely destroy the intelligibility of speech.

Figure 4 shows a chain of operations performed on an arbitrary periodic waveform. The double differentiation indicated prior to clipping and the double integration following frequency division were found by graphical methods to yield the output signal which most closely approximated the one desired.

EXPERIMENTAL FREQUENCY DIVISION SYSTEM

Principle of Operation

The comments given in the preceding sections would lead the reader to believe that binary and "phase-angle" frequency division systems would work poorly, if at all. Nevertheless, the intelligibility of predifferentiated infinitely clipped speech and the similarity of the "phase-angle" frequency divider output to a desired waveform (Figure 4) made the prospect of actually trying these two schemes too tempting to resist. A block diagram of the experimental system, which can operate in either the binary or "phase-angle" frequency division mode, is shown in Figure 5.

The two cascaded differentiators shown are identical, and can be switched in or out of the circuit at will. The clipper amplifier which follows the differentiators supplies a fairly large trapezoidal wave to the Schmitt trigger circuit, whose output is a rectangular wave with very sharp leading and trailing edges.

An amplifier and split load phase inverter convert the infinitely clipped speech output of the Schmitt circuit into a pair of 180-degree out of phase rectangular wave trains. Each of these

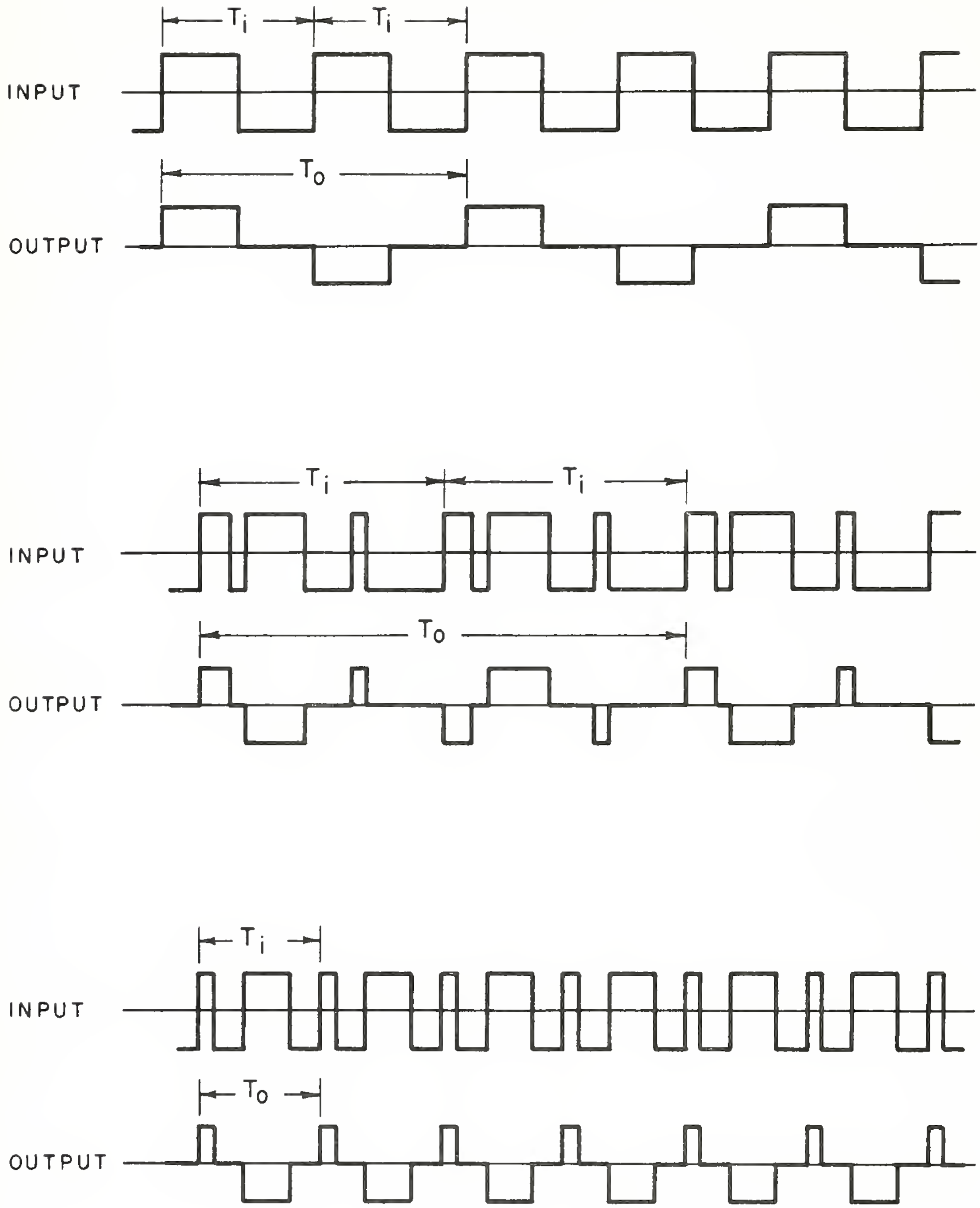


Figure 3. "Phase-Angle" Frequency Divider Output Waveforms.

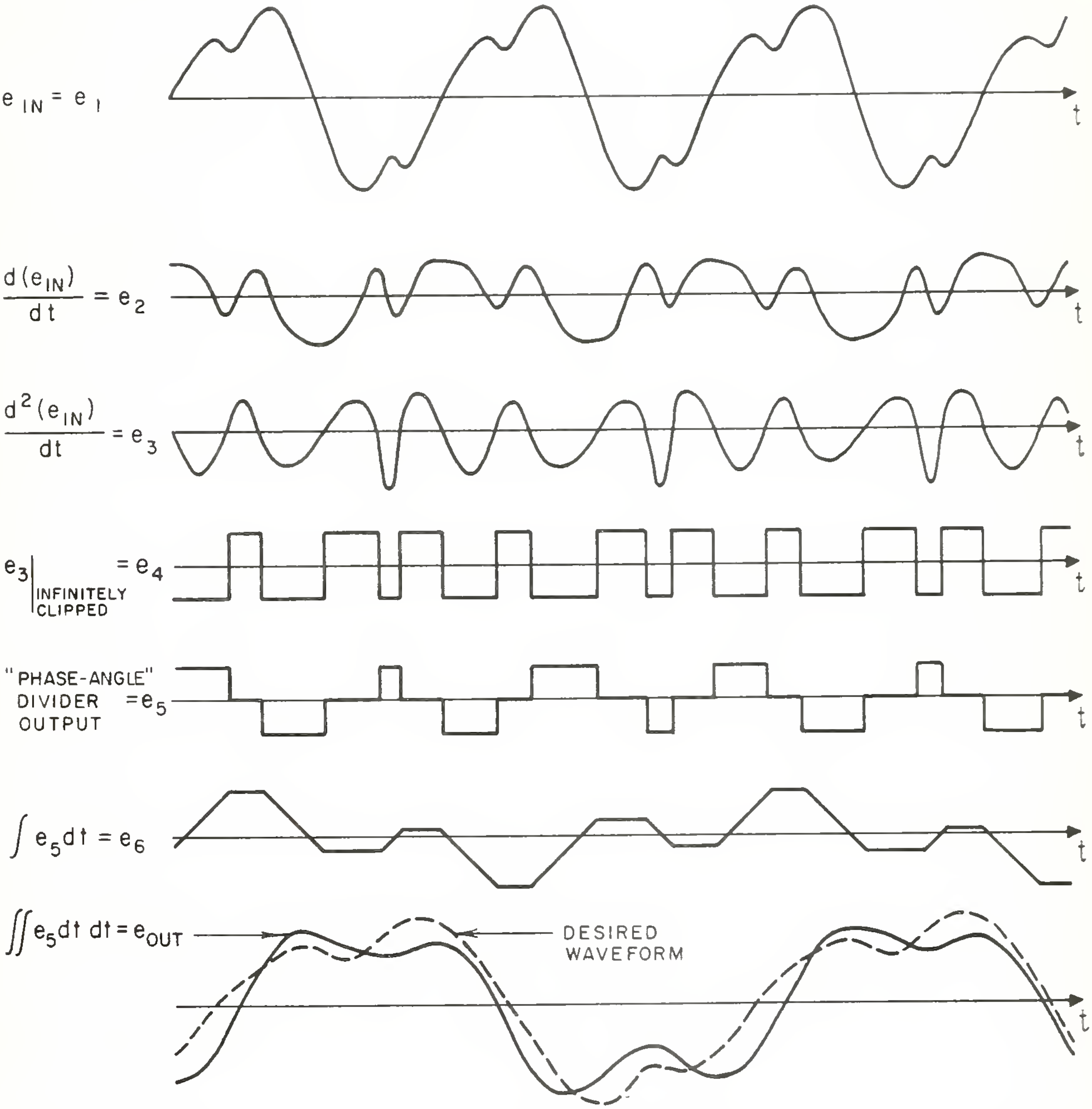


Figure 4. Signal Processing by the "Phase-Angle" Frequency Divider.

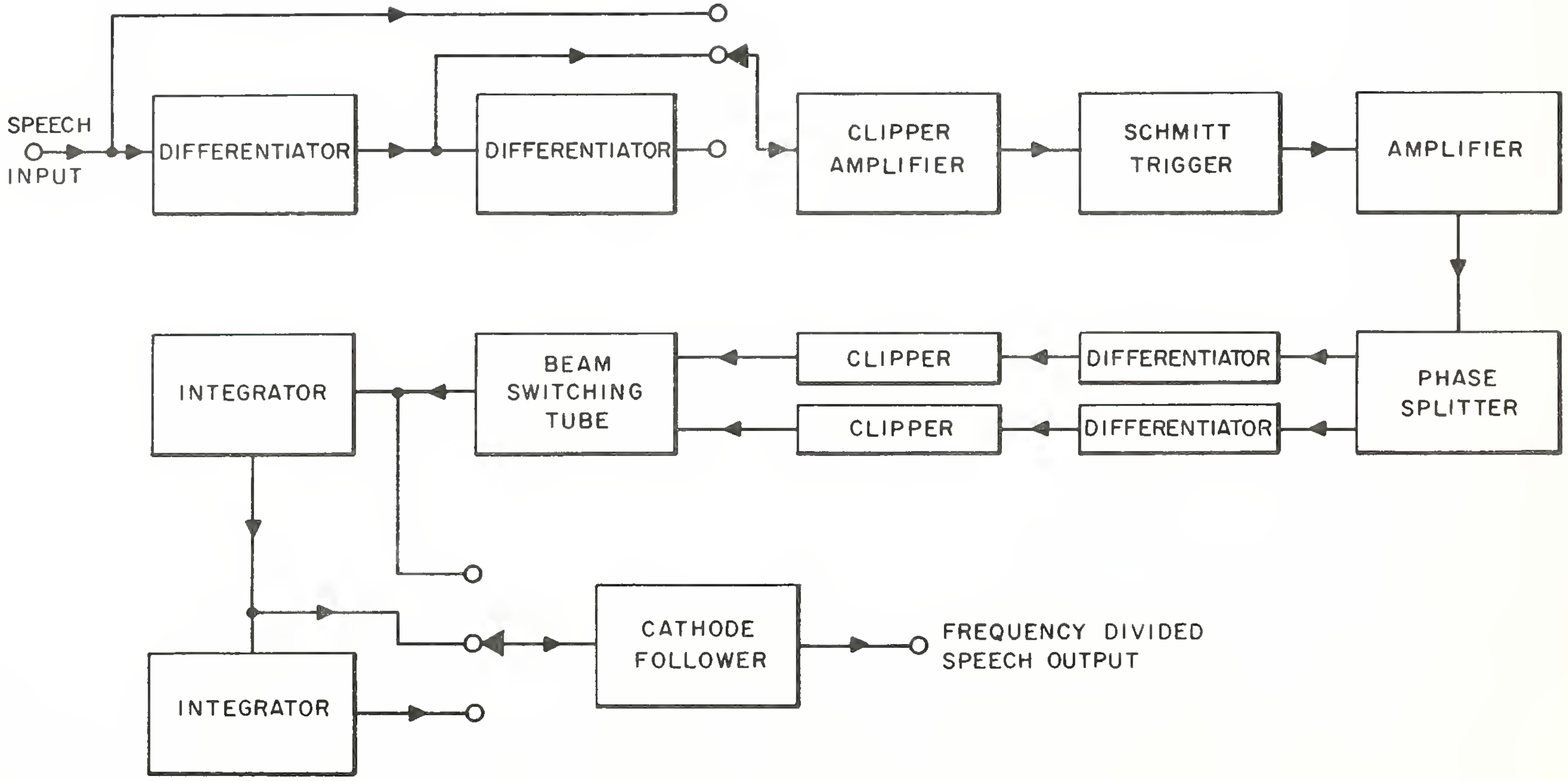


Figure 5. *Block Diagram of Experimental Frequency Division System.*

is differentiated separately, and the positive spikes in the resulting pair of impulse trains are clipped off. The result is a pair of negative impulse trains; the impulses in one train mark positive going axis-crossings of the original speech signal (or its first or second derivative), and the impulses in the other mark the negative going axis-crossings.

The two impulse trains are applied to alternate grids of a beam switching tube. This tube has ten separate beam targets, the voltage at each being independently variable. By adding all of the target voltages together, and since there is no signal from a nonconducting target, the sum signal is equal to the level at the one "live" target. Each time a pulse (from either train) arrives, the beam advances one position and the summed output signal varies accordingly. If successive target voltages correspond to the progression +1, +1, -1, -1, +1, +1, etc., a transition in the summed output signal will occur at every other axis-crossing of the original speech signal. Hence the system acts as a binary frequency divider. (Obviously, we could accomplish binary frequency division by 2 in a much simpler manner.) If sucessive target voltages correspond to the progression +1, 0, -1, 0, +1, etc., the overall system acts as a "phase-angle" frequency divider.

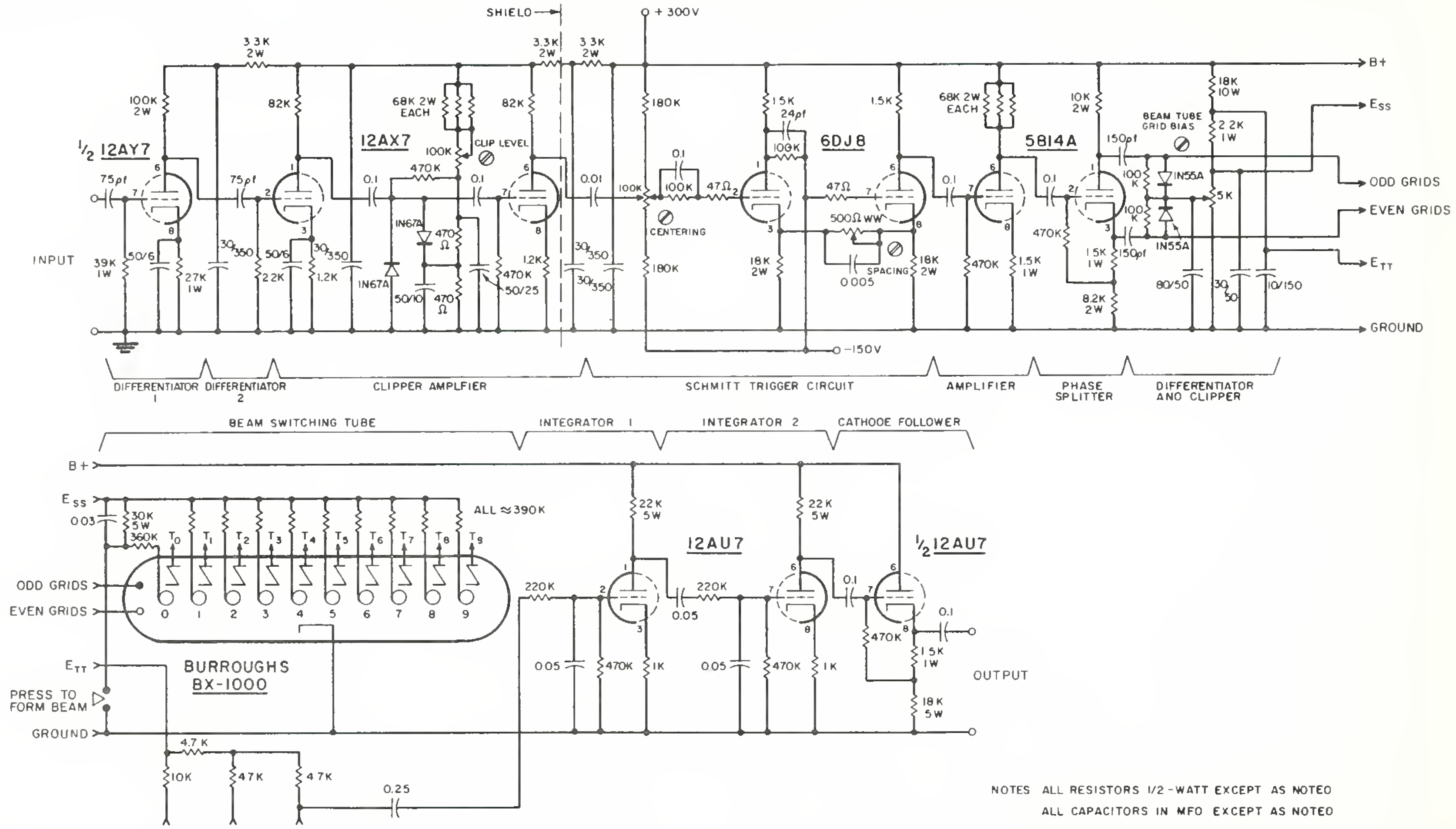
The integrators which follow the beam switching tube can, like the input differentiators, be switched in or out at will.

A complete circuit diagram of the two-mode frequency divider is shown in Figure 6. The circuit occupied one 10 inch by 19 inch relay rack panel.

Performance of Subsystems

Figure 7 is a plot of the pure tone frequency response of a single input differentiator and of the pair in cascade. Figure 8 shows the frequency response of a single output integrator and of the two in cascade.

Considerable difficulty was experienced with the beam switching tube because of external wiring capacitance associated with the target and spade (beam locking) electrodes. Transitions between states in the summed output signal were slow, and the maximum input frequency (i.e., switching rate) was limited to about 7000 cycles per second. Placing the target load resistors in close proximity to the tube socket and wiring the spade load resistors directly to the socket pins reduced the external capacitance so markedly that the maximum input frequency rose to over 50,000 cycles per second.



NOTES ALL RESISTORS 1/2-WATT EXCEPT AS NOTED

ALL CAPACITORS IN MFD EXCEPT AS NOTED

ALL POTS 2-WATT CARBON EXCEPT AS NOTED

MODE	CONNECT TARGETS		
BINARY	$T_{0,1,4,5,8}$	———	$T_{2,3,6,7,8}$
"PHASE-ANGLE"	$T_{0,4,9}$	$T_{1,3,5,7,8}$	$T_{2,6}$

Figure 6. Circuit Diagram of Experimental Frequency Division System.

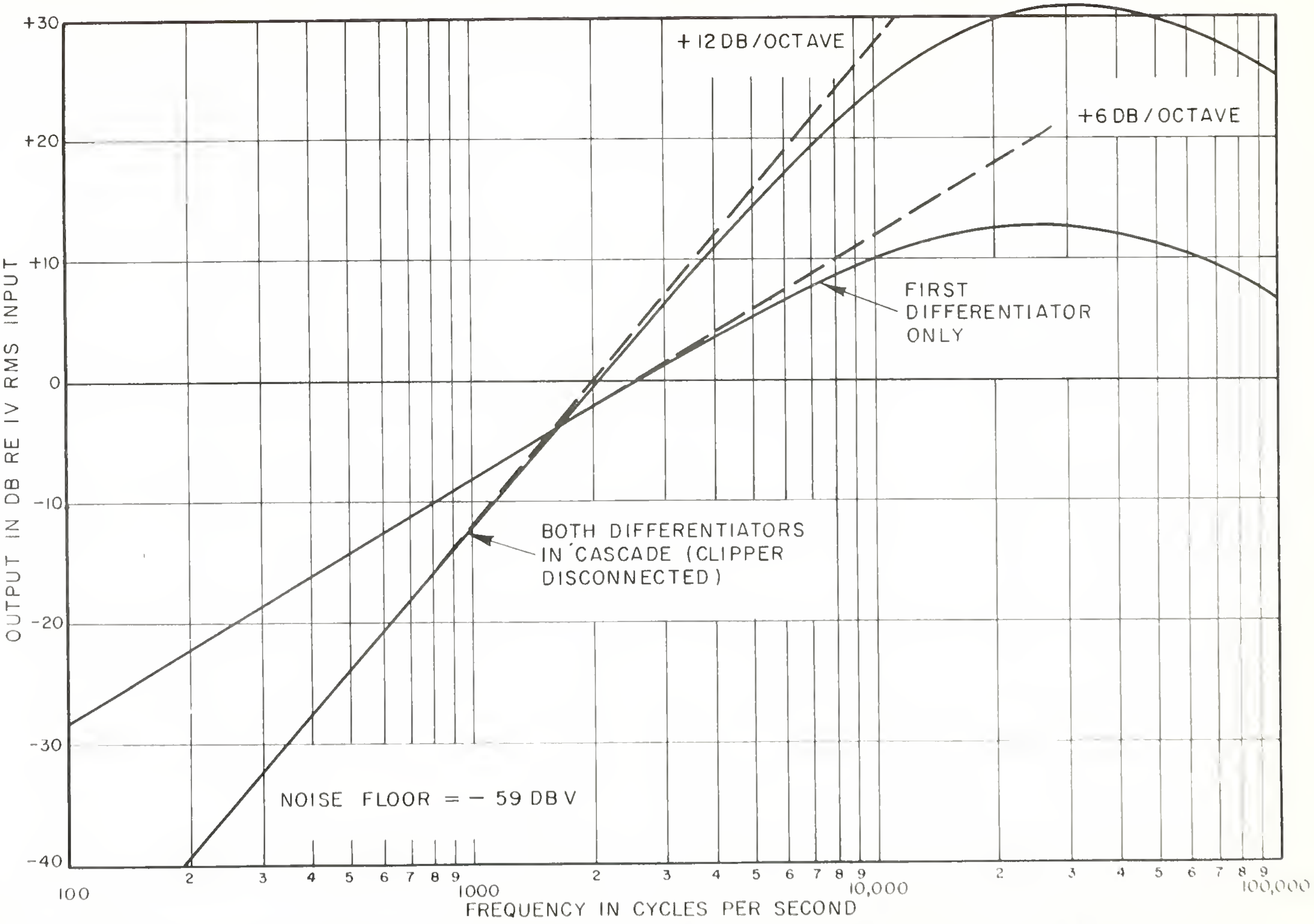


Figure 7. Frequency Response of Single and Cascaded Input Differentiators.

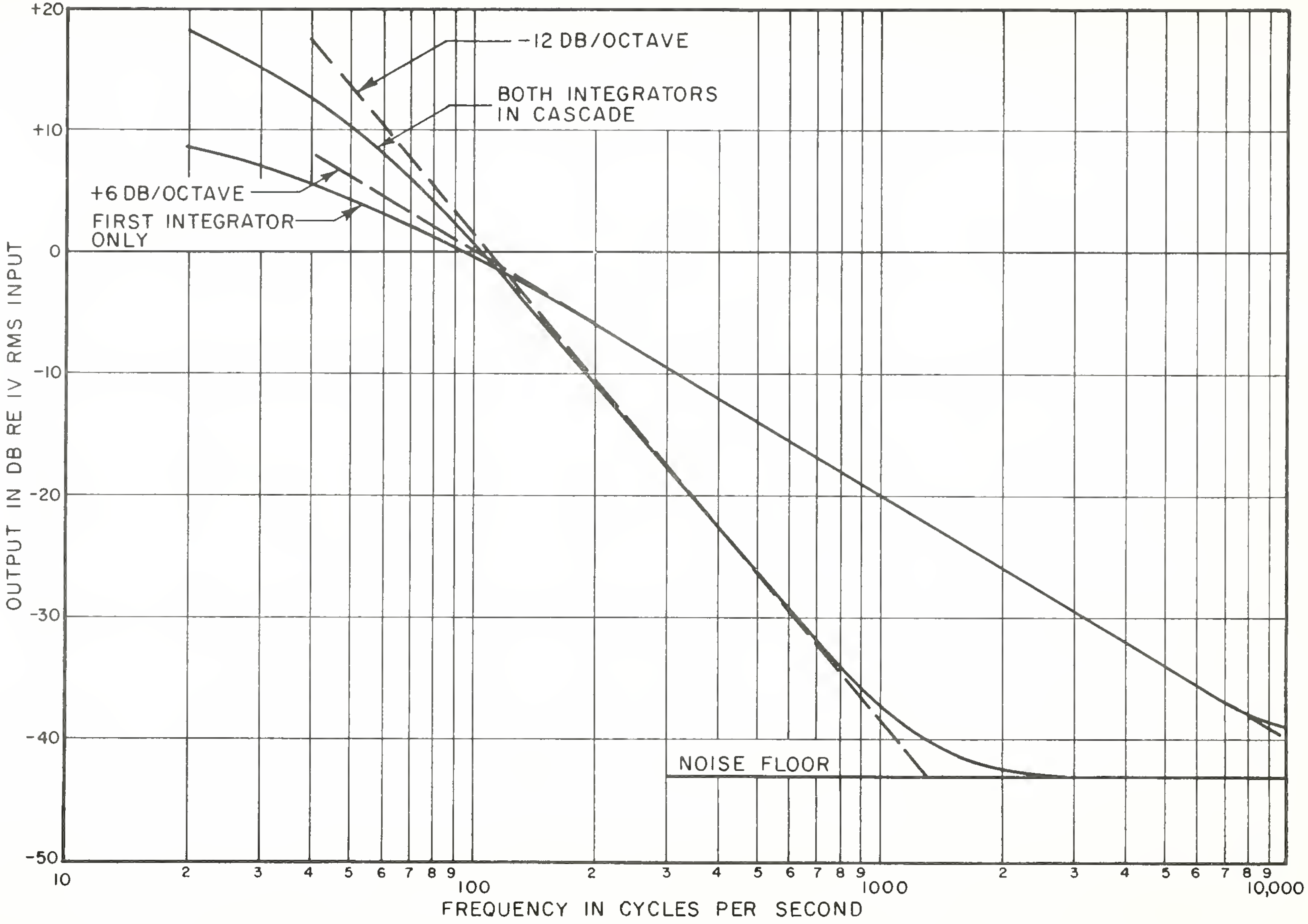


Figure 8. Frequency Response of Single and Cascaded Output Integrators.

INTELLIGIBILITY TESTS

Introduction

After "debugging," both modes of system operation were evaluated by ear. The performance of the "phase-angle" mode was judged so poor that no controlled intelligibility tests were run. This matter will be discussed in some detail below.

To evaluate the frequency divider operating in the binary mode, and to determine separately the effects of shortened phoneme duration and spectrum shift on intelligibility, a series of tape recordings was prepared of PB (2) words. A sample PB list will be found in Appendix A.

Recording System and Techniques

The recordings were made in a low reverberation time (although not strictly anechoic) room which had been especially constructed for maximum acoustic isolation from the adjoining spaces. The ambient acoustic background level was measured and found to be well below the NC-20 curve of Beranek (3) in all octave frequency bands from 20 to 10,000 cycles per second.

Special pains were taken to keep the electrical noise floor of the recording system, shown in Figure 9, as low as possible. The ratio of speech peaks to combined electrical and acoustic noise was measured at the input to the tape recorder and found to be in excess of 56 db overall in the range of frequencies from 75 to 10,000 cycles per second.

The talker was a man of 28 with a typical General American accent and considerable speaking experience. The word lists were recorded without the usual carrier phrases ("The next word is ..." or "Now you will write..."), but were preceded by the list number of the word and a short pause (e.g., "Number seven...catch"). The words were spoken at regular intervals, with timing provided by a flashing neon lamp.

The 50-word lists were recorded according to the timing schedule shown in Table I. The lists read with other than the normal 3.3 second interword interval were "stretched" - i.e., the talker pronounced both the word identification numbers and the words themselves more slowly, so that the ratio of word duration to interword interval remained relatively constant between lists.

Rerecording System and Techniques

The more slowly spoken word lists were rerecorded by means of the system shown in block form in Figure 10. The hysteresis synchronous motors in the playback machines could, by changing the

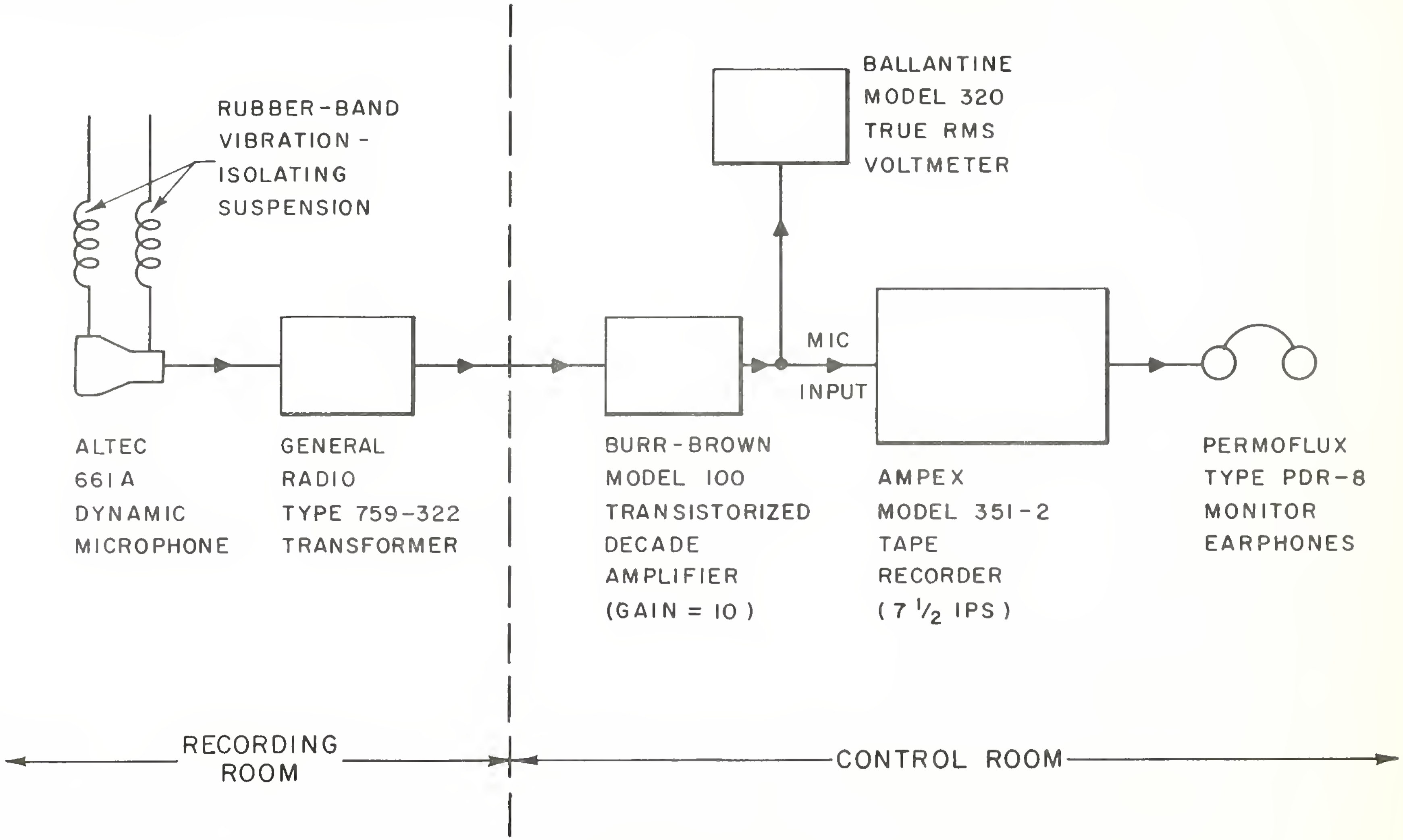


Figure 9. Block Diagram of Recording System.

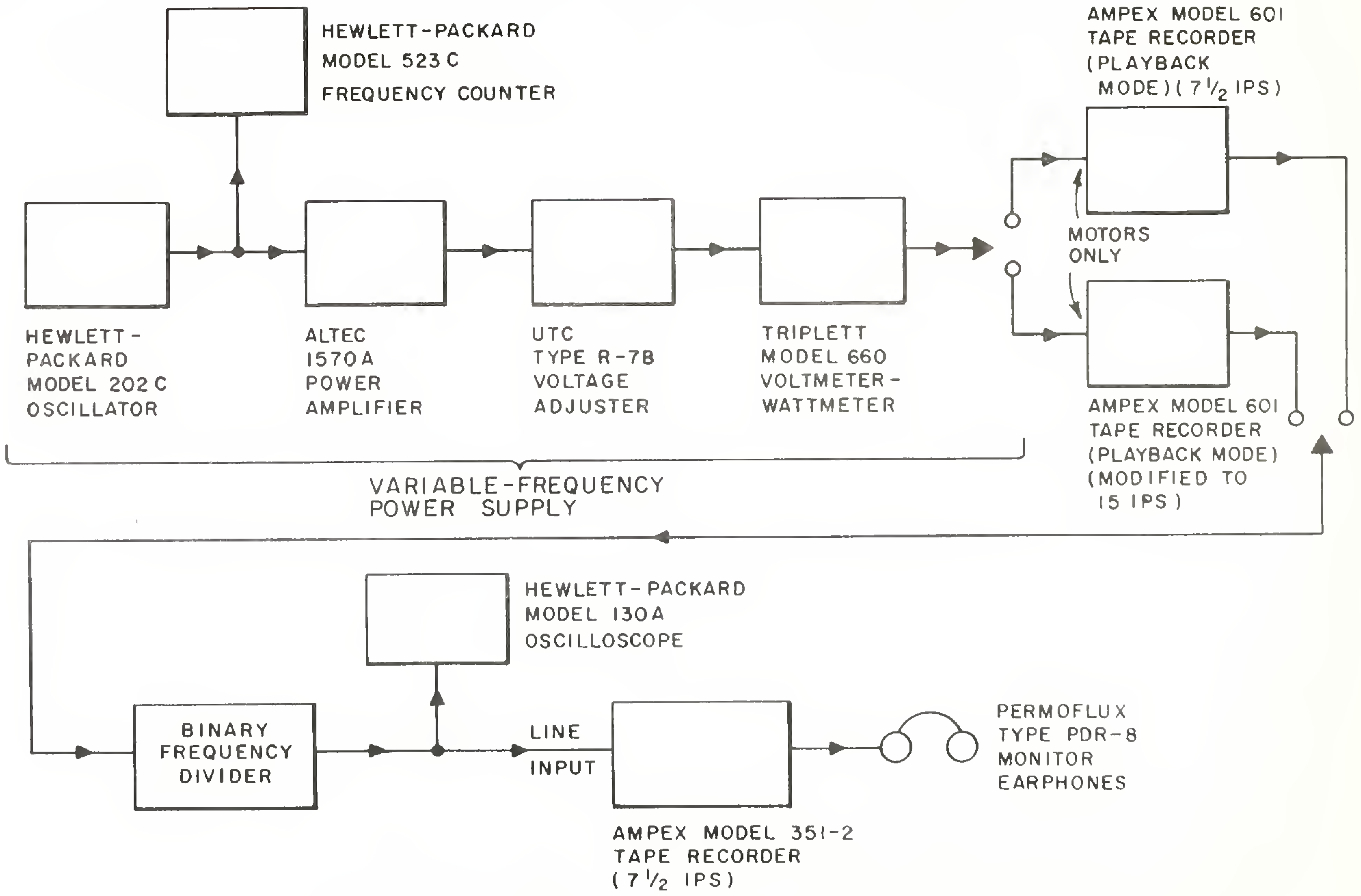


Figure 10. *Block Diagram of Rerecording System.*

TABLE I

PB WORD LIST RECORDING TIMING SCHEDULE

Test No.	Time between Words (as spoken), Sec	Test No.	Time between Words (as spoken), Sec
1	3.3	8	3.3
2	3.3	9	6.6
3	3.3	10	3.3
4	3.3	11	3.3
5	4.4	12	6.6
6	5.5	13	3.3
7	6.6		

frequency of the 117-volt driving signal, be made to run synchronously over a moderate range of speeds. By correctly choosing the playback machine and adjusting the oscillator, the ensemble of recordings listed in Table II was obtained.

It was necessray, in the case of the 7-1/2 inches per second machine, to reduce the starting winding capacitance (by adding a 5-microfarad capacitor in series with the existing 2-microfarad unit) in order to achieve synchronous operation at a "line" frequency of 80 cycles per second.

The binary frequency divider was located, as shown, between the playback machine and the rerecording machine. Input differentiators and output integrators were used or not as indicated in Table IV. Infinitely clipped speech was derived from the output of the Schmitt circuit.

Frequency Response of Overall Record Playback System

The frequency response of the entire recording, rerecording and playback system was determined through the use of a series of pure tones, each of about 5 seconds duration, recorded on the original master tape at a level of -10VU. These test tones were 1/3-octave apart, located at the frequencies given in Table III.

TABLE II

RERECORDING SCHEDULE [a,b]

Test No.	Speed of Playback Machine on 60 cps Line, in./sec	Oscillator Frequency, cps	Ratio of Playback Frequencies to Recorded Frequencies
1	7.5	60	$\frac{7.5}{7.5} \times \frac{60}{60} = 1.00$
2	7.5	80[c]	$\frac{7.5}{7.5} \times \frac{80}{60} = 1.33$
3	15	50	$\frac{15}{7.5} \times \frac{50}{60} = 1.67$
4	15	60	$\frac{15}{7.5} \times \frac{60}{60} = 2.00$
5	7.5	80[c]	etc. 1.33
6	15	50	1.67
7	15	60	2.00
8	7.5	60	1.00
9	15	60	2.00
10	15	60	2.00
11	7.5	60	1.00
12	15	60	2.00
13	15	60	2.00

Notes: a) Tape speed of original recorder = 7.5 in./sec

b) Tape speed of recorder onto which dubbings were made = speed of intelligibility test playback machine = 7.5 in./sec

c) 5 μf condenser required in series with existing 2 μf condenser for synchronous operation at 80 cps

TABLE III

SYSTEM FREQUENCY RESPONSE TEST FREQUENCIES

63	160	400	1000	2500	6300
80	200	500	1250	3200	8000
100	250	630	1600	4000	10000
125	320	800	2000	5000	cycles per second

These frequencies were chosen to correspond to the geometric means of the filter bands in an available 23-band program equalizer, since a possible need for equalization was foreseen.

The overall frequency response of the recording, rerecording and final playback system (except for the recording microphone and the listener's earphones) is shown in Figure 11. The program equalizer was not used in making these measurements, and the response curves of Figure 11 were deemed sufficiently flat so that it was never in fact employed.

Test Presentation Techniques and Instructions

The word lists were presented monophonically through a pair of PDR-8 earphones in type MX-41/AR sponge neoprene cushions. The frequency response characteristics of these earphones, as measured in a standard 6 cc coupler, are plotted in Figure 12.

The tests were presented at an rms sound pressure level of about 85 db *re* 0.0002 dyne/cm^2. The listening environment varied, but in no case was there any interference with the tests by ambient noise.

The subjects were supplied with ruled, prenumbered answer sheets. They were instructed to guess when uncertain, but the short time allotted for the recording of answers did not encourage a listener to stop and think for long. By virtue of the relative rapidity with which the words were presented, the listener was almost always forced to begin writing as soon as the test word had been uttered.

The philosophy behind this fairly high data presentation rate is simply that the 2-1/2 words per second average rate of normal speed connected speech does not allow a listener much time to stop

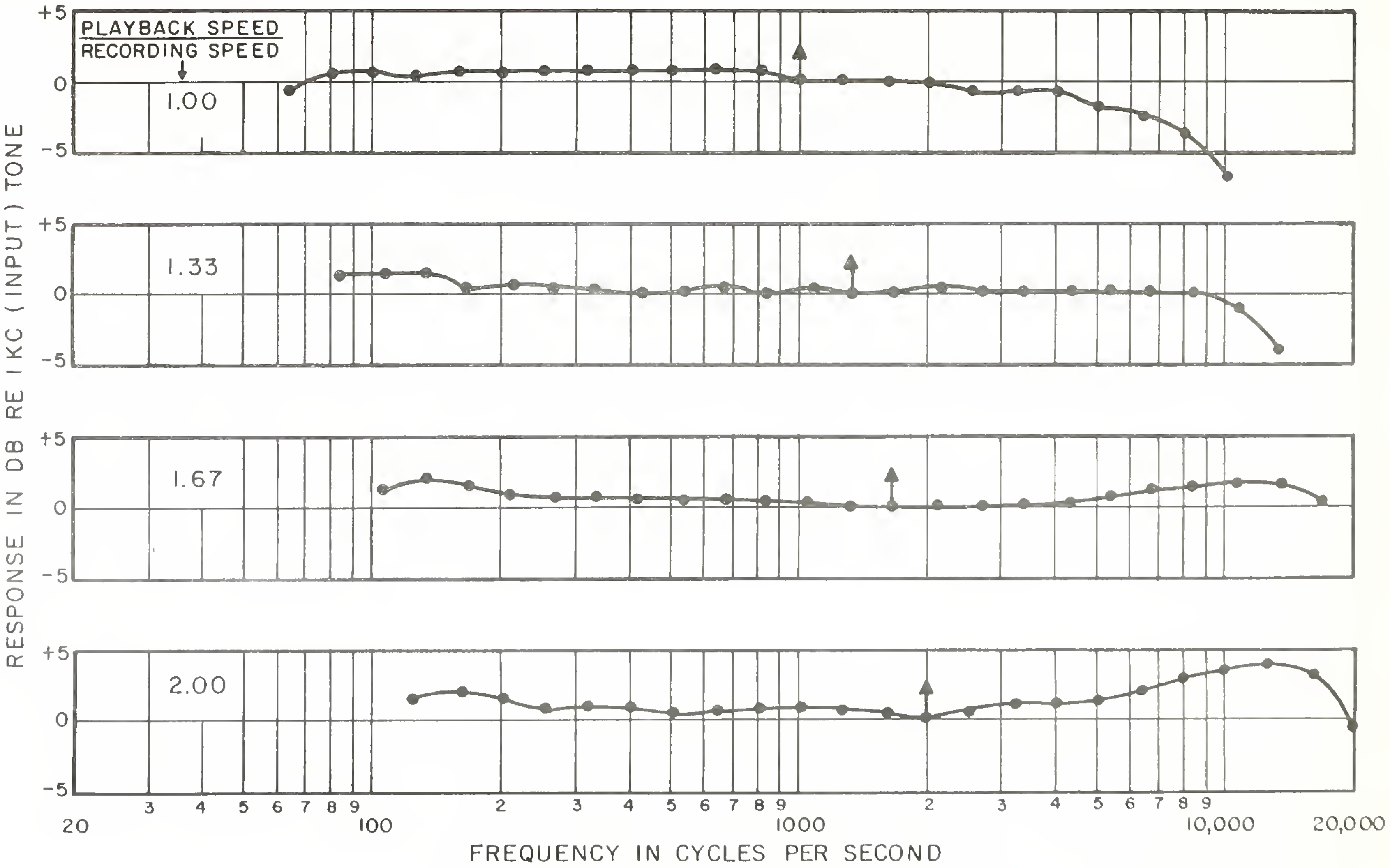


Figure 11. Overall Frequency Response of Recording, Rerecording and Playback Electronics.

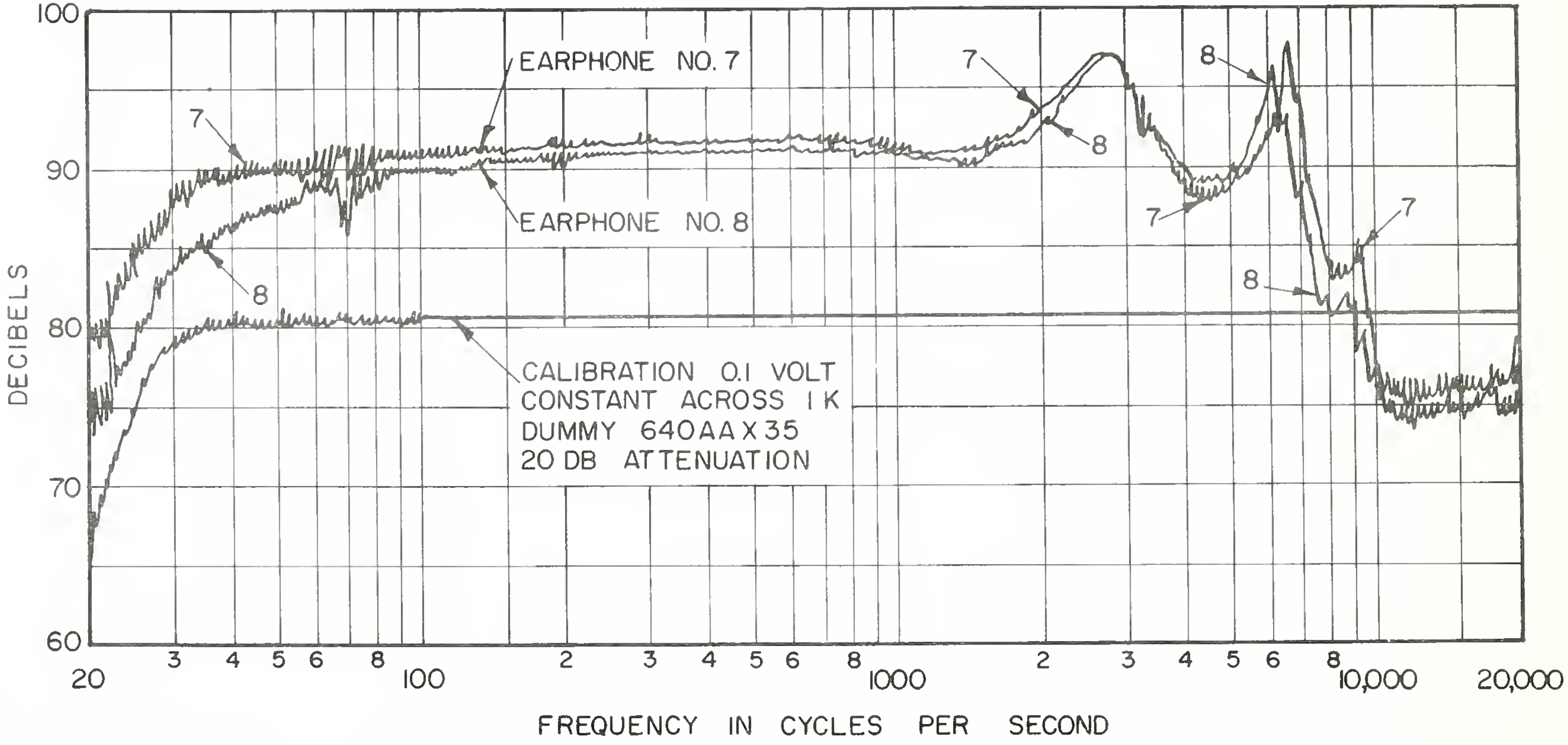


Figure 12. 6CC Coupler Calibration of PDR-8 Playback Earphones.

and think. The test word presentation rate, the reader should note, is something considerably removed from the concept of word duration. Clearly, isolated words such as those used in PB tests are not connected speech, but enough work has been done on the correlation of the two (13) so that the intelligibility of the latter can be predicted from that of the former with a fair degree of precision.

Returning to instructions - the subject was told that he was going to hear several 50-word lists of common monosyllabic words. He was also told that some of the words would be very difficult to understand and that guessing was recommended but not required. If, he was informed, the word he heard sounded like "SNURFTHL," then he should attempt to write "SNURFTHL" down. He was told that spelling did not count, nor did penmanship - except insofar as he was expected to be able to read his own writing.

The test battery was divided into sections consisting, except for the last, of three lists each. A short rest period was given between sections. The order of presentation was the same as in Table IV.

After the battery of tests had been completed, the listener read back his responses, which were then marked correct or incorrect by the author. By having the listener read back his response, problems of indecipherable handwriting were avoided.

RESULTS OF INTELLIGIBILITY TESTS

Test Scores

The scores of the five listeners on the intelligibility tests described above are listed in Table IV.

The average scores for tests 1, 2, 3, and 4 are plotted in Figure 13 as a function of the playback/record frequency ratio. Also plotted are data by Fletcher (20) on CVC (consonant-vowel-consonant) words and by Klumpp and Webster (23) on PB words. Fletcher's data were taken many years ago, using disc recordings and (by modern standards) poor quality equipment. The shape of Fletcher's curve is clearly the same, though, as the other two. The difference between these two curves could be due to different signal/noise ratios, system equalizations, vocabularies (PB words come in lists of various linguistic frequency), etc.

Importance of Spectral and Temporal Distortion

Figure 14 is a plot of the average word intelligibility scores of the entire test battery. Let us first consider the uppermost pair of curves, which compare the intelligibility of words of normal duration (regardless of spectrum) with that of words whose durations

TABLE IV

PB WORD INTELLIGIBILITY TEST SCHEDULE AND SCORES

Test #	Speed or Frequency Ratio	Word Duration Ratio	$\frac{d}{dt}$	Clip	Freq. Div.	$\int dt$	Test Score in % for Listener					Average Score
							KP	JH	HB	RB	SM	
1^a	1.00	1/1.00	no	no	no	no	100	100	100	98	98	99.2
2^a	1.33	1/1.33	↓	↓	↓	↓	100	98	98	96	98	98.0
3^a	1.67	1/1.67	↓	↓	↓	↓	98	90	74	80	70	82.4
4^a	2.00	1/2.00	↓	↓	↓	↓	88	72	40	50	34	56.8
5^b	1.33	1/1.00	↓	↓	↓	↓	96	88	96	96	90	93.2
6^b	1.67	1/1.00	↓	↓	↓	↓	94	86	74	66	66	76.8
7^b	2.00	1/1.00	↓	↓	↓	↓	90	74	48	52	20	56.8
8^c	1.00	1/1.00	one	yes	↓	one	94	78	86	84	84	85.2
9^d	2.00	1/1.00	↓	↓	↓	↓	82	48	38	40	16	44.8
10^e	2.00	1/2.00	↓	↓	↓	↓	74	36	26	24	12	34.4
11^c	1.00	1/1.00	↓	↓	binary	↓	6	6	4	6	0	4.4
12^d	2.00	1/1.00	↓	↓	↓	↓	48	6	6	14	14	17.6
13^e	2.00	1/2.00	↓	↓	↓	↓	28	12	6	10	2	11.6

Notes:
- a) Word duration decreases with increasing playback/record speed ratio
- b) Word duration stays constant as speed ratio changes
- c) Frequencies and word durations are normal
- d) Frequencies are doubled but word durations are normal
- e) Frequencies are doubled and word durations are halved

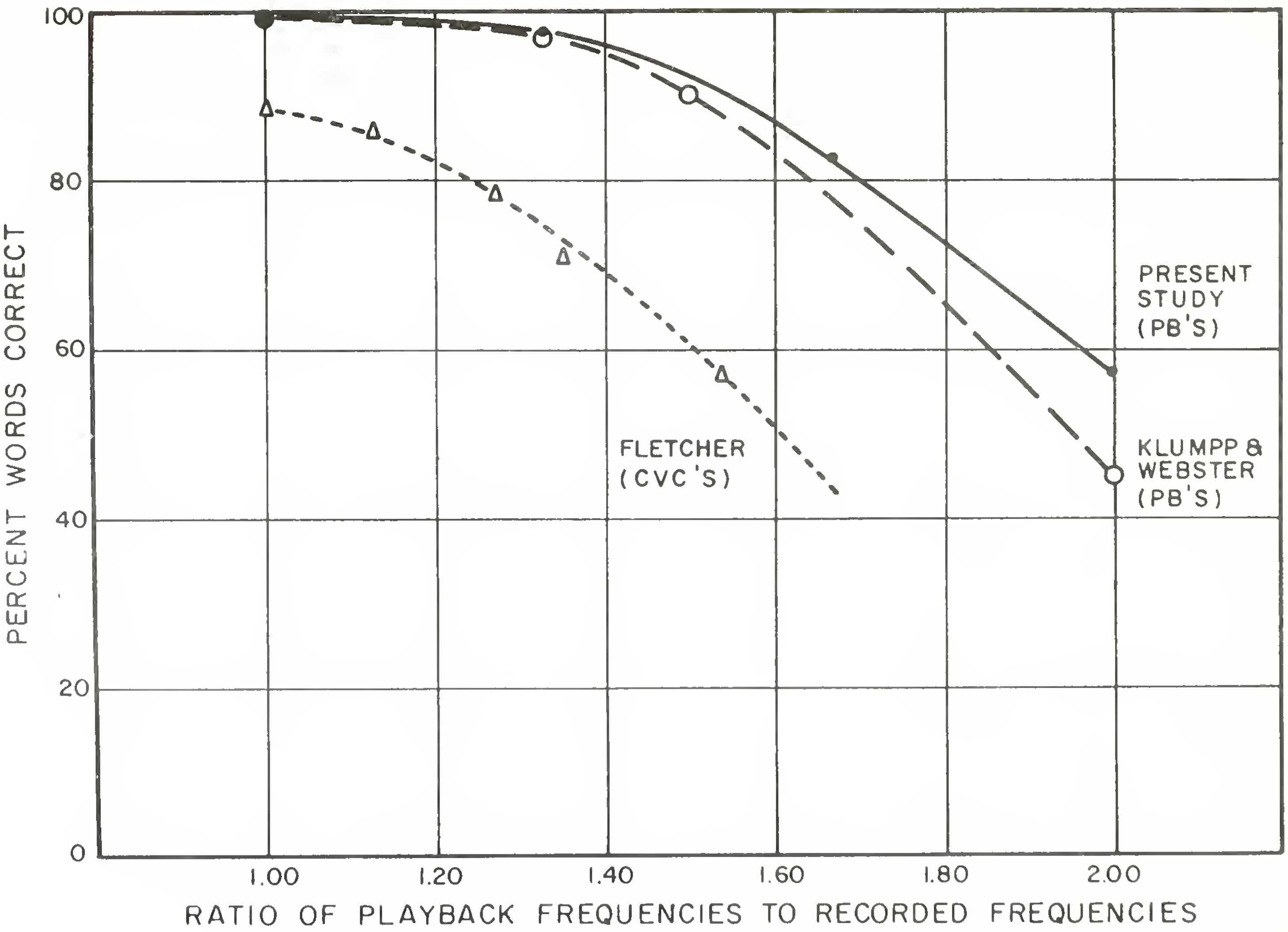


Figure 13. Intelligibility of Unmodified Speech as a Function of Playback-to-Record Frequency or Speed Ratio.

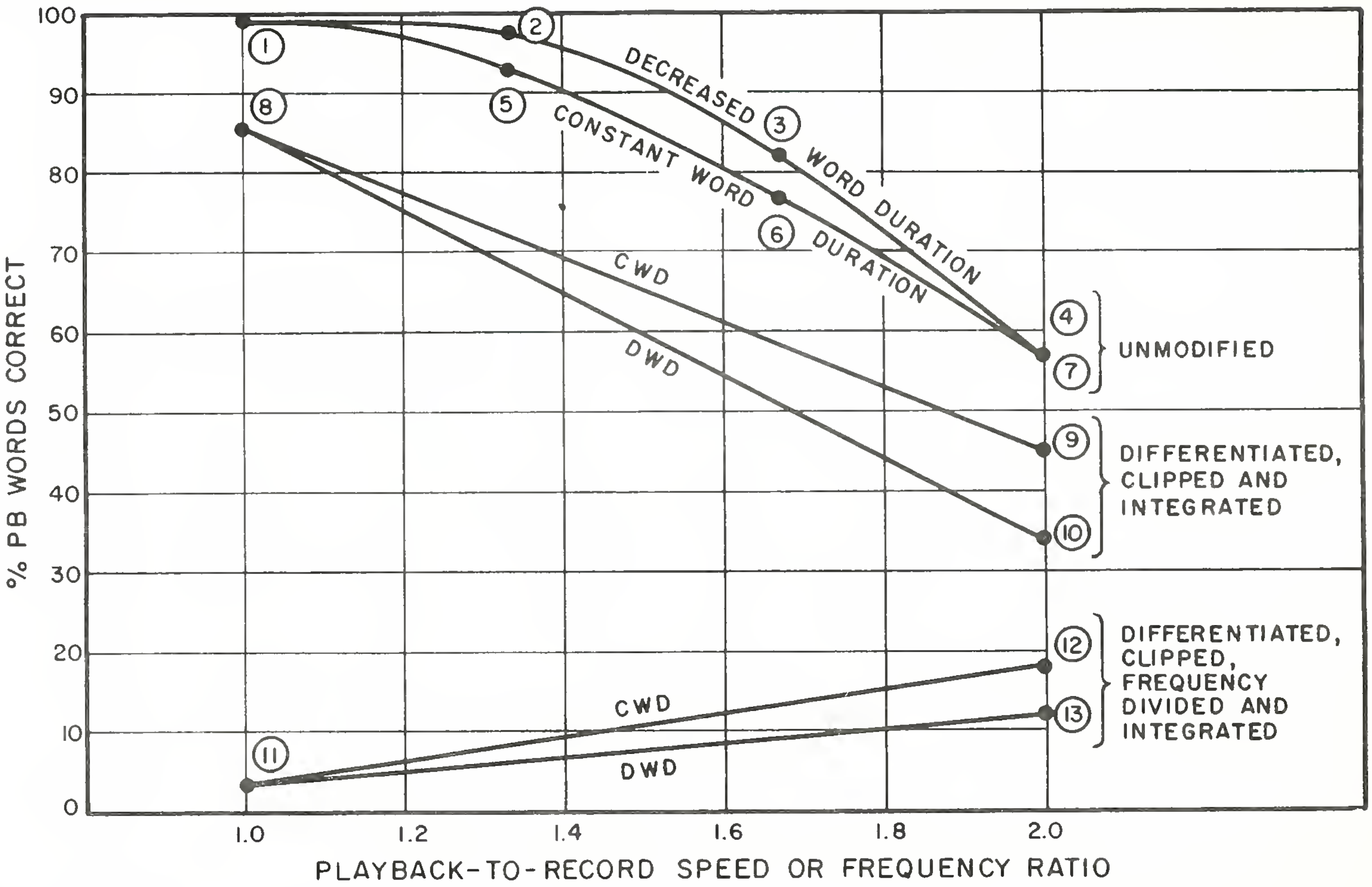


Figure 14. Intelligibility of Unmodified and Modified Speech as a Function of Playback-to-Record Frequency or Speed Ratio.

become shorter as the playback/record speed (and frequency) ratio is increased. The upper curve represents tests 1, 2, 3, and 4, while the lower one represents tests 1, 5, 6, and 7.

I had expected that, if there had been any discernible difference at all, keeping the word duration constant would have improved intelligibility. The fact that the opposite situation was observed might only have been due, I think, to the distortion introduced in the original recording by the talker's attempt to "stretch" his words. In the other sets of curves, which we shall discuss in a moment, the constant duration words were more intelligible than those whose durations and spectra were related in the normal manner. The difference might conceivably lie in that the talker improved his performance at "word stretching" as the recording session progressed. In any event, the maximum difference (6 percent) between the uppermost pair of curves in Figure 14 is not, in view of the small amount of data which it represents, sufficiently significant to warrant further discussion.

Differentiation, Infinite Clipping, and Integration

The middle pair of curves in Figure 14 represents the scores of tests 8, 9, and 10. The words in these tests were predifferentiated, infinitely clipped, and postintegrated. In this case, as previously mentioned, the data at ratio = 2.00 show that having a normal word duration can slightly improve the intelligibility at accelerated playback speeds. The difference between the datum points is 11 percent, which I would venture to say is significant.

The datum at ratio = 1.00 (and hence, perhaps, both curves) is low compared to Licklider's (32) measurement (91 percent words correct on the first trial). The discrepancy could be attributed to differences in listener experience and/or to differences in clipper discrimination level. The clipping thresholds must be adjusted very close to and symmetrically about the zero axis. If they are asymmetric, or symmetrical about and too far from the zero axis, the quality of the clipped speech will be seriously impaired. In our intelligibility tests, the spread between positive and negative clipping thresholds was minimized, after which the pair of thresholds was shifted up and down together for maximum noise output (signifying that the thresholds were located symmetrically about the axis). It is possible that because of power supply instability one or both threshold levels drifted, thus upsetting the required balance conditions.

Binary Frequency Division

The lowest pair of curves in Figure 14 represents the average intelligibility scores for tests 11, 12, and 13, in which the binary frequency divider was used. Once again, the constant word duration datum lies slightly above that of proportionately shortened words. It is gratifying to note that the average number of words correctly perceived is greater at ratio = 2.00 that at ratio = 1.00 when the binary frequency divider is used. What is not so gratifying is the very low intelligibility of both constant duration and decreased duration words at either ratio.

Conclusions

The conclusions which can be drawn from these results may be summarized as follows:

1. The intelligibility of monosyllabic PB words decreases with increasing ratio of playback speed to recording speed, except when binary frequency division is employed.

2. The shift in the speech spectrum corresponding to the playback/record speed ratio is, in general, the determining factor in the loss of intelligibility observed at ratios greater than unity. The difference in intelligibility is negligible between words spoken slowly on recording, so that they have normal duration on playback, and words spoken normally on recording so that upon fast playback they are shorter than normal.

3. The technique of binary frequency division as applied does not improve the intelligibility of accelerated playback speech. It causes, in fact, a serious decrement in intelligibility from that of non-frequency-divided accelerated playback speech. Some improvement might be expected as a result of practice (32, 47), but even with practice and connected speech material consisting of known message sets (43), the intelligibility of binary frequency divided speech could not hope to equal that of nondivided speech.

4. The "phase-angle" frequency division technique described does not appear to be a satisfactory solution either. Speech processed by the experimental system constructed to perform this function was judged so poor in quality, although the system was operating properly, that no quantitative evaluation was made. The technique was described because of its circumvention of the "discarded information" problem inherent in the binary frequency divider.

5. If I may, I should like to interject a few subjective observations which do not really fit well anywhere else in this report. The binary divider does not work well, I feel, because of the wideband noise which it generates. This noise is the result of frequency division of the "rectangular noise" output of the Schmitt circuit. Now infinitely clipped speech, especially when it has been predifferentiated, is highly intelligible even when this "rectangular noise" appears between phrases and words.

Why the frequency divider output should sound noisy *during* speech sounds, then, is not clear, except that perhaps the perturbation of the spectrum by "alternate crossing only" recognition is so extremely stochastic in nature that only the few infinitely clipped speech sounds which spend nearly equal times in successive states for extended periods of time are successfully handled. If one listens closely to the output of the binary divider when the input is double speed speech, the original vocal pitches are occasionally very apparent. In some instances, one can even recognize the talker, a feat which is impossible when listening to the double speed speech input of the divider.

A STORED-SAMPLE FREQUENCY DIVIDER

Introduction

Several days after the preceding report had been typed in final form, another possible frequency division technique came to mind. There was not enough time remaining to investigate it in detail, but it was simulated on a digital computer and the effect of varying one of three parameters was determined.

Principle of Operation

The suggested system which operates in a manner analogous to that of the rotating-head tape playback described on pages 26 and 27 is depicted in block diagram form in Figure 15. The instantaneous amplitude of the double speed speech input is sampled at a frequency which is at least twice the highest frequency of interest. Since this frequency is 6400 cycles per second in double speed speech (3200 cycles per second in normal speed speech), we must take our samples at a rate of at least 12,800 samples per second.

The sample pulses are quantized, stored in a two-dimensional shift register chain (i.e., several shift register chains in parallel, each of which carries one of the sample bits), and read out at 6400 samples per second, or one-half the rate at which they were read in. When the registers are filled* the sampling process stops (though the speech input does not) until the registers are empty, at which time the process is repeated. We may think of this procedure as akin to looking at the double speed speech through a "window" for a length of time dictated by the width of the window. We then store what we see of the speech signal through the window, and read it out of the storage mechanism in a time

---

*If bits are put into one end of an n-address chain at the rate of b bits per second and removed from the other at b/2 per second, the $2n_{th}$ bit will not be able to get in. Sampling must be ceased, therefore, after the $(2n-1)_{th}$ bit.

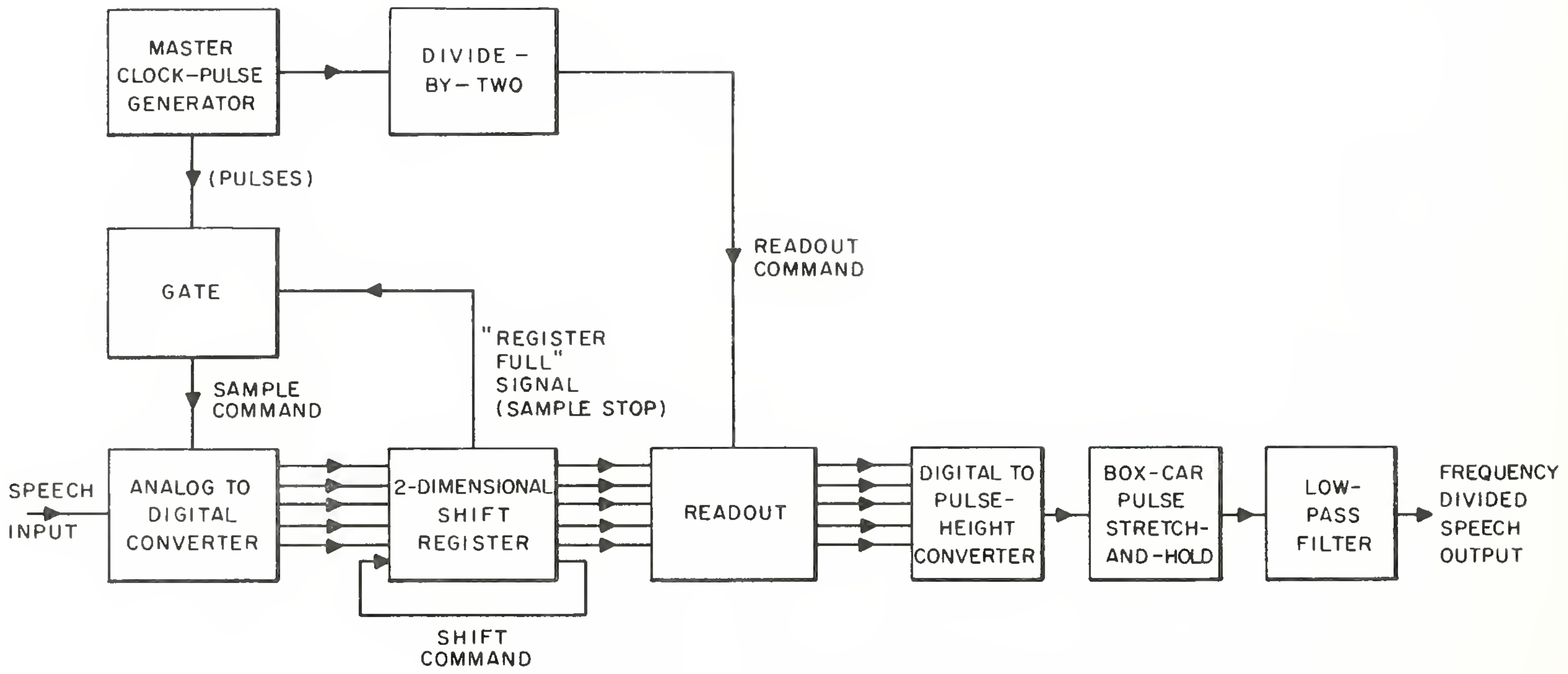


Figure 15. Proposed Stored-Sample Frequency Divider.

equal to two window widths. To avoid confusion, let us define our terms in the following manner:

a) The *sampling frequency* is the rate at which the instantaneous amplitude of the input signal is measured (12,800 samples per second for double speed speech).

b) The *window time* is the period for which we look at the sampled input signal.

c) The *number of samples stored* is equal to the product of the sampling frequency and the window time. For a sampling frequency of 12,800 samples per second and a window time of 10 milliseconds, the number of samples that would be seen and stored would be 128.

d) The *required register chain size* is, for a double speed speech input, simply one-half the number of samples stored. To store the 128 samples in the example above we would need 64 registers for each bit into which the samples were quantized.

Now it is known from Shannon's work that the minimum sampling frequency for double speed speech is about 12,800 samples per second. But what should the window time be and into how many bits must we quantize our samples? As far as the window time is concerned, we know that it must be short compared to the shortest phoneme (at double speed) and at least equal to the period of the lowest frequency of interest (again at double speed). Since some phonemes are as short as 37 milliseconds at double speed (75 milliseconds in real time [37]) and the lowest speech frequency of interest is about 600 cycles per second (300 at normal speed), the optimum window time for double speed speech should be somewhere between 1.6 and 37 milliseconds. The minimum number of bits required to adequately describe a sample would have to be determined empirically, though it is undoubtedly some number less than seven (9).

Intelligibility Tests

The proposed system was simulated on a DEC model PDP 1-B digital computer. There was not enough time available to study the effects of varying the sampling frequency and quantization level, but a study was performed of the effect of window time variation upon intelligibility.

The speech material consisted of phonetically balanced sentences, rather than isolated PB words. The rationale was that connected speech would be a better yardstick by which to appraise the system performance, since our blind students would be listening to connected speech material. A sample PB sentence list will be found in Appendix B.

The sentence lists were recorded in the same environment and with the same system described on page 43. Because the PDP 1-B computer could not sample at a rate much above 8200 samples per second, these recordings were played into the computer at normal speed. The computer input signal was bandpass (300 to 3000 cycles per second) filtered, and the output was limited by a second filter to the frequency range of 150 to 1500 cycles per second. The computer output was recorded at 7-1/2 inches per second and played back to the listeners at 15 inches per second. The net effect was identical to what would have happened had we been able to sample a double speed speech input at 16,400 samples per second.

The window times investigated were 5, 10, 20, 40, 60, and 80 milliseconds. The sampling frequency, as previously stated, was 8200 samples per second, and sampling was done with 11-bit accuracy (10 bits plus a sign bit). One 20-sentence list was processed at each window time, and a "control" list was rerecorded without benefit of computer processing.

List playback to the listeners was in random order, to avoid any monotonic effects due to learning. The "control" list of double speed sentences was played twice, the first time in a random position in the seven-list sequence and the second time at the end of the test session. The scores of the six listeners are tabulated in Table V.

The data of Table V are plotted in Figure 16 as a function of the *normal speed speech input* window time. The average first trial score for unmodified double speed sentences is also plotted.

Results and Conclusions

Figure 16 shows the striking improvement in intelligibility afforded by the stored sample frequency divider. It provides conclusive proof that spectrum shift is indeed the major factor contributing to loss of intelligibility at playback/record speed ratios greater than one. This was a conclusion that was reached only by inference in tests employing the binary frequency divider, since we showed there that keeping word duration constant did not significantly improve intelligibility.

The shape of the curve of Figure 16 is close to that predicted above since window times near 3.2 and 75 milliseconds with a normal speed input result in significant degradation in performance.

To build a practical frequency divider, we must investigate the effect of coarser quantization of the speech samples. The PDP 1-B computer quantized its samples into eleven bits, but we

TABLE V

DOUBLE-SPEED SENTENCE INTELLIGIBILITY TEST SCORES

List No.	Window Time (Normal-Speed Input)	% Key Words Correctly Perceived by Listener*						
		HA	RB	KP	SM	RG	AI	Average
1	5 milliseconds	93	79	89	88	79	91	86.50
2	10 milliseconds	99	93	97	96	95	98	96.33
3	20 milliseconds	96	92	98	96	93	94	94.83
4	40 milliseconds	97	87	97	90	90	87	91.33
5	80 milliseconds	66	59	81	68	73	76	70.50
6	60 milliseconds	94	88	96	97	92	98	95.83
7	no computer	59	35	89	44	49	80	59.33
7	second trial	79	60	-	68	-	-	69.00

---

* The key words in each sentence are underlined (see Appendix B). Each sentence contains five key words. If the key words are correctly perceived, we may say that the sense or meaning of the sentence has been fully perceived.

know that many pulse code modulation (PCM) systems work extremely well with only 7 bits.

There are two criteria for a practical device of this sort. First, it must improve the intelligibility of a double speed, connected speech input to 95 percent or more, and it must be reasonably priced. The cost of the device would depend in large measure upon the required storage capacity, which in turn depends upon the sampling frequency, window time, and quantization accuracy. We do not know the minimum number of bits which would result in acceptable speech quality, but Figure 16 clearly shows that a 5 millisecond window (for a double speed input or 10 milliseconds for the normal speed input) is close to the minimum size consistent with high intelligibility.

The ideal system, then, would accept a double speed speech

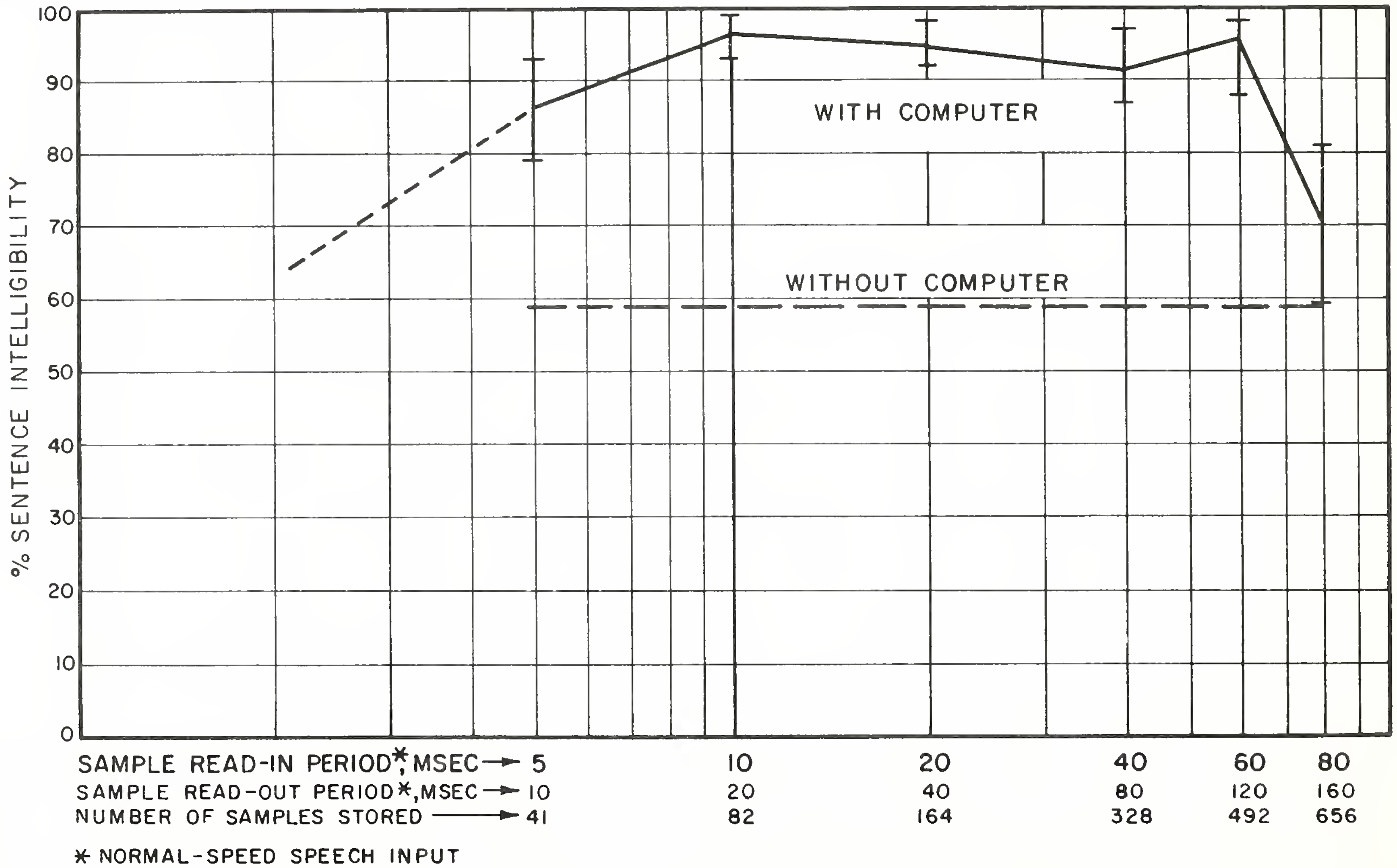


Figure 16. Sentence Intelligibility of Double-Speed Speech with and Without Digital Computer Frequency Division.

input, sample it at 12,800 samples per second for 5 milliseconds, store these 64 samples, and read them out in 10 milliseconds.

RECOMMENDATIONS FOR FURTHER RESEARCH

General Recommendations

On the basis of the almost completely negative results obtained with the binary and "phase-angle" frequency dividers, I would suggest that a very different approach be made to the problem. I think it is clear that the root of the matter lies in restoring the originally recorded speech spectrum. Various modulation techniques (VoBanC [4], the Marcou-Daguet system [34]) seem to hold some promise of accomplishing this objective. Experimental work with a formant tracker (7, 16, 18, 19, 28, 36, 48) should also be carried out, for although I do not think it represents a practical solution to the problem, the probability of improving the intelligibility of accelerated playback speech through the use of such a device seems high.

APPENDIX A

SAMPLE PB* WORD LIST

1. claws
2. fade
3. lip
4. rob
5. cat
6. crab
7. chip
8. bale
9. hush
10. chaff
11. his
12. camp
13. axe
14. sieve
15. waste
16. trod
17. dice
18. grab
19. loud
20. got
21. art
22. chain
23. fool
24. grew
25. share
26. gush
27. lunge
28. gray
29. thorn
30. weed
31. dot
32. cub
33. ouch
34. hide
35. thine
36. freeze
37. fat
38. thaw
39. debt
40. sash
41. sack
42. note
43. lynch
44. chill
45. lime
46. flare
47. bless
48. claw
49. rose
50. aims

---

* Phonetically Balanced

APPENDIX B

SAMPLE SENTENCE LIST

1. The *dune rose from* the *edge* of the *water.*
2. *Those words* were the *cue* for the *actor* to *leave.*
3. *Farmers hate* to *use* a *hoe* or *rake.*
4. A *yacht slid around* the *point* into the *bay.*
5. The *two met while playing* on the *sand.*
6. *It's foolish* to *make* a *pass* at *Jane.*
7. The *ink stain dried* on the *finished page.*
8. *Fail once* on this *job* and *be discharged.*
9. *Scotch can't* be *bought today* at *all.*
10. The *walled town* was *seized without* a *fight.*
11. The *lease ran out* in *sixteen weeks.*
12. *They pulled* a *fast one* on the *deacon.*
13. The *lewd face stared out* of the *window.*
14. A *fine starry night greets* the *pair.*
15. *I* am *speaking dumb* and *vain words.*
16. A *tame squirrel makes* a *nice pet.*
17. The *throb* of the *car woke* the *sleeping cop.*
18. *George* the *second* was *then queen* of the *May.*
19. *Great men are* the *worst husbands.*
20. The *heart beat strongly* and with *firm strokes.*

SUMMARY

Tape recordings of speech become essentially unintelligible when played back at ratios of playback speed to recording speed greater than about 1.6. The possible causes of this phenomenon are enumerated, and several schemes are described for improving the intelligibility of accelerated playback speech.

Two frequency dividers are described which operate upon a principle of recognition of the axis-crossings of the double speed speech signal. The better of these is evaluated by means of listening tests and is found to be wholly unsatisfactory.

In the final section, a frequency divider based upon sampling and storage principles is described. Its strikingly successful performance is demonstrated by means of connected speech listening tests. Spectrum shift is thereby proved to be the major cause of decreased intelligibility in accelerated playback speech.

ACKNOWLEDGMENTS

I wish to acknowledge my indebtedness, first and foremost, to Professor Samuel J. Mason, without whose help and guidance this project might never have been completed. I would also like to thank Professor Kenneth N. Stevens for his sound advice and the use of his speech spectrogram facility.

Also due a vote of thanks are my talkers, Henry Beller and Carl Williams, and my listeners:

Herbert Amster
Ruth Ball
Rene Beller
Henry Beller
Rolfe Goetze
Judith Hart
Alice Isbell
Susan Millman
Karl Pearsons

I must also thank Bill Fletcher for his assistance in programming the PDP1-B computer, Marion Giurleo for typing the report, and Polly Horan for preparation of the figures. My acknowledgments would be incomplete, though, if I failed to express my thanks to my wife, Ricky, whose continuing encouragement was perhaps the most tangible asset of them all.

REFERENCES

1. Ball, J.H. "The Perceptability of Seven Distinctive Features of English Consonants in the Presence of Time and Frequency Distortion." Unpublished research report, Massachusetts Institute of Technology, 1958.

2. Beranek, L.L. Acoustic Measurements. New York: John Wiley and Sons, Inc., 1949, pp. 770-773.

3. Beranek, L.L., "Revised Criteria for Noise in Buildings," Noise Control, Vol. 3, No. 1 (1957), pp. 19-27.

4. Bogert, B.P., "The VoBanC - A Two-to-One Speech Band-Width Reduction System," J. Acoust. Soc. Amer., Vol. 28, No. 3 (1956), pp. 399-404.

5. Borst, J.M., "The Use of Spectrograms for Speech Analysis and Synthesis," J. Audio Eng. Soc., Vol. 4, No. 1 (1956), pp. 14-23.

6. Chang, S.H., "Portrayal of Some Elementary Statistics of Speech Sounds," J. Acoust. Soc. Amer., Vol. 22, No. 6 (1950), pp. 768-769.

7. Chang, S.H., "Two Schemes of Speech Compression System," J. Acoust. Soc. Amer., Vol. 28, No. 4 (1956), pp. 565-572.

8. Cooper, F.S., "Spectrum Analysis," J. Acoust. Soc. Amer., Vol. 22, No. 6 (1950), pp. 761-762.

9. David, E.E., M.V. Mathews, and H.S. McDonald. "Experiments with Speech Using Digital Computer Simulation." Bell System Monograph No. 3405, 1959.

10. Davis, K.H., R. Biddulph, and S. Balashek, "Automatic Recognition of Spoken Digits," J. Acoust. Soc. Amer., Vol. 24, No. 6 (1952), pp. 637-642.

11. Delattre, P.C., A.M. Leberman, and F.S. Cooper, "Acoustic Loci and Transitional Cues," J. Acoust. Soc. Amer., Vol. 27, No. 4 (1955), pp. 769-773.

12. Denes, P., "Effect of Duration on the Perception of Voicing," J. Acoust. Soc. Amer., Vol. 27, No. 4 (1955), pp. 761-764.

13. Egan, J.P., et al. "Articulation Testing Methods II." OSRD Report No. 3802, 1944.

14. Fairbanks, G., W.L. Everitt, and R.P. Jaeger, "Method for Time or Frequency Compression-Expansion of Speech," Trans. Inst. Radio Eng., Vol. AU-2, No. 1 (1954), pp. 7-12.

15. Fairbanks, G., and F. Kodman, "Word Intelligibility As a Function of Time Compression," J. Acoust. Soc. Amer., Vol. 29, No. 5 (1957), pp. 636-641.

16. Fant, C.G.M., "Speech Communication Research," Iva, Vol. 24 (1953), pp. 331-337.

17. Flanagan, J.L., "Effect of Delay Distortion upon the Intelligibility and Quality of Speech," J. Acoust. Soc. Amer., Vol. 23, No. 3 (1951), pp. 303-307.

18. Flanagan, J.L. "A Speech Analyzer for a Formant-Coding Compression System." Unpublished Doctor of Science thesis, Massachusetts Institute of Technology, 1955.

19. Flanagan, J.L., "Automatic Extraction of Formant Frequencies from Continuous Speech," J. Acoust. Soc. Amer., Vol. 28, No. 1 (1956), pp. 110-118.

20. Fletcher, H. Speech and Hearing. New York: D. Van Nostrand Company, Inc., 1929, pp. 291-293.

21. Gabor, D., "Theory of Communication," J. Inst. Elec. Eng., Vol. 93, Part III (1946), pp. 429-457.

22. Jakobson, R., C.G.M. Fant, and M. Halle, "Preliminaries to Speech Analysis," Technical Report No. 13. Cambridge, Massachusetts: Massachusetts Institute of Technology Acoustics Laboratory, 1952.

23. Klumpp, R.G., and J.C. Webster, "Intelligibility of Time-Compressed Speech," J. Acoust. Soc. Amer., Vol. 33, No. 3 (1961), pp. 265-267.

24. Kryter, K.D. "Speech Communication in Noise." Air Force Cambridge Research Center TR-54-52, 1955.

25. Kryter, K.D., "On Predicting the Intelligibility of Speech from Acoustical Measures," J. Speech Hear. Disord., Vol. 21, No. 2 (1956), pp. 208-217.

26. Kryter, K.D., private communication.

27. Latham, W.S., "Variable-Speed Scanning of Recorded Magnetic Tapes," J. Audio Eng. Soc., Vol. 6, No. 1 (1958), pp. 26-34.

28. Lawrence W., in Communication Theory. London: Butterworth's Scientific Publications, 1953, Chap. 34.

29. Licklider, J.C.R., "Effects of Amplitude Distortion upon the Intelligibility of Speech," J. Acoust. Soc. Amer., Vol. 18, No. 2 (1946), pp. 429-434.

30. Licklider, J.C.R., "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise," J. Acoust. Soc. Amer., Vol. 20, No. 2 (1948), pp. 150-159.

31. Licklider, J.C.R., "The Intelligibility of Amplitude-Dichotomized, Time Quantized Speech Waves," J. Acoust. Soc. Amer., Vol. 22, No. 6 (1950), pp. 820-823.

32. Licklider, J.C.R., and I. Pollack, "Effects of Differentiation, Integration and Infinite Peak Clipping upon the Intelligibility of Speech," J. Acoust. Soc. Amer., Vol. 20, No. 1 (1948), pp. 42-51.

33. Licklider, J.C.R., D. Bindra, and I. Pollack, "The Intelligibility of Rectangular Speech-Waves," Amer. J. of Psychol., Vol. LXI, No. 1 (1948), pp. 1-20.

34. Marcou, J., and J. Daguet, "New Methods of Speech Transmission," in Third London Symposium on Information Theory. London: Butterworth's Scientific Publications, 1955, pp. 231-244.

35. Mason, S.J., private communication.

36. Meeks, W.W., J.M. Borst, and F.S. Cooper, "Syllable Synthesizer for Research on Speech," J. Acoust. Soc. Amer., Vol. 26, No. 1 (1954), p. 137(A).

37. Miller, G.A., and J.C.R. licklider, "The Intelligibility of Interrupted Speech," J. Acoust. Soc. Amer., Vol. 22, No. 2 (1950), pp. 167-173.

38. Peterson, E., "Frequency Detection and Speech Formants," J. Acoust. Soc. Amer., Vol. 23, No. 6 (1951), pp. 668-674.

39. Peterson, G.E., and H.L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Amer., Vol. 24, No. 2 (1952), pp. 175-184.

40. Peterson, G.E., E. Sivertsen, and D.L. Subrahmanyam, "Intelligibility of Diphasic Speech," J. Acoust. Soc. Amer., Vol. 28, No. 3 (1956), pp. 404-411.

41. Potter, R.K., G.A. Kopp, and H.C. Green. Visible Speech. New York: D. Van Nostrand, Inc., 1947.

42. Potter, R.K., and G.E. Peterson, "The Representation of Vowels and their Movements," J. Acoust. Soc. Amer., Vol. 20, No. 4 (1948), pp. 528-535.

43. Rubenstein, H., L. Decker, and I. Pollack, "Word Length and Intelligibility," Lang. Speech, Vol. 2 (Oct. - Dec. 1959), pp. 175-178.

44. Sakai, T., and S. -I. Inoue, "New Instruments and Methods for Speech Analysis," J. Acoust. Soc. Amer., Vol. 32, No. 4 (1960), pp. 441-450.

45. Schiesser, H., "A Device for Time Expansion Used in Sound Recording," Trans. Inst. Radio Eng., Vol. AU-2, No. 1 (1954), pp. 12-15.

46. Sharf, D.J., "Intelligibility of Reiterated Speech," J. Acoust. Soc. Amer., Vol. 31, No. 4 (1959), pp. 423-427.

47. Spogen, L.R., H.N. Shaver, and D.E. Baker, "Speech Processing by the Selective Amplitude Sampling System," J. Acoust. Soc. Amer., Vol. 32, No. 12 (1960), pp. 1621-1625.

48. Stevens, K.N., R.P. Bastide, and C.P. Smith, "Electrical Synthesizer of Continuous Speech," J. Acoust. Soc. Amer., Vol. 27, No. 1 (1955), p. 207(A).

49. Vilbig, F., "An Apparatus for Speech Compression and Expansion and for Replaying Visible Speech Records," J. Acoust. Soc. Amer., Vol. 22, No. 6 (1950), pp. 754-761.

50. Vilbig, F., "Frequency-Band Multiplication or Division and Time-Expansion or Compression by Means of a String Filter," J. Acoust. Soc. Amer., Vol. 24, No. 1 (1952), pp. 33-39.

AN EXPERIMENTAL STUDY OF VIBROTACTILE APPARENT MOTION*

William Hopkin Sumby
Decision Sciences Laboratory
Electronic Systems Division
L.G. Hanscom Field
Bedford, Massachusetts

INTRODUCTION AND HISTORY

In many military situations in which communication is necessary for the successful completion of the operation, the amount of information that can be effectively transmitted concerning a particular situational component of the mission frequently is dangerously limited. One reason is that the engineering complexity of the equipment used in many such operations requires greater physiological channel capacities for its efficient use than are possible with the two sense modalities traditionally used in communications; the auditory and the visual. A possible attack on the problem would be to attempt to increase the overall informational receiving capacity of the communicating individual by directing information through an additional sense modality. The Psychological Laboratory of the University of Virginia has been concerned for the past few years with the investigation of such a possibility.

Several studies have recently been completed at that laboratory in which the communicatory potential of vibrotactile sensitivity has been investigated (17, 19, 18, 20). Vibrotactile stimulation has been delivered, rather than static tactual signals, because of its greater manipulability (more dimensional properties) and its slower adaptation process (23). Each of the previously mentioned studies (17, 19, 18, 20) has been concerned with either the determination of differential thresholds or the number of absolute judgments that readily can be made for a specific stimulus dimension of vibration. An experiment is now in progress in which the encoding potential of these different perceptions is being investigated (8).

The present study is not directly concerned with the recognition or identification of points on the dimensional continua

---

* This publication is based on a thesis submitted in partial fulfillment of a Master of Arts degree at the University of Virginia, Graduate School.

per se, but rather with the possible perceptual effects of such stimulation. There is a clinical neurological practice of tracing geometrical patterns with a stylus on the skin of a patient to determine the relative degree of "neural integrity" present. Pattishall (15) used this technique in a recent exploratory study. He traced letters of the alphabet and other patterns on the backs of five normal subjects. Tracing areas of two different sizes were used, one 5 inches by 7 inches and the other 3 inches by 3 inches. In both situations it was possible for the subject to report almost errorlessly the letter or pattern traced. The possibility of using some such a method for the transmission of simple directional or positional information became immediately apparent. A most obvious difficulty of incorporating such sensitivity into a military communications system, however, is the extreme difficulty encountered in the design of a practical movable transducer. If it were possible to induce tactual apparent movement, and thereby eliminate the need for a moving stylus, it might be feasible to develop a signaling system which would be as effective as the pressure tracing technique. This notion led directly to the construction of two vibratory matrix stimulators to be used in an investigation of such a possibility, one by Pattishall and one by the present author. Furthermore, the notion led ultimately to the major research objective of this study.

The author, using a 3-vibrator by 3-vibrator square matrix, investigated the reliability with which direction and position of stimulation could be reported when an area of the back was successively vibrated by a series of transducers. In addition, incidence of apparent movement was noted when it was experienced. The intensity of stimulation was maintained at approximately 200 μ (amplitude when fully damped by the skin), and a separation between vibrators of 2 inches was built into the apparatus. The rate of stimulation, duration of the stimulus bursts, and the duration of the silent interval between bursts were hand controlled. This resulted in considerable variability of both stimulus dimensions. The average silent interval was estimated to be about 80 msec, but no estimate can be given for the range of variability about the average. Using these stimulus values, direction of movement was always reported correctly when the matrix was centered laterally on the thoracic region of the back. The identity of the vibratory column or row activated was likewise correctly reported in all cases, although when the lumbar region of the back was included a small percent of errors did occur. Movement perceptions were reported in only a few of the total responses. Movement, during this phase of the work, was considered to have been aroused by a chance occurrence of a silent interval the duration of which approximated the optimal duration. Because of the more or less random occurrence of movement, it was apparent that, prior to the construction of a vibratory matrix, it would be necessary to manipulate systematically the critical variables to arrive at the optimal stimulus values for apparent movement. The purpose, then, of the research here reported was to determine the relationships among the stimulus properties optimal for the arousal of tactual apparent movement when induced by spatially discrete vibratory stimuli.

A survey of the pertinent literature revealed that apparent motion induced by vibrotactile stimulation has been previously reported by only three authors. Petzoldt (16), the first of these, reported apparent movement traveling from one hand to the other or from the hand to the foot when vibrating units were successively energized. Katz (11) found that synthetic motion could be aroused when an intensitive differential was established between two vibrating units simultaneously activated, the movement always traveling in the direction of the less intense stimulus. Bice (3), reporting the results of a vibratory tracking task, indicated that it was possible to confine the subject in an apparent belt of moving vibration by positioning a series of vibrators around the chest and activating the transducers successively and with equal intensity. None of these authors, unfortunately, correlated the effects of the manipulation of the stimulus properties with the arousal of the movement perception, nor did they report quantitatively the characteristics of the stimulus dimensions used in their research. These appear to be the only works reporting an experiment concerned with vibrotactually induced apparent motion.

Tactual apparent movement aroused by static or nonvibrating stimuli has been reported independently by several investigators. The first two of these (10, 22) reported illusions of movement when tactual stimuli were delivered to the skin, but they did not specify any of the stimulus conditions used in their research. Wertheimer's study (24), reporting the visual "phi" phenomenon, appeared to afford the impetus for the trend to specify quantitatively the stimulus dimensions for apparent motion when he stated that a temporal interval of 60 msec between the successive presentation of radiant stimuli would be optimal for its arousal. Korte's confirmation of Wertheimer's results (12) and his statement of the laws for the arousal of apparent visual movement followed almost immediately. It was at this point in time that studies specifically concerned with the arousal of tactual apparent movement were initiated. Most of these studies have attempted either to confirm or refute the major variable conditions for movement as formally stated by Korte: 1) the spatial separation for apparent movement (the phi phenomenon) varies directly with the intensity of the stimuli; 2) intensity of the stimuli varies inversely with the temporal interval between stimuli; and 3) the temporal interval between stimuli varies directly with the spatial separation between stimuli.

The first attempt to specify quantitatively the optimal stimulus dimensions for the arousal of tactual apparent movement was carried out by Benussi (2). His intention was to duplicate in the field of touch the work of Wertheimer in the field of vision. Movement was reported by his subjects in situations in which the temporal interval between static stimuli ranged from 240 to 260 msec, and spatial separations ranged from 3 to 130 cm. It is interesting to note that Benussi, in opposition to Wertheimer's major conclusion, considers the spatial separation between stimulators, rather than the temporal separation, to be the most critical variable for perceived movement. Burtt (5), using a stimulus duration of 60 msec throughout his series of experiments, obtained reports of movement from his subjects when he varied the temporal silent interval from 15 to 72 msec and the spatial separation between transducers from 8 to 16 cm. His results tend to confirm the notion that the relationships expressed in Korte's laws for visual movement may be generalized to other sense modalities. Furthermore, his work on auditory apparent movement (4) again offers evidence supporting the possible universality of Korte's laws. Whitchurch (25), in her effort to determine the elementary conditions for the arousal of cutaneous apparent movement, found that if a stimulus duration of 150 msec was used, the optimal interval between stimuli appeared to be between 100 and 75 msec, depending upon the stimulus pressure applied. She reports that, when using her optimal stimulus conditions, only 67 percent of the observations included any reports of movement, and concludes that, "the cutaneous perception is less fundamental and compulsory than the corresponding visual perception." Andrews (1) reports that his subjects all experienced movement of some kind during his investigation, but that the optimal conditions varied tremendously from subject to subject.

A most comprehensive paper reporting a systematic study of tactual apparent motion is that by Hulin (9). Hulin summarizes the analysis of 13,500 judgments, and reports that in 29.7 percent of the responses some evidence of apparent movement was reported. The duration of the static stimulus contact was held at 150 msec throughout the investigation. He varied temporally the successive presentations of the stimuli from simultaneity to a positive silent interval of 300 msec. In other words, the onset of the second stimulus succeeded the ending of the first by 300 msec. In addition, a stimulus overlap of 75 msec was included. In this situation the onset of the second stimulus preceded the termination of the first by 75 msec, i.e., the second stimulus was applied before the first was removed from the skin.* His publication includes the interesting observation that "the only definite quantitative relation shown by the data is that the minus 75 msec temporal interval is exceptionally favorable for the arousal of tactual apparent movement." It is within this temporal interval that optimal movement was reported in approximately 64 percent of the observations.

---

* Referred to by Hulin as a "minus silent interval."

Since Hulin's paper two other relevant investigations have been reported, one by Neuhaus (14), the other by Tschlenoff (21). Both of these papers were concerned with the stimulus properties optimal for the arousal of apparent movement. Neuhaus found that movement could be aroused with intervals between stimulation ranging from 0 to 500 msec, the spatial interval between transducers varying directly with the interval. Tschlenoff was concerned primarily with the nature of the experience aroused by tactual stimulation. He noted, as did Neuhaus, that several types of perceptual patterns could be aroused between those of succession and simultaneity.

Each of the studies reported has the common characteristic of extreme variability of results. The findings do not permit a conclusive statement of the tactual analogues to Korte's law. The evidence accumulated to this point indicates that there is no specific set of stimulus conditions which, when delivered to the skin, will invariably arouse the perception of motion. There are wide individual differences among subjects in the reporting of motion, and equally wide differences among studies as to the dimensions of the optimal stimulus conditions. The variability of results among the several studies reported suggests the possibility that apparent movement may be aroused by many different stimulus combinations. It indicates that, although apparent motion may be induced with a number of relationships among stimulus properties, there is no combination which will invariably arouse that perception.

Some authors, including Gilbert (6) and Whitchurch (25), suggest the possibility that the arousal of tactual motion may partially be a function of such factors as set, past experience, motivation, and other loosely defined variables. Factors such as these were not specifically controlled in the present work, except insofar as they were influenced by the pre-experimental instructions. It is believed that the range of physical stimulus conditions must be sufficiently specified before the effect of such extra stimulus variables can in any sense be investigated. Hulin's work suggests that such specifications can be eventually accomplished.

Hulin was the first investigator to utilize a pair of static transducers, the onset of which were neither simultaneous nor separated by a positive silent interval, but included a stimulus overlap. Hulin's major finding led to the notion that possibly the range of durations examined prior to the Hulin work either did not include the optimal interval or the temporal steps investigated were not adequately refined. The possibility of a temporal stimulus overlap arousing compulsory movement when vibratory stimuli were substituted for the static led to the selection of a wide temporal interval range to be tested. It was decided that exploratory work would include intervals from simultaneity of onset to a positive interval (time between termination of one stimulus and onset of the other) of 200 msec in steps of 100 msec, using a constant stimulus duration of 200 msec. Figure 1 graphically illustrates this design. When the results of the exploratory work were analyzed the range of intervals could be further restricted about the interval at which the greatest frequency of movement was reported.

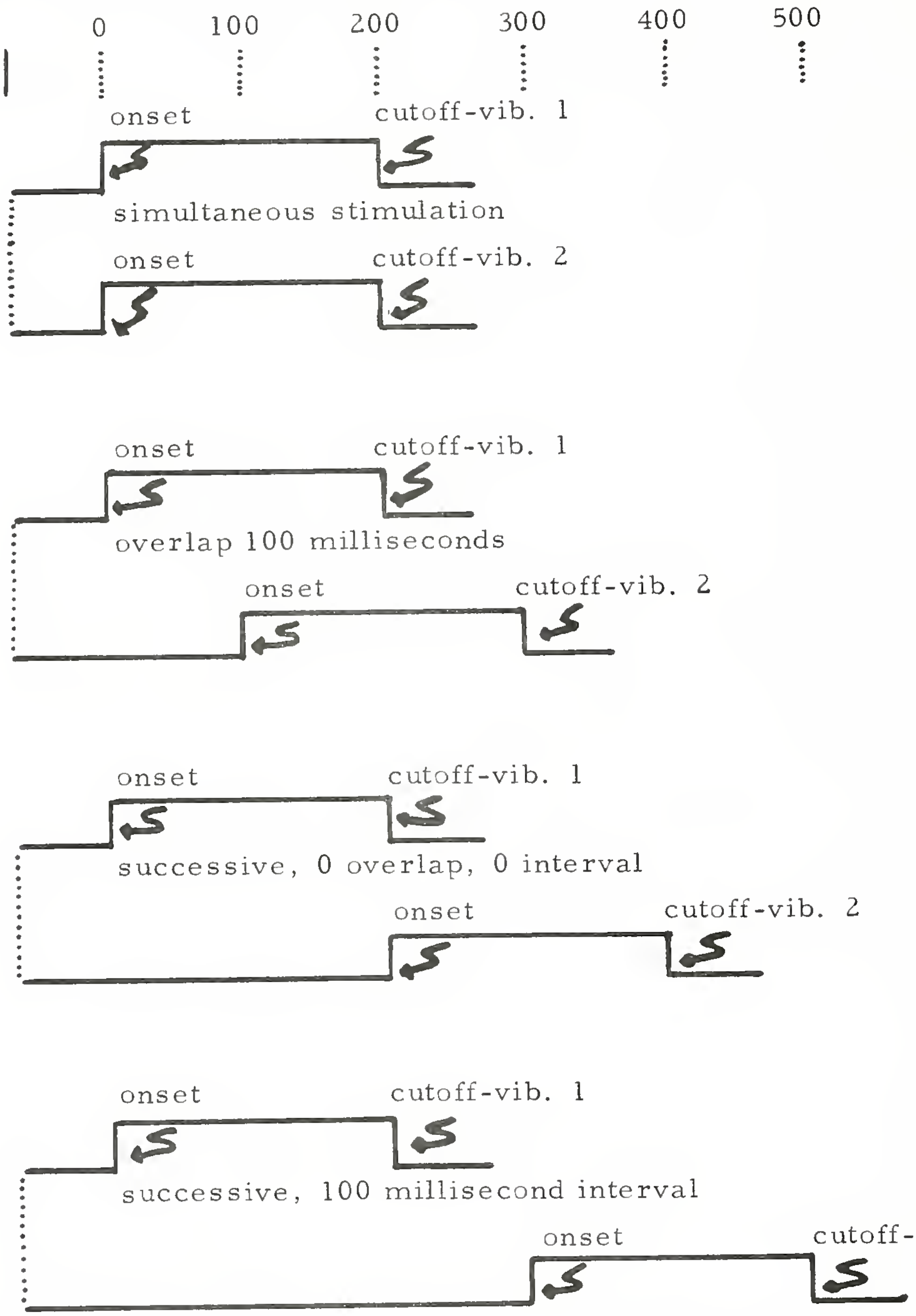


Figure 1. Schematic showing various temporal intervals between onset and cutoff of stimuli. Duration in milliseconds shown at top.

Again, the immediate purpose of this research was to determine, if possible, the optimal vibrotactile stimulus conditions for the arousal of apparent tactual movement. The long term objective of the research is in the direction of the development of a vibratory communication system designed for the transmission of relatively simple information using the phenomenon of apparent motion in the perceptual signal.

APPARATUS AND PROCEDURE

The transducers used throughout this experiment were constructed to oscillate with the action of a 6 V ac relay coil (Guardian, type 200-6A) when the coil was energized by 60-cycle current (Figure 2). They were modified versions of the vibrators used by Spector (20). A strip of highly tempered steel spring (2-1/2 inches by 1/4 inch) was soldered to an angle iron, upon which the coil had been factory mounted. An iron armature, 7 mm in thickness, was soldered to the strip, directly in line with the core of the coil. A separation of 2-1/2 mm was maintained between the armature and the core. The actual skin contractors were plastic hemispheric buttons, 17 mm in diameter, which were cemented to the spring in line with the core of the coil and the armature. The resonance frequency of this construction was approximately 60 cps. To reduce unreliability of performance, because of the correspondence between the resonance frequency of the vibrators and the energizing current, a lead clamp was attached to the spring thereby changing the resonance frequency to 37 cps.

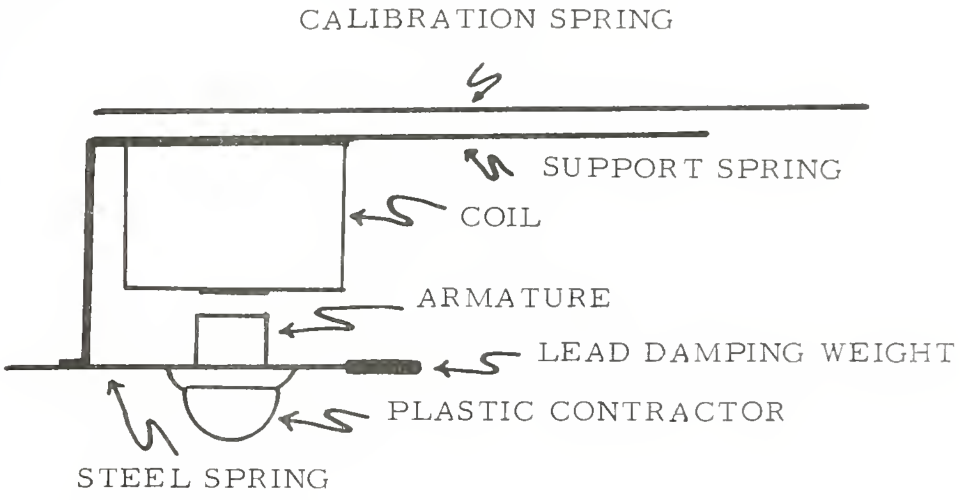


Figure 2. Diagram of the Vibrator Assembly Used in the Experiment.

The angle iron mount of the coil was soldered to a hacksaw blade, 31 cm long, which served as a spring mounting for the vibrator. The spring was then bolted through a short piece of wood, 5 mm thick and about 20 cm from the vibrator, to a superposed second blade. By this construction skin contact pressure could be readily adjusted. When the plastic contact button was positioned on the skin of a subject, and sufficient pressure applied so that the tips of the two blades at the transducer end just made contact, the static pressure on the skin was 100 grams. The blade attachments were then clamped to a 62-cm metal gooseneck to facilitate vibrator positioning. The gooseneck, in turn, was clamped to a 4-foot steel rod support held rigidly in position by a heavy cement base.

The variables manipulated in the experiment were the intensity of vibration, the temporal interval between stimulations, and the spatial separation between transducers. Since frequency of vibration was to be held constant, 60-cycle house current was used throughout.

The intensity of stimulation, measured in terms of amplitude of vibration, was set at 120, 240, and 360 μ. These amplitudes were selected because the threshold differences between successive settings represent equal numbers of j.n.ds.; approximately 4 (19). These amplitude measurements were correlated with voltage settings read after being stepped down from 110 V to 17 V by a transformer, and further varied by an autotransformer between 12 and 17 V. The amplitude calibrations were optically determined, using a magnification of 17X in a binocular microscope in conjunction with a Strobolux (general Radio, 648A) as an intense, pulsating light background. The vibrators were damped for calibration on a piece of sponge rubber 1 inch thick and mounted on wood. This combination gave evidence of having comparable damping characteristics to the surface of the back in the thoracic region.* The vibrators were recalibrated twice during the experiment and no significant change was apparent in any of the average amplitudes.

The duration of each vibratory stimulus burst was constant at 200 msec. This was accomplished by a calibration of segments of exposure sections of a pendulum-type (Whipple) tachistoscope (Figure 3). The exposure time was measured by allowing a 100-cycle audio signal, generated by an oscillator (Hickok, model 198) to pass through a relay to a Potter counter. When the tachistoscope was released from its normal ready position, each cutout

---

* Vibrator action was observed on several different damping materials, and compared with calibrations made, with difficulty, on the back. It was found that the sponge rubber and the back tissue damped the vibrators in about the same manner throughout the intensitive scale used.

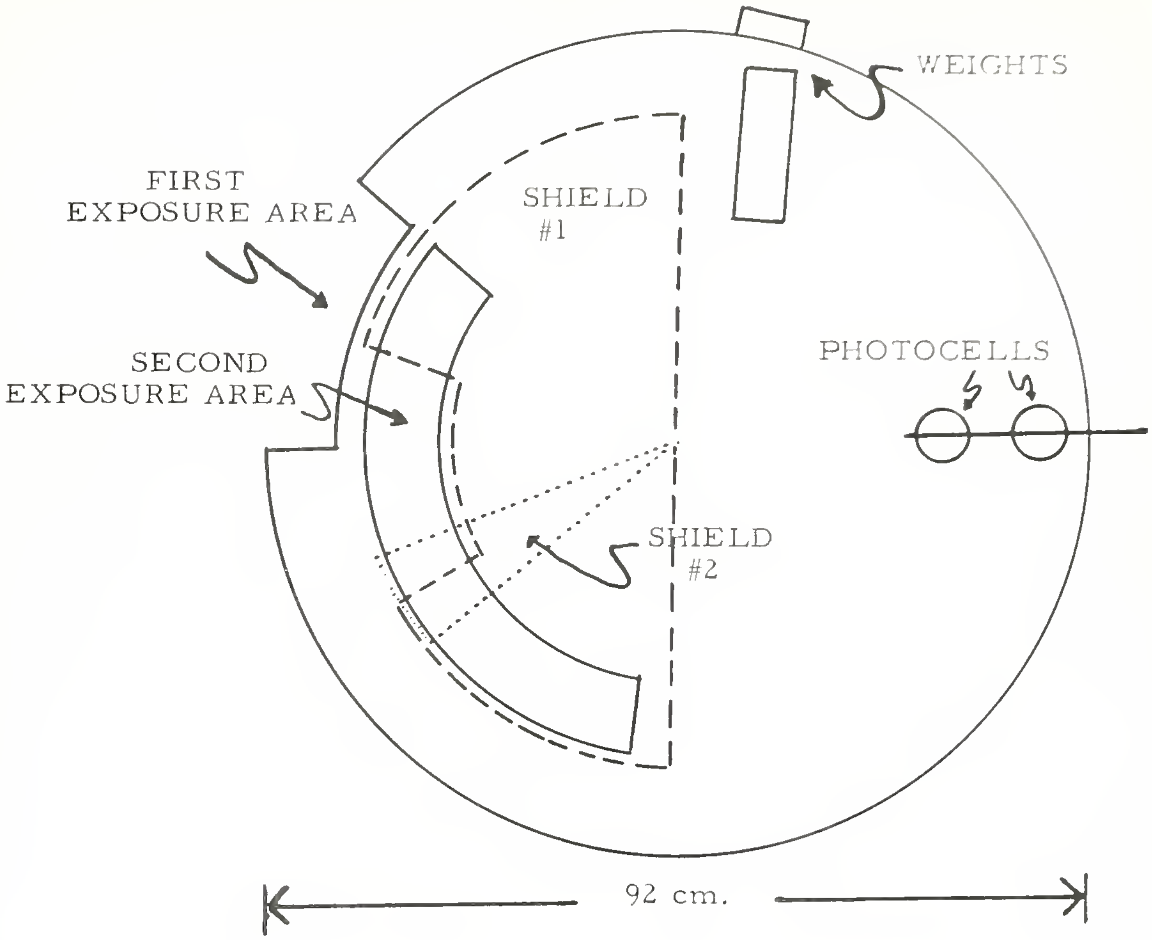


Figure 3. Diagram of the modified Whipple tachistoscope used. The tachistoscope was mounted on the side of a heavy table, and turned in a clockwise direction. The photocells were mounted on metal supports, and were approximately 2 inches in front of the disc.

For a full description of the tachistoscope see - Whipple, G.M. Manual of Mental and Physical Tests, 2nd ed., Pt. 1. Baltimore: Warwick and York, 1914.

section of the disc allowed light from a 100-W projection lamp to fall briefly upon a photoelectric cell as the sector passed the cell (Weston, model 594). The exposure time was a function of the size of the cutout section and the position of the wheel relative to its starting point. Position was a variable because of the gradual acceleration and deceleration of the rotation. The output of the photocell was sent to an amplifying system, including a transistor (Raytheon CK 722) and a vacuum tube (6J5), where the current was amplified sufficiently to activate a relay and allow current to pass to the vibrator. The tachistoscope afforded an extremely rapid and reliable exposure onset and cutoff of the photocell, allowing presentation of the stimulus with a total variation of but 2 msec.

The spatial separation of the transducers was carefully measured each time the vibrators were positioned on the skin. The area of the body used was always the thoracic area of the back. Three different separations were used, 4, 12, and 22 cm, the latter being about the maximum "flat" lateral surface of the average back. Vibrators were placed on opposite sides of the spinal column during the main phase of the research. Evidence coming from exploratory work indicated that when the vibrators were positioned longitudinally on the thoracic surface the incidence of movement reported was equivalent to that when the vibrators were arranged laterally.

The temporal interval between bursts was controlled by the manipulation of two aluminum masking shields attached to the axle of the disc. The first (outermost) exposure section of the disc was cut so that it would allow a 200 msec vibratory burst when the wheel was released in a clockwise direction from its locked position. The second section was so cut that its maximum exposure period was 450 msec. Shield No. 1, having an exposure section of 220 msec at the slowest position of its arc, was then swung from the center of the disc in such a manner that the duration between the onset of bursts could be controlled and at the same time set the exposure area of the second stimulus. For example, if the beginning and end points of one exposure section were along the same radii of the circle, as the beginning and end points of the other section, the exposure periods would be simultaneous. If the beginning point of the inner slip was opposite the center of the outer slit there would be a stimulus overlap of approximately 100 msec. In this way it was possible to calibrate the overlap interval in steps of 20 msec, from a 20-msec overlap to a 180-msec overlap. The duration of the second burst was kept constant by changing the position of a second shield. In this manner it was possible to increase or decrease the size of the opening, and thus control the exposure period, the duration varying with the shifting of the weight of the shields.

Five male subjects, ranging in age from 19 to 20 years, participated in the experiment. W and G were graduate students in psychology, D, A, and J were undergraduates. The two graduate students were familiar with the proposed plan of research, including the purpose of the work. The three undergraduates were not familiar with the work in any way prior to their signing as subjects. The first few sessions held with the undergraduates simply involved the presentation of successive stimuli. In each case, after many presentations, the subject reported a movement perception from one vibrator to the next and was asked to describe the experience qualitatively. After this initial experience of movement such reports became quite common. They were then informed briefly as to the purpose of the experiment and were reminded of commonplace occurrences of visual apparent movement, such as witnessed in motion pictures and directional signs. With this brief introduction they were introduced to the experimental situation.

A total of 2025 stimulus presentations were given, 405 to each of the 5 subjects. Since the experimental design included 9 different intensities, 3 different overlap times, and 3 spatial separations, there were 81 different possible stimulus complexes. Each of these complexes was presented to each subject 5 times.

Exploratory work suggested that it would be possible to distinguish three gross categories of report. The first category included those perceptions in which there was no movement experienced. The possibilities here were a) the perception of two spatially discrete vibratory bursts whether temporally discrete or not, without any evidence of movement between them, and b) the perception of a single structured cutaneous pattern following simultaneous presentation, in which direction of movement could not be specified. The second of the three categories included perceptions of movement, but in which the movement did not completely cover the distance separating the two transducers, or in which a break occurred in the movement, such as a "dead" spot at about the middle of the spatial separation. The third type of perception was a full movement between the vibrators, in which the direction of movement could be definitely specified, and in which the movement was uninterrupted in its course. The subjects were informed of the three gross possibilities and reported each presentation as it led to one of the three categories of report. Thus, the responses were; 1) no directional movement, 2) movement, but not complete, and 3) good, full movement between the vibrators.

Paired stimulations were presented at a rate of approximately one every 30 seconds. The intensitive and temporal variables were randomized throughout each experimental session. The spatial variable was held constant for each session because of the inconvenience of changing vibrator positions. The sequence of spatial separations was different for each subject, however. Each session lasted approximately 45 minutes, including at least one 5-minute rest during the period. The subject was seated in a chair and rested his head on a cushioned table top in such a manner that his back surface was about parallel with the floor. The vibrators were positioned horizontally on the back, equidistant from the spine. The experimenter gave a verbal ready signal and then released the disc. The subject's response was recorded, the intensity and interval time were changed, and the ready signal was given again for the next presentation.

RESULTS

The data to be reported and discussed in the following pages are presented in Tables I, II, and III and are further represented graphically in Figures 4, 5, and 6. In each of these figures the abcissa represents the temporal overlap of the two stimuli, while the ordinate shows the number of responses of full movement. It will be recalled from the previous section that the subjects were instructed to categorize their perceptions under one of three headings; 1) no movement, 2) discontinuous or partial movement, and 3) complete or full movement. Each of these tables represents the responses made by the subjects for a particular spatial separation, e.g., 4 cm, while the intensitive and temporal variables were randomly manipulated. Each row of each of the five subject columns shows the five responses made by one subject for each of the stimulus complexes. The three columns at the right show the frequency of the combined responses for each category. Complete summary totals are given in Table IV. All subjects reported full movement for some of the presentations, and reports of full movement were made at least once for practically all of the stimulus complexes.

When a stimulus dimension from one continuum was put into combination with each possible pairing of single dimensions from the two remaining continua (3 intensities, 9 temporal overlaps, and 3 spatial separations) there were 81 stimulus complexes available. There were 2025 judgments made during the experiment, 405 by each of the 5 subjects. A simple frequency count of the results revealed that responses of complete movement were made in 24.6 percent of the cases. When the frequency of partial movement responses was combined with the full movement response totals, 62.4 percent of the toatl presentations resulted in some experience of movement. Table V gives the percent of movement response and the combined full and partial movement responses. The obvious extreme subject variability is probably attributable to the different criteria used in judging the perception by the individual subjects.

The results demonstrate quite conclusively that the optimal temporal interval lies between 80 and 120 msec of stimulus overlap regardless of the intensity or spatial separation used. With the amplitude of vibration set at 360 μ the incidence of full movement is greatest for all spatial separations when the temporal overlap is 100 msec. The results are not as dramatic when the other amplitudes are delivered. It is true, however, that the optimal interval does fall within or close to the range of 80 to 120 msec. Figure 7 shows frequency of movement for each temporal overlap when the intensitive and spatial variables are combined. Again, it is made apparent that the optimal interval

TABLE I

TABULATION OF RESPONSES MADE BY SUBJECTS
4 CENTIMETERS SPATIAL SEPARATION

Temporal overlap	Subjects					Totals		
	W	D	A	J	G	1	2	3
Amplitude 120 microns								
180	21111	12121	21111	11111	11111	21	4	0
160	11121	33222	13112	11121	12111	15	7	3
140	13221	21333	31111	13111	12212	13	6	6
120	33312	33333	22231	13221	22221	5	10	10
100	33322	13323	22232	11112	32213	6	10	9
80	23111	21331	22222	11221	21213	10	11	4
60	21111	22331	12122	12211	22222	10	13	2
40	22111	33222	31132	21211	22323	8	11	6
20	22321	22322	22133	22111	23322	5	14	6
						93	86	46
Amplitude 240 microns								
180	11111	32211	22112	11111	11111	19	5	1
160	33311	33111	21113	11211	12111	16	3	6
140	31311	31232	11221	12311	11111	15	5	5
120	33333	33333	21223	33233	32323	1	6	18
100	33333	32233	23132	23333	32323	1	7	17
80	32333	33333	33233	22223	33223	0	8	17
60	22222	32322	32323	22213	23232	1	16	8
40	12231	12132	33212	11122	22222	8	13	4
20	21211	22321	23233	21112	12222	8	13	4
						69	76	80
Amplitude 360 microns								
180	11111	11211	11111	11111	11111	24	1	0
160	13331	22212	12112	12211	11111	14	8	3
140	11133	31232	12113	33121	11112	13	5	7
120	33333	23323	33332	33333	33112	2	4	19
100	33332	23332	33333	33332	33323	0	5	20
80	23332	23333	32323	12223	12323	2	10	13
60	23232	23333	22213	12222	12223	3	14	8
40	23222	22222	32133	21123	22322	3	16	6
20	22221	22222	22333	22111	12212	6	16	3
						67	79	79

overall totals 229-241-205

TABLE II

TABULATION OF RESPONSES MADE BY SUBJECTS
12 CENTIMETERS SPATIAL SEPARATION

Temporal overlap	Subjects					Totals		
	W	D	A	J	G	1	2	3
Amplitude 120 microns								
180	11111	11211	11111	11121	11111	23	2	0
160	31112	22211	21122	12121	11121	14	10	1
140	11311	33222	12121	21211	11122	13	9	3
120	33221	32312	11132	13222	12121	9	10	6
100	21311	33322	12231	11211	22232	9	10	6
80	12123	33333	12332	11221	22212	7	10	8
60	13111	22132	22322	23221	32222	6	14	5
40	21112	22223	22122	22212	12212	7	17	1
20	13112	13322	13212	12222	22222	7	14	4
						95	96	34
Amplitude 240 microns								
180	11111	11111	11111	11211	11111	24	1	0
160	11331	11111	11111	11111	11111	23	0	2
140	33313	32121	32122	33211	21313	8	7	10
120	33333	33232	33321	13332	11233	4	5	16
100	33322	33222	23232	32223	33223	0	13	12
80	33332	32213	32223	23322	33323	1	10	14
60	32222	23222	22222	33312	33233	1	15	9
40	22222	12123	32222	11222	22222	4	19	2
20	12212	23322	21323	22212	22322	4	16	5
						69	86	70
Amplitude 360 microns								
180	11111	11111	11111	11111	11111	25	0	0
160	11111	11111	21112	32111	11111	21	3	1
140	33113	23312	33111	22332	12111	10	6	9
120	33323	33322	22221	23323	11313	4	9	12
100	33333	32322	33333	22333	33133	1	5	19
80	33333	33223	33233	22323	23233	0	8	17
60	23322	32213	21232	23222	22333	2	14	9
40	22222	22222	31233	11112	22332	5	15	5
20	23212	23232	12222	11221	22333	5	14	6
						73	74	78

overall totals 237-256-182

TABLE III

TABULATION OF RESPONSES MADE BY SUBJECTS 22 CENTIMETERS SPATIAL SEPARATION

Temporal overlap	Subjects					Totals		
	W	D	A	J	G	1	2	3
Amplitude 120 microns								
180	11111	11111	11111	11111	11111	25	0	0
160	13211	12111	11111	11111	21111	21	3	1
140	22221	11211	11111	11111	22121	17	8	0
120	13121	11323	21122	12211	12221	12	10	3
100	22211	23322	13331	11212	22122	8	12	5
80	21211	13221	23122	11111	23222	11	11	3
60	23121	22122	22331	12211	33222	7	13	5
40	12212	33322	12222	22221	22222	4	18	3
20	11211	22222	22221	12221	22222	7	18	0
						112	93	20
Amplitude 240 microns								
180	11111	11111	11111	11111	11111	25	0	0
160	13222	11211	21111	11111	11111	19	5	1
140	33111	11111	11111	11111	12111	22	1	2
120	33213	13311	21111	33212	12222	10	8	7
100	33333	12223	23232	32332	22222	1	13	11
80	22222	12321	22122	23321	22322	4	16	5
60	22322	32312	33232	23223	23223	1	14	10
40	22223	22222	12221	23312	22222	3	19	3
20	22121	23212	31221	11221	22222	8	15	2
						93	91	41
Amplitude 360 microns								
180	11111	11111	11111	11111	11111	25	0	0
160	11313	13111	11111	11111	11111	22	0	3
140	33313	13111	31111	13211	21311	15	2	8
120	33322	12211	11131	31111	22222	11	9	5
100	33332	23322	33222	31223	22223	1	13	11
80	2222	13312	23322	22222	22332	2	17	6
60	33223	22122	32331	22231	23232	3	13	9
40	13232	12222	11222	22221	32233	5	15	5
20	12312	22212	32122	23211	22322	6	15	4
						90	84	51

overall totals 295-268-112

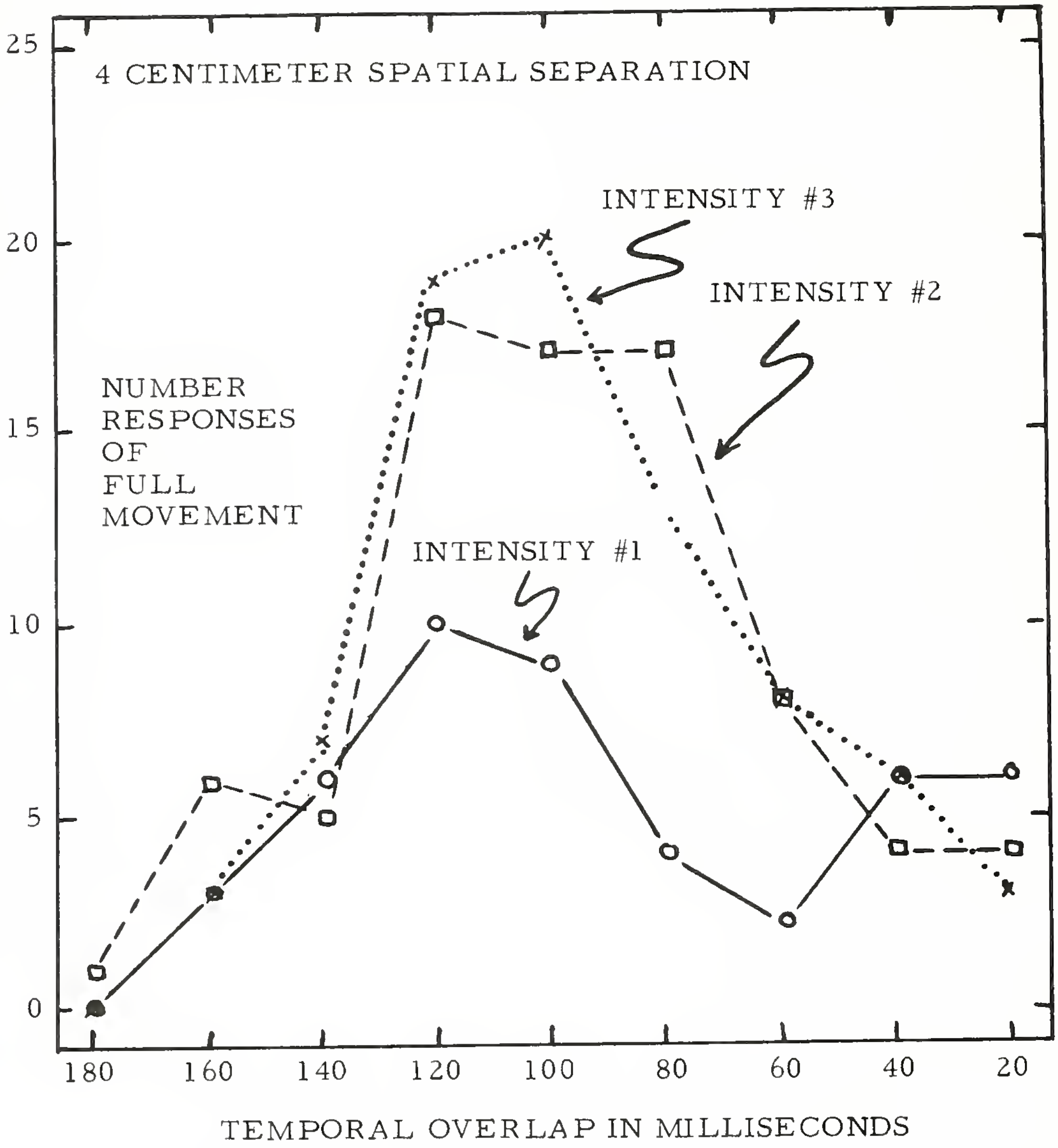


Figure 4. A graph showing the number of responses of full movement as a function of the temporal overlap of the stimuli. The parameter is the intensity of the vibration.

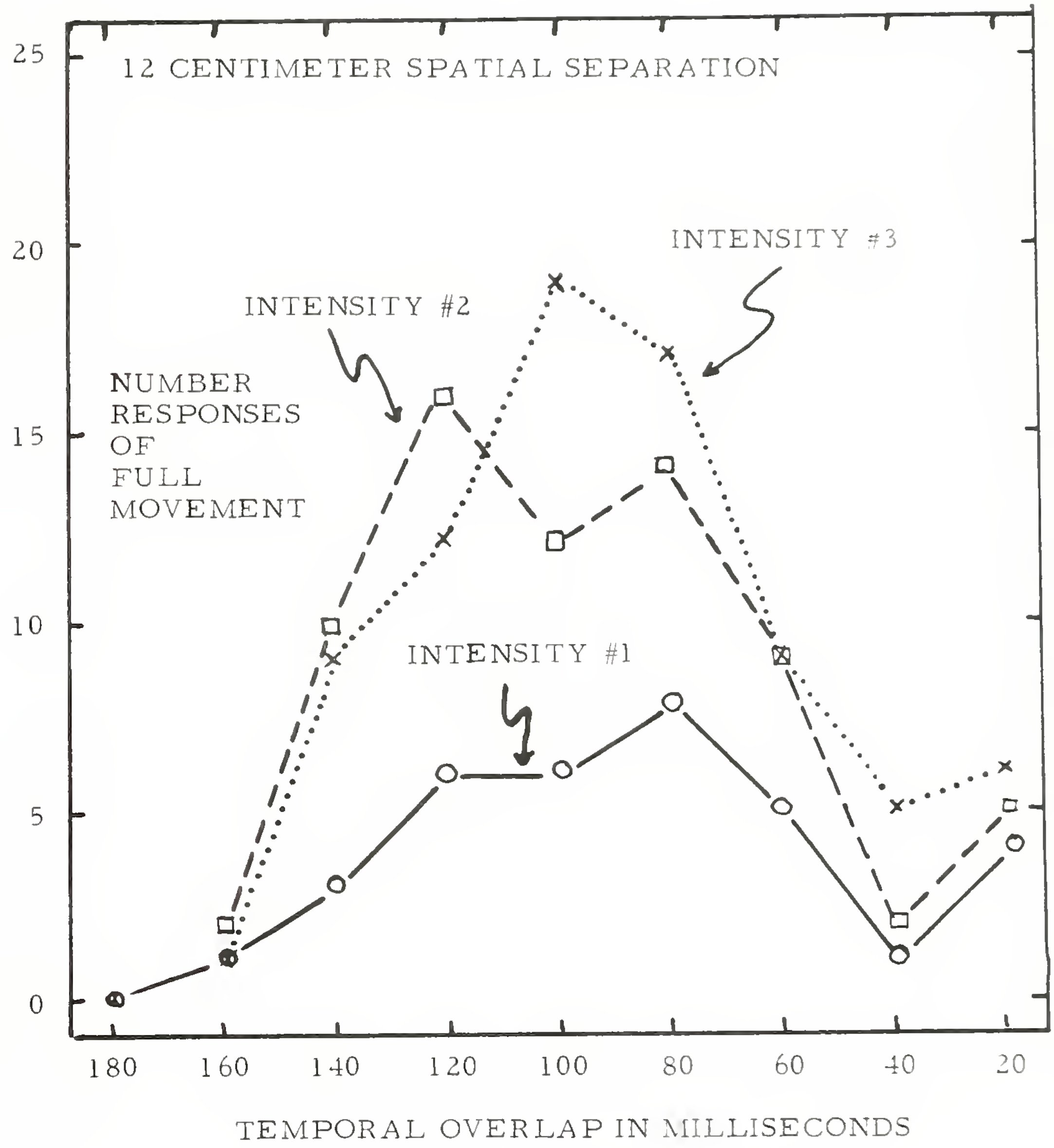


Figure 5. A graph showing the number of responses of full movement as a function of the temporal overlap of the stimuli. The parameter is the intensity of the vibration.

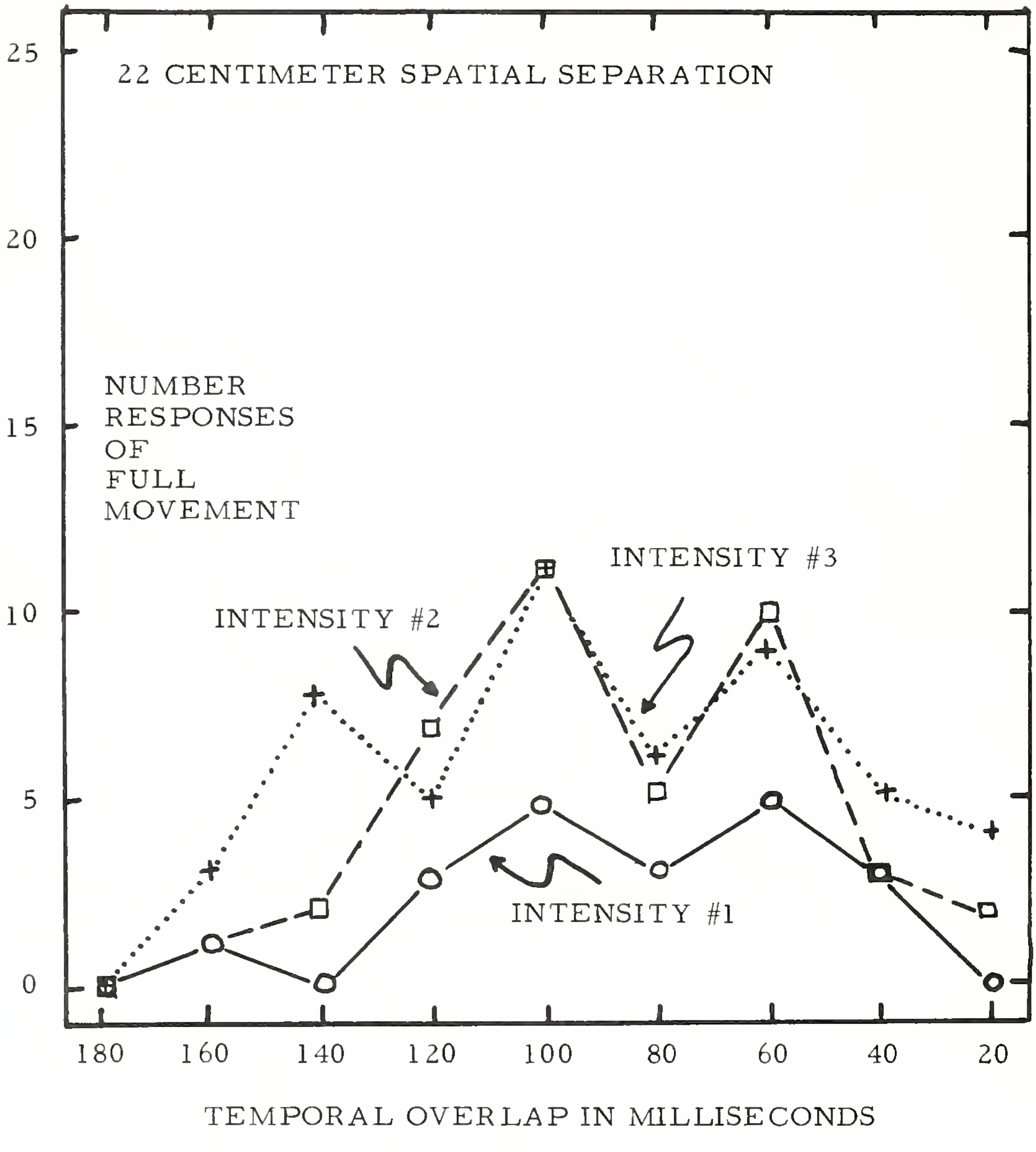


Figure 6. A graph showing the number of responses of full movement as a function of the temporal overlap of the stimuli. The parameter is the intensity of the vibration.

TABLE IV

A SUMMARY TOTAL OF EACH MANIPULATED VARIABLE

Milliseconds of overlap	Responses 1	2	3
180	211	13	1
160	165	39	21
140	126	49	59
120	58	71	96
100	27	88	110
80	37	101	87
60	34	126	65
40	47	143	35
20	56	135	34
	761	765	499

Centimeters of spatial separation			
4	229	241	205
12	237	256	182
22	295	268	112
	761	765	499

Microns of amplitude			
120	300	275	100
240	231	253	191
360	230	237	208
	761	765	499

TABLE V

A SUMMARY OF THE FREQUENCY OF OCCURRENCE OF RESPONSE 3 AND RESPONSE 3 + 2 FOR EACH SUBJECT

Subjects	W	D	A	J	G
Response 3	132	120	98	69	79
Response 2 + 3	256	286	249	210	263

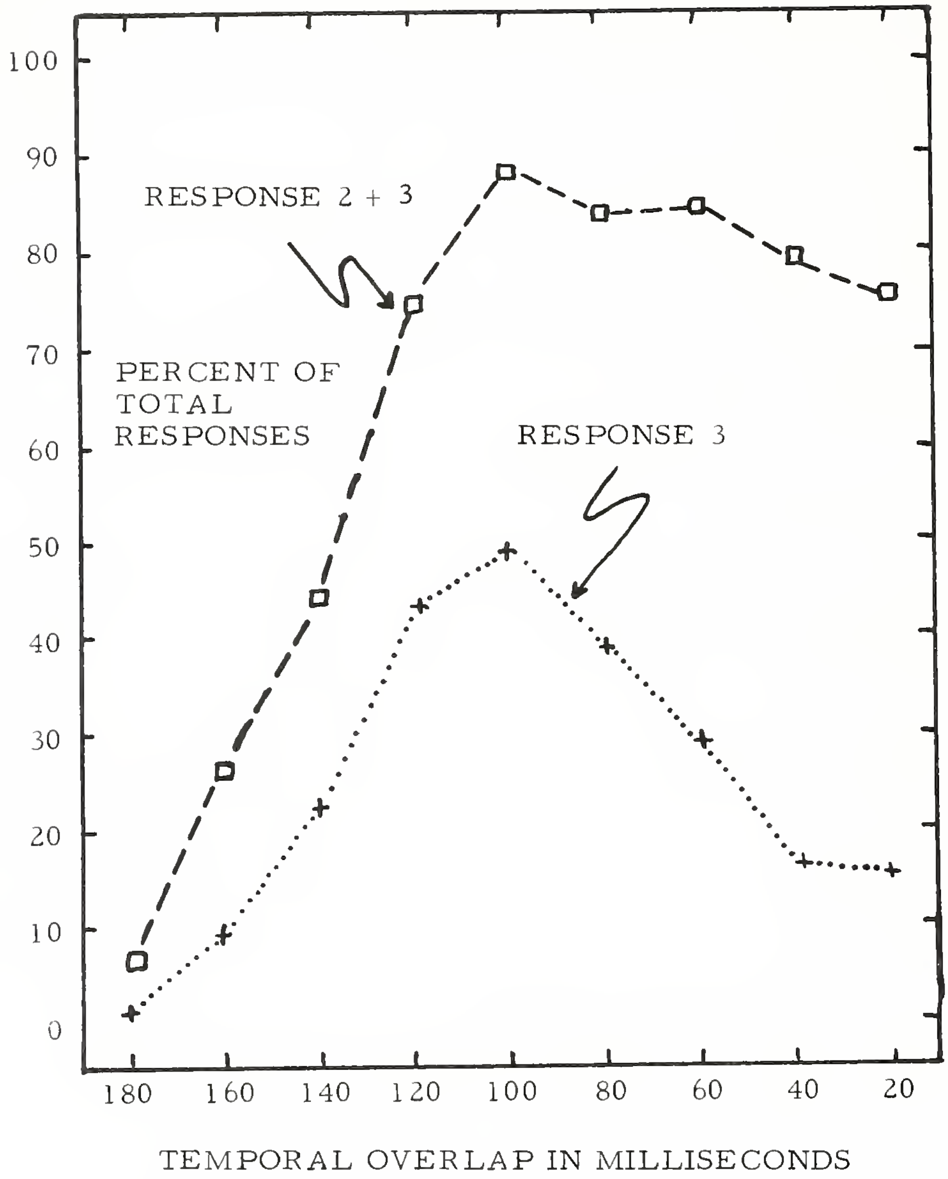


Figure 7. A graph showing the percent of responses falling in Category 3, and the percent falling in Category 2 plus 3.

probably lies somewhere between 80 and 120 msec of overlap. The mode for the combined responses occurs at the 100-msec overlap for both the number of full movement responses, and the combined full and partial movement responses. The amount of movement reported at the 100-msec overlap is 48.9 percent and for the combined, full and partial, it is 88.0 percent.

The results somewhat support the conclusion that the spatial separation is not as critical as the temporal interval, although Figure 8 does show that the incidence of movement decreases as the spatial separation between transducers is increased. The optimal temporal interval, however, remains approximately the same for each of the spatial settings. The functions indicate that the spatial range in which full movement may be aroused is large, with very little difference, if any, in the number of movement responses between 4- and 12-cm separations.

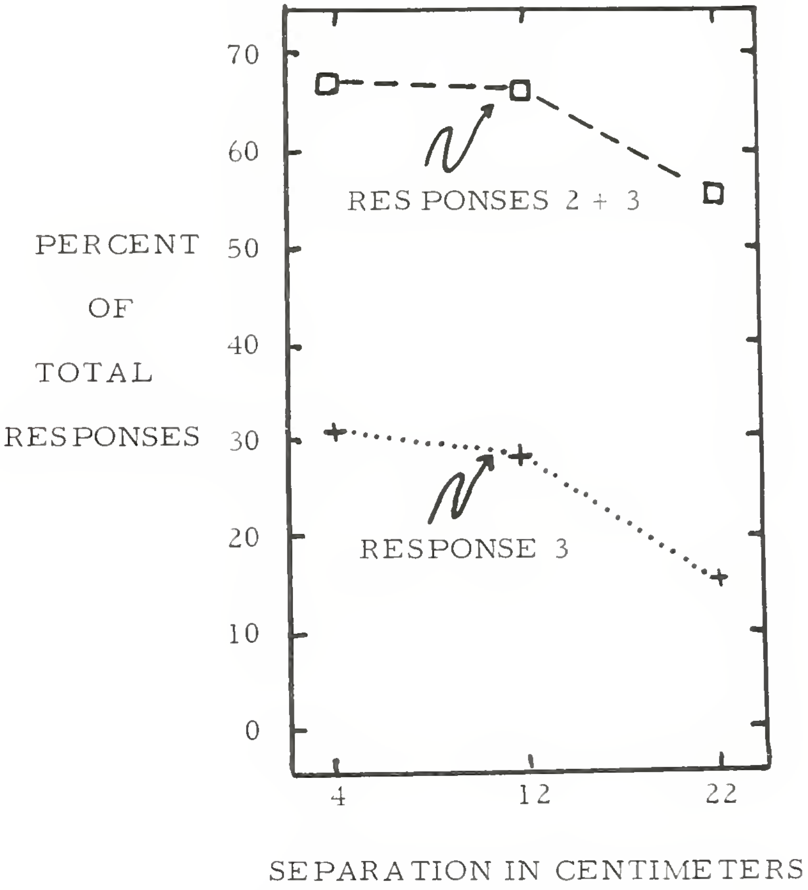


Figure 8. *A graph showing the percent of responses made as a function of the distance of the spatial separation between transducers.*

The results also lend support to the conclusion that the intensitive variable is not as critical for the arousal of apparent movement as is the temporal interval. The results indicate that once the intensity is raised sufficiently above the 100 percent absolute threshold the amplitude of vibration is of little importance (Figure 9). In many cases, in which the lowest amplitude was delivered, the subjects reported that they could not detect both vibrators when they were energized. When the amplitude was then doubled this report was never given.

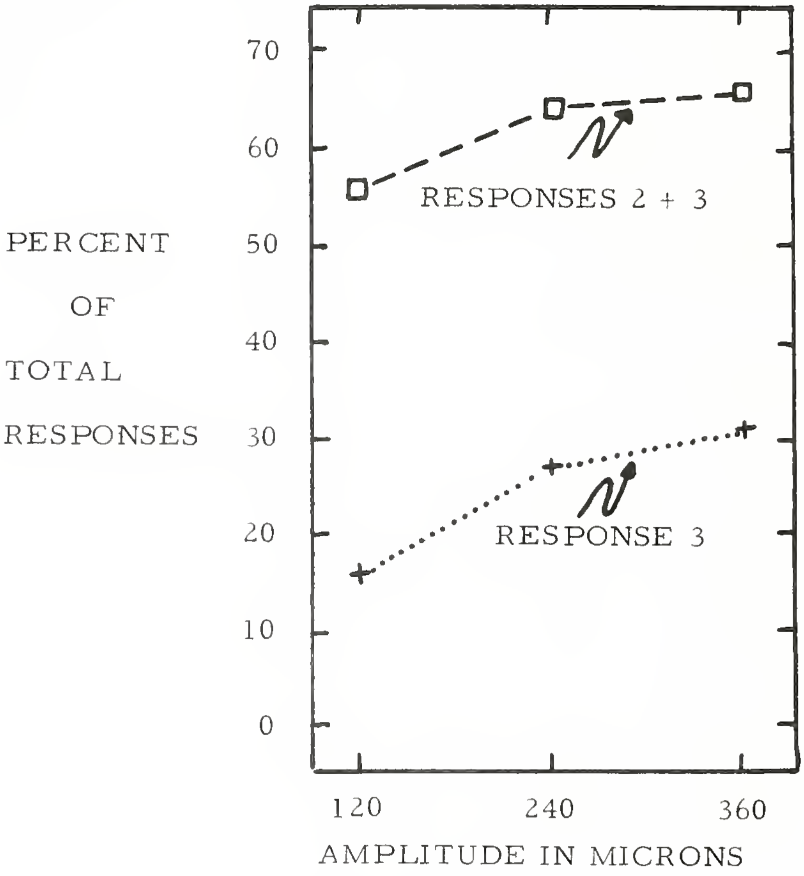


Figure 9. A graph showing the percent of responses made as a function of the amplitude of vibration.

DISCUSSION

The specific purpose of this research was to determine the stimulus conditions optimal for the arousal of vibrotactually induced apparent movement; to define, if possible, the vibrotactile analogues of Korte's laws for visual movement. It was postulated that, since the stimulus relationships critical for the arousal of visual synthetic motion can be specified quite adequately (7), it would likewise be possible to state the stimulus relationships necessary for the arousal of apparent movement through other sense modalities. Several experimenters investigating such possibilities have expressed the opinion that Korte's laws are valid both for visual perception and for other senses, especially Burtt in audition and touch (4, 5). Hulin claims, however, that the verification of Korte's laws for touch is doubtful, and Mathiesen's results from an auditory study suggest that the laws are not valid in that sensory area (13). These conclusions were reached because of low incidences of reported movement for literally thousands of stimulus presentations. Korte, in his statement of the laws for the arousal of perceived movement, describes the effect of the manipulation of each of several stimulus variables on the perception of movement: spatial separation between stimuli, temporal interval between stimuli, intensity and duration of stimulation. Wertheimer and Korte considered the interval between stimulation to be the prepotent variable, although they did consider the quantitative specification of the others to be essential for the arousal of visual apparent movement.

Before a discussion of the findings can be undertaken, the limitations imposed by the somewhat practical orientation of the experiment and the reasons for such limitations will be explained briefly. It will be recalled from the introductory section that the ultimate goal of this research is the development of a vibratory matrix to induce apparent motion and to be used as a transducer mechanism for the transmission of brief, simple information via the cutaneous pathways. Such a matrix to be maximally effective would have to be positioned on a relatively homogeneous, large surface of the body so that the stimulators would be equally effective over the area, and so that it would be possible to space them farther apart than the limits imposed by the psychophysical two point threshold. Furthermore, the area selected should not be such that the positioning of the matrix would interfere with other necessary operations, such as might be involved in military occupations. For these reasons, the back was selected. The thoracic area of the back was finally chosen because of its relatively high sensitivity to tactual stimuli (19). The experimental area, then, was confined to the thoracic area of the back, which limited the maximum spatial separation to be investigated to 22 cm.

The intensitive variable, too, was greatly restricted. The vibrators used were built around 6-V, ac relay coils which would overheat with excessive current flow or when energized for long durations. The amplitude of the vibration (when damped by the skin) was therefore limited to 360 μ to minimize the possibility of overheating, and the temporal burst was limited to 200 msec.

There is a good possibility that the stimulus ranges used in this work, therefore, are not of sufficient magnitude to demonstrate the complete effects of the manipulation of certain variables. It is believed that the restrictions are so constraining that no sweeping generalizations can be made from the results. It is with knowledge of these limitations that the results of this experiment are discussed.

The analysis of the results does not reveal any consistent general relationships existing among the stimulus properties which might be interpreted as suggesting the vibrotactile analogues of Korte's laws. As the intensitive and spatial variables are manipulated no change in the duration of the optimal temporal interval can be noted, which would be the predicted outcome if such relationships, as Korte has stated, held for the vibratory mode of stimulation. If the intensitive variable is increased, at least within the limits imposed by this experiment, an increase in the total number of reports of movement does increase. However, it can be observed that the number of movement reports is not obviously related to the intensity of the stimulus, nor is there any change in the optimal temporal interval as the intensity is increased. As the amplitude was raised above the absolute threshold an asymptote for the frequency of movement reports was quickly reached. This suggested the possibility of plateaus caused by different stimulus relationships which will result in identical responses over relatively wide stimulus ranges. No significant change in the frequency of reported movement was noted between two of the three intensitive settings used.

The same result is obtained when the spatial separation between the transducers is varied. The number of reports of movement does not consistently change over the entire spatial range studied but quickly reaches an asymptote (at the shortest spatial separation used), maintains this position over a relatively wide range, and then rapidly declines regardless of the manipulation of the intensitive variable. In other words, the optimal intensitive and spatial variables are specifiable within a range of values, but once the limits of the range are exceeded for one variable, no compensatory manipulation of the other will result in the arousal of movement. One can speculate that the limits imposed upon the stimulus range were such that the functional relationships among the variables could not be adequately observed.

The results indicate that the most critical variable for vibrotactile apparent movement is the temporal interval between presentations of stimuli. It is with the manipulation of this variable that a specific stimulus setting will result in the maximum amount of reported movement, regardless of the stimulus dimensions of the other variables. It also suggests that the range of effective temporal intervals is quite restricted, in that relatively slight deviations from the optimal time will sharply reduce the reports of movement. The results show that the frequency of perceived movement is greatest when a temporal overlap of approximately 100 msec is fixed, and other variables simply are set within a rather wide range.

With these results it must be concluded that the vibrotactile analogues of Korte's laws cannot be stated within the limits of this study, but rather merely the specification of an optimal setting of the temporal interval. There is the further knowledge that the remaining variables, spatial separation and intensity, are uncritical; their values may be selected from wide ranges. It is highly probable that apparent movement can be aroused using stimulus dimensions outside the limits of this experiment. There are no experimental data in the vibratory field at the present time which might be used as reference either to confirm or deny such a possibility. Hulin's work, in the closely related area of tactual movement induced by static transducers, tends to corroborate the one specification of the experiment, in that he considers the temporal interval to be the critical variable for movement (9). Hulin, attempting to determine the optimal stimulus relationships for tactual apparent movement, found that the temporal interval between stimulations was the only stimulus property he could quantitatively specify. He found it necessary to say that the only quantification he could make from his experimental results was that a temporal overlap of 75 msec "is exceptionally favorable for the arousal of apparent tactual movement" (9, p. 320). He concluded from this that it was impossible for him to state the tactual analogues of Korte's laws. Hulin, however, like the present writer, confined himself to limited stimulus ranges.

The comparability of the present results with those of Hulin, where a minus temporal interval proved to be optimal, led to an extremely pertinent question. What temporal part of the stimulus burst is critical for the arousal of tactual movement? Is it aroused because of the sustained nature of the signal, vibratory in the present instance and static in the Hulin experiment, or is perceived movement aroused primarily with the initial impacts of the transducers? The question was raised because of the difference in the optimal temporal interval for apparent visual and tactual movement, a positive interval for the visual and a negative one for the tactual. If successiveness of vibratory bursts were the critical stimulus feature for tactual movement, then the optimal interval would be quite comparable to the interval optimal in vision, according to Wertheimer's finding. Stimulation would actually involve a temporal delay between the termination of the first stimulus and the beginning of the second.

The author undertook to answer the question in a brief supplementary study. Two of the subjects used in the major phase of the experiment, W and D, again acted as subjects in this work. Vibrotactile stimulation was delivered using three temporal intervals, including the optimal interval,* one spatial separation (4 cm) and all three of the previously described intensities. In addition to the usual stimulus burst duration of 200 msec, bursts of 20 msec were included, to determine whether or not a short "jab" would be as effective as the sustained vibratory burst or sustained static application. Both subjects reported that they could not detect a vibratory characteristic in the brief stimulation. The supplementary results are presented in Table VI.

This minor study indicates, as do four experiments in which static transducers were used (1, 9, 14, 25), that it is possible to arouse apparent tactual movement with a silent interval between stimulations. The evidence points to the possibility that the critical variable is the time between onsets of the two stimuli, rather than a silent or overlapping interval between the two. The frequency differences between the two types in reports of "no movement" are sufficiently large to afford some evidence that a burst of short duration is not as compulsory in producing apparent movement as the more sustained burst. The results further suggest that a sustained vibratory signal is an important stimulus property in the arousal of synthetic movement.

The study has shown, then, that apparent movement can be aroused using vibrotactile stimuli. Although it is impossible to state the vibrotactual analogues of Korte's laws, this experiment has defined a set of stimulus dimensions which can be incorporated into a pair of transducers and thus arouse movement consistently. The question which remains unanswered is, what will occur when these dimensions make up the stimulus characteristics of a vibratory matrix using more than two vibrators? This should be the next question answered. It is possible that the addition of other vibrators will tend to lower the threshold for perceived movement. As Bice (3) has pointed out, it was almost impossible for a subject to deny movement when six vibrators, equally spaced around the chest, were successively activated.

SUMMARY

The present study was designed to determine the stimulus conditions optimal for the arousal of apparent movement induced by vibrotactile stimulation. The implications which may be drawn from the results cannot be generalized since the stimulus ranges used were highly restricted, and the sensitivity of only one body area was investigated, that being the thoracic area of the back.

---

* The duration of the silent interval in this case is shown in Table VI. The 80-msec silent interval is the same as the optimal 100-msec overlap. The onset of the first stimulus precedes the onset of the second by 100 msec which was true in the earlier reported work. Compare Figure 1, p. 76.

TABLE VI

A TABULATION OF THE DATA COMPARING THE FREQUENCY OF MOVEMENT RESPONSES AROUSED WITH A SUSTAINED VIBRATORY BURST WITH A SHORT BURST

	Subjects			
Silent interval	W		D	
0 milliseconds	Vib.	Jab	Vib.	Jab
Amplitude				
120	21111	11111	12121	11111
240	11111	11211	32211	23122
360	11111	11112	11211	21111
80 milliseconds				
120	33322	21112	13323	32112
240	33333	33232	32233	23331
360	33332	33333	23332	22122
140 milliseconds				
120	22111	12121	33222	12212
240	12231	11212	12132	11111
360	23222	12232	22222	11112

Combined totals

Silent interval	Vib.			Jab		
0 milliseconds	1	2	3	1	2	3
Amplitude						
120	7	3	0	10	0	0
240	7	2	1	5	4	1
360	9	1	0	8	2	0
	23	6	1	23	6	1
80 milliseconds						
120	1	3	6	5	4	1
240	0	2	8	1	3	6
360	0	3	7	1	4	5
	1	8	21	7	11	12
140 milliseconds						
120	3	5	2	5	5	0
240	4	4	2	8	2	0
360	0	9	1	5	4	1
	7	18	5	18	11	1
Totals	31	32	27	48	28	14

The ultimate goal of this research, however, is the construction of a vibratory matrix to be positioned on the back of an operator, and to be used as a receiving mechanism for relatively simple, encoded information. This practical application is of primary importance, and for this reason no effort was made to increase the stimulus ranges beyond those dimensions which can be practically incorporated into the construction of such a communicatory device.

The vibratory transducers used were constructed around a 6-V, ac relay coil and could be driven at amplitudes ranging up to 360 μ. The intensities selected from this range for investigation were 120, 240, and 360 μ. Stimulus bursts were uniformly 200 msec in duration. In addition, nine temporal overlaps of stimulus bursts were used, each separated from the preceding by 20 msec. An exploratory study revealed that if the temporal interval was absolute succession, that is, no overlap or silent period, and the intensitive and spatial variables were the same as in the major part of the overall experiment, reports of movement were extremely rare, succession being reported. The third variable manipulated was the spatial separation of the transducers on the back. Three separations were used: 4, 12, and 24 cm.

Five subjects were each tested five times with every possible stimulus complex, amounting to 405 perceptual specifications for each subject. Every subject identified movement during the exploratory work without being told specifically to make such an observation.

The results indicate that the variable which can be most precisely specified quantitatively is the temporal interval between successive stimulus bursts. With a temporal overlap of 100 msec, the subjects reported some type of movement in 90 percent of the stimulus presentations, and 50 percent of the time good, full movement was reported. The vast majority of these reports of full movement occurred at the two highest intensities and the two smallest separations. The interval of optimal movement, in other words, was an overlapping of 100-msec of the two 200-msec vibratory bursts. The spatial and intensitive variables could not be quantitatively defined with such precision mainly because of limitations imposed by the practicalities of the situation, viz., the thoracic area of the back cannot be transcended, and excessive currents cannot be used to drive the transducers at higher amplitudes. The two shortest spatial separations were responded to as movement on approximately an equal number of occasions, the same result being true for the two greatest intensities. From this it was concluded that, within the limitations imposed upon this experiment, the temporal interval of 100 msec is prepotent, and the specification of the other variables is not critical. When the spatial separation is set between 4 and 12 cm, the intensitive variable between 240 and 360 μ, and the temporal interval at 100 msec of overlap, good synthetic movement will frequently be aroused.

A supplementary study, in which 200-msec stimulus bursts were compared with 20-msec bursts or "jabs," was carried out. The results revealed that even though it was possible to induce apparent motion with the short burst, using the same duration between the onsets of the two stimuli as with the more sustained bursts, the frequency of movement responses was greater using the long vibratory stimulation.

These optimal stimulus characteristics now will be incorporated into a vibratory matrix having at least a 3 by 3 vibrator design to determine the effectiveness of apparent movement in the transmission of simple directional and positional information.

ACKNOWLEDGMENT

The author wishes to express his most sincere appreciation to Professor Frank A. Geldard who contributed so much of his invaluable time to the progress of this study. For Professor Geldard's willingness to offer much needed advice and direction, the writer is further indebted. The author, too, is extremely grateful to Professors Raymond C. Bice and John F. Hahn whose tenacity, generosity, and knowledge were responsible for the solution of numerous electronic difficulties.

This research was completed under the auspices of the Office of Naval Research. It was conducted under Contracts N7-onr-372-02 and N-onr-474-06.

REFERENCES

1. Andrews, W.A., "Haptical Illusions of Movement," Amer. J. Psychol., Vol. 33 (1922), pp. 277-284.

2. Benussi, V., "Versuche zur Analyse taktil erweckter Scheinbewegunger," Arch. f. d. ges. Psychol., Vol. 36 (1917), pp. 84-104.

3. Bice, R.C., "Tactual Apparent Movement," Virginia J. Sci., Vol. 4 (1953), p. 279.

4. Burtt, H.E., "Auditory Illusions of Movement," J. Exp. Psychol. Vol. 2 (1917), pp. 63-75.

5. Burtt, H.E., "Tactual Illusions of Movement," J. Exp. Psychol., Vol. 2 (1917), pp. 371-385.

6. Gilbert, G.M., "Dynamic Psychophysics and the Phi Phenomenon," *Arch. Psychol., N.Y.*, No. 237 (1939).

7. Graham, C.H., "Perception of Movement," in S.S. Stevens, (ed), *Handbook of Experimental Psychology*. New York: John Wiley & Sons, Inc., 1951, pp. 895-901.

8. Howell, W.C. "Vibrotactile Communication and Training." Paper read at Virginia Academy of Science, Harrisonburg, Virginia, May 1955. (Abstract in *Virginia Journal of Science*, Vol. 6, Sept. 1955.)

9. Hulin, W.S., "An Experimental Study of Apparent Tactual Movement," *J. Exp. Psychol.*, Vol. 10 (1927), pp. 293-320.

10. Judd, C.H., "Uber Raumwahrnehmungen im Gebiete des Tastsinnes," *Phil. Stud.*, Vol. 12 (1896), pp. 425-428.

11. Katz, D., "The Vibratory Sense and Other Lectures, *Univ. of Maine Stud.*, 2nd ser., No. 14 (1930), pp. 90-114.

12. Korte, A., "Kinemntoskopische Untersuchungen," *Z.f. Psychol.* Vol. 72 (1915), pp. 193-296.

13. Mathiesen, A., "Apparent Movement in Auditory Perception," *Psychol. Monog.*, Vol. 41, No. 4 (1931), pp. 74-131.

14. Neuhaus, W., "Taktile Scheinbewegung," *Arch. f.d. ges. Psychol.*, Vol. 83 (1932), pp. 519-562.

15. Pattishall, E.G. "Operation Back-Track." Rep. 23, Proj. 16A, onr N7-372-02. University of Virginia, June 1, 1954.

16. Petzoldt, S., "Experimentelle Beitrage zur Lehre vom Vibrationssinn," *Z.f. Psychol.*, Vol. 108 (1928), pp. 155-194.

17. Sherrick, C.E. "Measurement of the Differential Sensitivity of the Human Skin to Mechanical Vibration." Unpublished master's thesis, University of Virginia, 1950.

18. Sherrick, C.E., "Variables Affecting Sensitivity of the Human Skin to Mechanical Vibration," *J. Exp. Psychol.*, Vol. 45 (1953), pp. 273-282.

19. Spector, P. "An Investigation of the Sensitivity of the Human Skin to Mechanical Vibration in Various Body Areas. Unpublished master's thesis, University of Virginia, 1952.

20. Spector, P. "Cutaneous Communications Systems Utilizing

Mechanical Vibration." Unpublished doctor's dissertation, University of Virginia, 1954.

21. Tachlenoff, L.G., "Sensibilitatsstudien an Nervenkranken I. Uber taktile Scheinbewegungswahrnehmungen," _Dtsch. Z.f. Nervenhk._, Vol. 121 (1931), pp. 180-212.

22. Von Frey, M., and S.R. Metzner, "Die Raumschwelle der Haut bei Successivereizung," _Z.f. Psychol._, Vol. 29 (1902), pp. 179-180.

23. Wedell, C.H., and S.B. Cummings, "Fatigue of the Vibratory Sense," _J. Exp. Psychol._, Vol. 22 (1938), pp. 429-438.

24. Wertheimer, M., "Experimentelle Studien uber das Sehen von Bewegung," _Z.f. Psychol._, Vol. 61 (1912), pp. 161-265.

25. Whitchurch, Anna K., "The Illusory Perception of Movement on the Skin," _Amer. J. Psychol._, Vol. 32 (1921), pp. 472-489.