Effects of Delayed Speech Feedback

BERNARD S. LEE
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
(Received August 9, 1950)

Delayed speech feedback may act as a governor of the average speed of speech operating on the phoneme unit. Certain aurally monitored manual tasks resemble the speech mechanism. The trigger-like nature of aurally monitored voice on the more rapid articulation element is described. A speech mechanism model is portrayed which satisfies phenomena of natural speech, artificial stutter, and motor aphasias.

INTRODUCTION

WHEN a person's own voice is returned to his ears by technique of the multiheaded magnetic tape recorder and earphones a startling disturbance of his speech may be noted. A delay of about $\frac{1}{4}$ sec. may cause the subject to speak very slowly but if he maintains normal speed, stuttering characterized by repetition of syllables or fricatives may occur. The level of the returned speech must be somewhat above that heard through bone conduction in order to be effective. This phenomenon seemed especially worthy of attention because of the important role played by the aural monitor of speech and because of the unusual opportunity afforded to experimentally hybridize a neural and electronic network. A similar psychological upset of speech in certain auditoriums equipped with public address systems has been noted in the past. The effect referred to herein is more pronounced because the earphones stop the normal monitoring system and permit any level of sound to be fed back to the subject. The terms delayed side-tone, dominating echo, and delayed feedback are used synonymously. Since the first communication[1] other characteristics of this phenomenon have been noted.

AS A GOVERNOR OF SPEECH SPEED

Delayed feedback may act as a governor of the overall speed of speech. This was first noted as a tendency to slow down under the influence of the delayed speech feedback. Subjects were instructed to read a passage of moderately difficult text at a comfortable speed without error. The average time required to read this selected passage under normal conditions was 31 sec. They were then instructed to repeat the same passage while their speech was returned with the delay of 0.04, 0.14, and 0.28 sec.; and the total time to read the passage was recorded. Cooperation with the slowing down effect was urged on the subjects. This is accomplished automatically by some subjects while others may require practice to achieve the cadence demanded by the delayed feedback. Failure to achieve the proper cadence results in artificial stuttering or erratic speed, defeating the instructions given to the subject; namely, to maintain accurate speech. If delayed speech feedback acts as a governor on the speed of speech, then the question is, "On what unit does the delay operate?" A count of the phonemes in the passage numbered 372 while the number of spaces between words was 65. Arbitrarily assigning the length of time for a space between words equal to that of a phoneme the total units in the passage was 437. If the person were truly machine-like in performance (and some are) his total time T would be

$$T=n(d+t),$$

where n is the number of units; that is, phonemes plus spaces; t is the unit time for unencumbered speech; and d is the delay artificially inserted in the monitoring loop. Figure 1 presents the data for this experiment. In another test subjects were asked to repeat as rapidly as possible without error a single syllable 40 times. (See Fig. 2.) Here the error, of which more will be said later, was of a type which produced more syllables than the subject wished. It must be pointed out again that cooperation on the part of the subject is required to achieve these results since partial evasion of the effect of the feedback can be made by some individuals by concentrating on lip movements for monitoring purposes.

Black[2] showed that speech is slower in live, large rooms than in dead or small rooms. Examination of his published data reveals that a difference in wall distance of 7.9 ft.* would provide the acoustic delay to account for the time difference of 0.23 sec. on a 16-phoneme message.

AURALLY MONITORED MANUAL TASKS

A very skillful machine gun operator can squeeze off one or two rounds from an automatic weapon which shoots at 10 per sec., but an untrained person may shoot one or two more than desired. A delicate touch and an acute ability to aurally monitor the required number of shots is demanded for this feat. A study of the errors in speech induced by the delayed feedback reveals that the speech mechanism is similar in operation to the machine gun. For example, the word aluminum may become aluminum-num; "lum-lum" becomes "lum-lum-lum," etc. These stuttering errors occur when the subject attempts to speak at his normal rate and the delay is of sufficient duration to permit an extra syllable or fricative. Persons endowed with a sense of rhythm are

[1] Bernard Lee, J. Acous. Soc. Am. 22, 639 (1950).

[2] John W. Black, J. Acous. Soc. Am. 22, 174 (1950).
* $(0.23\times1100)/(16\times2)=7.9$.

incapable of rhythmic tapping when listening to their output over a delayed circuit. Skilled wireless telegraph operators perform erratically under the same conditions; for example, *C* may result in —··—·· rather than —·—·.

RELATION OF ARTICULATION AND VOICING

Speech may be divided into the components of articulation and voicing. The elements of articulation are commonly accomplished at a rate[3] of 14 per sec. while distinct utterances of the vocal chords alone can be performed at only one-quarter this speed. Musical instruments analogous to the speech mechanism are the violin and trumpet where the bow and vibrating lips provide voice to the variations of the fingerboard and valves, respectively. In speech the vocal chords produce the basic tone which is modified by rapid variations of the vocal tract. The duration or number of the high speed elements is controlled by aurally monitoring the slower of the two components. This is most convenient and reasonable because the aural monitor provides the same information to the speaker that he wishes to impart to a listener. I do not mean to say that monitoring of speech is exclusively aural; on the contrary, kinesthetic and tactile monitoring of the articulation elements is a necessary part of the speech model described in the following paragraph.

MODEL OF THE SPEECH MECHANISM

Figure 3 is a conveniently portrayed model for a speech mechanism which will accept the following phenomena: (1) phenomena of delayed speech feedback; (2) normal speech; (3) motor aphasias; (4) stuttering, both natural and artifically induced.

Properties of the model are: (1) Consistent with neuro-anatomy, the paths must be unidirectional closed loops. (2) There is a common junction for all of these loops. (3) Volitional or reflex switching is provided at this junction. (4) The length of each loop is roughly proportional to the time required for its particular function. (5) The hearing system is in series with or is inductively coupled to the voice loop for the aural monitoring function.

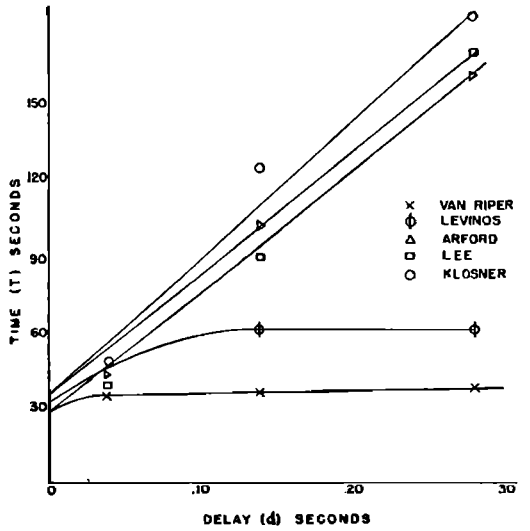

FIG. 1. Time required to read a text of 437 units (372 phonemes plus 65 spaces).

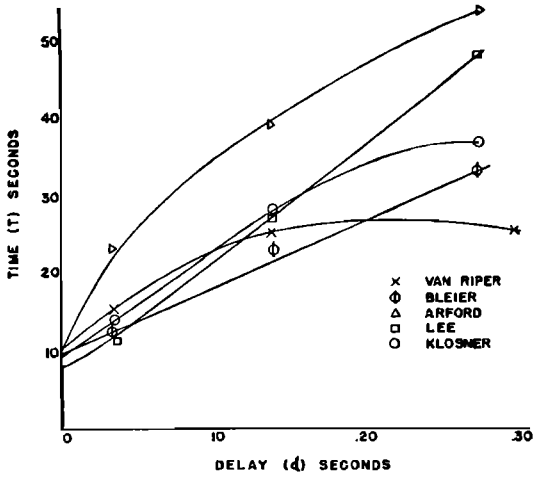

FIG. 2. Time reduced to recite 40 syllables (120 phonemes).

A great variety of portrayals having these properties could be shown. This one is selected because it is simple to draw and easy to remember. Instead of smooth circles, actual anatomy might very well use intricate multibranched courses. Although only two lines are shown at most in the sketch, it is more likely that a dozen parallel fibers carry the signals necessary to produce articulation by means of the several parts of the vocal tract, operation of the breath for volume, tension of the vocal chords for pitch and inflection, etc. It is the inner set of loops labeled articulating and voice which is the speech machine proper and to which delayed speech feed-back phenomena offer most supporting information. The other two loops, word and thought, are admittedly based on a casual study of people's speech habits.

By suitable switching at the common junction the following effects may be accomplished: (1) abstract thinking—uses outer loop only; (2) thinking in words—uses thought and word loop; (3) articulating without voice—all but voice loop; (4) speech—all loops.

A property of the model is that any element down to syllable may be repeated if the hearing monitor is not satisfied. If the monitor *is* satisfied, however, the signal proceeds into the next larger loop where it forms a part of the next larger component. Monitoring of the smallest loop (articulating) is by kinesthetic or tactile means, the next (voice) by aural means; and both are at reflex level. Monitoring of the remaining loops (words and thoughts), the decision to correct, repeat, or let stand, is volitional. A word may be repeated for correction, an entire thought rephrased for clarity, corrected for ambiguity, grammar,

[3] Dale J. Lundeen, J. Speech and Hearing Disorders (March, 1950).

etc. The word loop also provides a device for stalling or maintaining the speech flow as a defense against interruption. This gives a nervous quality to speech which all speakers resort to more or less but which when used excessively may be called a Class I stutter. The articulating or phoneme loops might be the 40 neural paths well-learned in infancy for English speech. The function of the signals carried around the voice loop is to tighten the vocal chords for pitch and inflection while regulating the breathing and thus provide the sound or carrier signal which is modulated by the vocal tract. Fricatives seem to be monitored by a combination of aural and kinesthetic feedback. The sibilants especially carry a high amount of energy and are subject to both artificial and natural stuttering.

ARTIFICIAL AND NATURAL STUTTER

As indicated above, the artificially induced stutter which occurs when one attempts to maintain normal rapid speech while monitoring over a delayed feedback results in undesired repetition of syllables or fricatives.

Speech therapists recognize two types of stutterers; the first, Class I or clonic,[4] involuntarily repeats syllables or words to an excessive degree. Practically all persons fall into this class more or less. The second type of stutterer, Class II or tonic, repeats a single phoneme or may stop halfway through a single speech element. It is this second type of stutter which is the more serious and is of a puzzling transitory nature.[5] Referring to the model it is the unsatisfied monitor of the voice loop under the conditions of artificially induced stutter which causes the voice loop to continue for an extra turn or two until the delayed feedback returns and releases the actuating trigger.

Since no voiced sound is involved in the individually articulated elements, the technique of the magnetic tape recorder cannot insert itself into this smallest loop of the speech machine for experimental purposes except as a speed governor. Perhaps the anesthetist or the neuroanatomist could experiment with these loops.

[4] Emil Froschels, *Speech Therapy* (Expression Company, Boston, Mass., 1933).

[5] Oliver Bloodstein, J. Speech and Hearing Disorders (June, 1950).

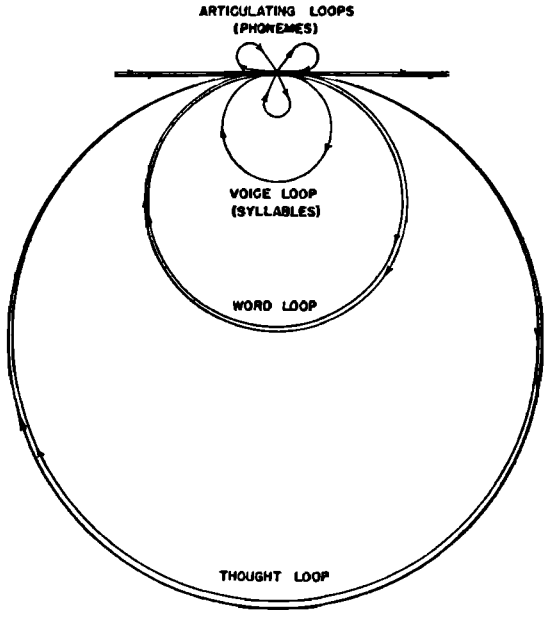

FIG. 3. Speech model.

APHASIAS

It is believed that most of the aphasias of the motor type described by Goldstein[6] can be compared with one or more defects in the speech model. Impairment of abstract attitude, finding of words, finding of small words, the ability to speak at all, and almost every conceivable combination of these defects may be typed as defects of the four-loop system, either with regard to switching, monitoring, or absence of any or all of the loops.

ACKNOWLEDGMENT

The writer wishes to acknowledge the many conferences and the valuable suggestions made by Mr. Neil G. Smith, Speech Instructor of Fort Monmouth Signal School.

[6] Kurt Goldstein, *Language and Language Disturbances* (Grune and Stratton, Inc., New York, 1948).