A precision electronic timer for the measurement of acoustic signals

HARVEL N. DAWIRS and VIRGINIA G. WALKER
Florida State University, Tallahassee, Florida

An electronic speech timer that uses only relatively inexpensive and readily available electronic components and common circuitry is described. This device rapidly and accurately measures the durational characteristics of speech and the durations of other acoustic signals over extended lengths of time. Its application to the measurement of total speaking time, articulation time, and phonation time are outlined. Accuracy, reliability, and validity are discussed and compared with other, more time-consuming and cumbersome methods of durational measurement.

In speech research, it is frequently necessary to determine the total amounts of time spent in phonation, articulation, and pause. Sophisticated and expensive instruments are available for making such measurements (Donahoo, Nettleton, & Bradshaw, 1973; Hargreaves & Starkweather, 1959; Jensen, Ruder, & Harrington, 1972; Starkweather, 1960; Steer & Hanley, 1957). It is possible, however, to use rather simple and relatively inexpensive electronic circuitry to make precise measurements.

Acoustically activated switches have been available for many years and are quite common. They are used to turn on such equipment as radio transmitters, tape recorders, and amplifiers when there is a voice signal. Intrusion detectors use such switches to indicate the presence of unusual acoustic noise and to turn on an alarm. Such switches are often referred to as "voice-operated relays" (Hanley & Peters, 1971).

Electronic speech timers used in research differ from ordinary voice-operated relays in that they respond more quickly to the acoustic signal, are more sensitive in order to provide the precision required, and have provisions for measuring the duration of such characteristics as articulation, phonation, and pause times. This paper is concerned with the use of inexpensive and readily available electronic components in reasonably common circuitry to make very accurate measurements.

BASIC CONCEPTS

Basically, all an acoustic switch (voice-operated relay) does is to convert an acoustic wave into a dc voltage, which is then used for control purposes. (In the case of speech timers used for research in speech science, the

dc voltage is used to operate timing circuits.) To do this, the acoustic wave is converted into a dc voltage by converting it into an electric signal that is then rectified. The electric signal comes from a microphone, usually by way of amplifiers and/or tape recorders. Maximum accuracy and reliability are obtained when the signal to be rectified is large. Therefore, a considerable amplification of the signal may be necessary to obtain good results. In fact, it is desirable to overamplify the signal (by overdriving the amplifier) to convert the signal to a series of "square waves," as illustrated in Figures 1a and 1b. This process is generally referred to as "clipping." Figure 1a represents a short interval of speech, and Figure 1b represents the corresponding square waves that result from overamplifying, or clipping, the speech waves.

Actually, the "square" waves obtained from the speech signal are always slightly trapezoidal, as indicated in Figure 1b, and slightly displaced from the speech

Figure 1. Conversion of speech wave to dc voltage.

Harvel N. Dawirs is with the Department of Speech Communication and Virginia G. Walker is with the Department of Audiology and Speech Pathology, both at Florida State University, Tallahassee, Florida 32306.

signal due to delays and time constants present in real circuitry. Such delays add to the overall response time of the switch and reduce its accuracy. They arise from the rise time of the components used and the time constants resulting from capacitance and inductance in the circuit. Consequently, to enhance accuracy, fast-acting components are used; stray capacitance and inductance are held to a minimum, and coupling capacitors and transformers are, in general, avoided if at all possible.

The speech signal may be clipped in a number of different ways, such as by overdriving an amplifier. In the circuit to be described, transistors are driven from cutoff to saturation.

Full-wave rectification of the clipped speech signal results in a dc voltage for the duration of the speech if the clipped wave is truly a square wave. In practice, however, there are always small gaps between the individual waves of the rectified speech signal, due to the trapezoidal shapes of the clipped waves, as depicted in Figure 1c. These gaps may be removed by means of a low-pass filter or an averaging circuit, both of which introduce unavoidable delay into the circuit.

An averaging circuit, such as the one shown in Figure 2, is better than a filter because the delay can be controlled and calibrated independently of frequency. The first wave (Figure 1c) of the rectified speech signal charges up the timing capacitance C in Figure 2. Care must be taken that the charging time-constant is small to ensure that there is negligible delay in charging the capacitance up to the full voltage of the rectified wave. This is important because the speed with which the timing capacitance charges up is one of the factors that determine how fast the overall switch will respond to the beginning of the speech signal. The initial squaring of the speech signal to minimize the trapezoidal shape, and hence the rise time, of the wave also minimizes the switching time by ensuring that a good, quick charging signal is supplied to the capacitance charging circuit in the first place.

The positive voltage across the timing capacitance C , applied to the base of the switching transistor through resistance R , saturates the transistor and causes the output voltage (dc control voltage in Figure 2) to drop to zero at the beginning of the "gate," as shown in Figure 1e).

Figure 2. Delay circuit and switching transistor.

At the end of the first wave, the diode in Figure 2 becomes back biased, and the capacitance discharges relatively slowly through the base of the transistor and the resistor R until either the capacitance discharges to the point that either the transistor is cut off or another wave charges up the capacitance, as indicated in Figure 1d. As long as the voltage across the capacitance remains above the saturation level of the transistor (see Figure 1d), the transistor remains saturated and its output (gate) voltage remains at zero, as shown in Figure 1e.

At the end of the speech wave, the rectified wave drops below the saturation level and quickly passes into the cutoff region of the switching transistor, as indicated in Figure 1d, at which time the output voltage of the switching transistor rises back up to the full power supply voltage V_{cc} and ends the voltage "gate," as shown in Figure 1e. Thus, the output (or dc control voltage) at the collector of the switching transistor always drops to zero whenever there is a speech signal, resulting in the voltage "gate" of Figure 1e.

Although the gate voltage can be made to start very soon after the speech wave starts, there is always some delay at the end of the speech wave, due to the unavoidable delay of the averaging circuit. This delay can be controlled by choosing the values of the capacitance and the base resistor of the discharge circuit, and may be made variable either by changing the value of the capacitor C or by using a potentiometer as the base resistance R . This time-constant must be long enough to fill in the gaps between the individual waves of the speech signal without causing undue delay at the end of the speech signal. If the time constant is made too short, the capacitance discharges too fast and dips below the saturation level of the transistor between the individual waves, thus leaving gaps in the dc control voltage and resulting in "holes" in the voltage gate.

The delay should also be long enough to eliminate the effects of such things as plosive gaps, which may be instrumentally confused with natural speech pauses such as occur when one breathes. The proper amount of time should not be so long as to increase the delay time to more than necessary to fill in the gaps between the individual waves. Even with relatively long delays, precision can be maintained if the delay circuit is calibrated or measured accurately, so that the amount of delay occurring with each pause is known, and the number of pauses is then counted. The number of pauses multiplied by the known delay introduced by the delay circuit for each pause will determine the total delay for any passage of speech. This value subtracted from the total articulation time will correct for all of the delay, resulting in very accurate values for the total pauses or articulation time.

AN OPERATING CIRCUIT

Figure 3 is a circuit diagram of a complete transistorized speech timer that has been operating reliably

Figure 3. Circuit diagram of a complete transistorized acoustic switch.

and accurately for a number of years in the Florida State University Speech Science Laboratory. Within the switch circuit itself, there are no coupling capacitors or transformers which can introduce unnecessary delays. The only coupling capacitors are C_1 and C_2 , which introduce the speech signal into the switch from the output amplifiers, and C_3 and C_4 , which are associated with the timing gate transistor Q_9 ; all of these capacitors are outside the acoustic switch proper.

Although the input capacitors C_1 and C_2 are not within the switch circuit itself, their values must be selected with some care to ensure that they work properly in conjunction with the associated resistances and transistors at the input of the switch. Otherwise, they either limit the operating bandwidth of the switch or charge up and then introduce excessive delays while they discharge. The latter is a distinct possibility when the inputs are as highly overamplified as they are here to achieve the clipping and squaring of the input waves to the switch. The diode across C_2 limits the voltage across that capacitor and maintains the proper polarity.

Capacitors $C_a, C_b \dots C_z$ are the timing capacitors of the delay circuit; they introduce the controlled delay necessary to fill in the gaps between the individual waves of the speech signal. In order to adjust the time delay for best operation, any one of a number of capacitors with different values may be selected by means of a switch. The delay introduced is about 2 msec for each microF of capacitors.

A diode bridge, the usual method of obtaining full-wave rectification, cannot be used in the absence of transformers or capacitors for coupling the balanced and unbalanced circuitry involved. In the switch described here, effective full-wave rectification is achieved with the following circuitry, which avoids coupling capacitors and transformers. A description of the operation of the switch follows:

Transistors Q_3 and Q_4 , along with their associated circuitry, are referred to in this diagram as the "negative channel," since they operate on the negative half cycles of the speech waves. The effect of this negative channel is to cause the voltage at the collector of Q_4 (which is also the collector of Q_5) to drop to near zero whenever the original speech wave is negative.

Transistors Q_1 , Q_2 , and Q_5 of Figure 3, along with their associated circuitry, are referred to as constituting the "positive channel," since they operate on the positive half cycles of the speech wave. The effect of the positive channel is to saturate Q_5 , dropping its collector voltage, along with that of Q_4 , to zero whenever the speech wave is positive.

The effect of both channels, positive and negative together, then, is to cause the collectors of transistors Q_4 and Q_5 (which are connected) to drop to zero whenever there is any speech wave; the positive channel causes the collector of Q_5 to drop on positive half waves, and the negative channel causes the collector of Q_4 to drop on negative half waves. Thus, transistors Q_4

and Q5, with the collectors connected, combine the effects of the two channels, which results in a combined wave that is positive when there is no incoming speech signal but is zero whenever there is a signal, except during the gaps between the individual waves. Note that this is just opposite to the wave in Figure 1c, in which the full-wave rectified signal is zero except when there is an incoming signal. Transistor Q6 of the actual switch (Figure 3) inverts the output of transistors Q4 and Q5 to obtain the same "full-wave" rectified wave of Figure 1c that would be obtained with a diode bridge. Transistor Q6 also provides a very fast charging circuit for quickly charging the timing capacitor in the delay circuits; to obtain the desired delay, one of the capacitors Ca, Cb ... Cz is selected. The action of the delay circuit of the actual acoustic switch (Figure 3) is then the same as that of Figure 2. The overall result of the complete switch, then, is the control voltage shown in Figure 1e. This is nearly zero whenever there is a speech signal at the input to the switch, but it is equal to the full power-supply voltage of V_{cc} at all other times, except during the small, controllable delay at the termination of each speech signal.

This control voltage may be used in any way desired. In Figure 3, for example, it is applied to the base of transistor Q8, which drives a light emitting diode. This diode is turned on when the control voltage of Figure 1e is zero, that is, whenever there is a speech signal. Thus, it serves as a monitor for the operation of the switch. It is also used as an indicator in setting the bias controls of transistors Q1 and Q3 at the switch input. If either transistor is underbiased, the light will stay on at all times.

To operate other circuits, the control voltage is also made available at the collector of Q8. This voltage can be used to drive a digital counter that will count the number of pauses between the sections of speech. This number can then be multiplied by the amount that the control voltage is delayed at the end of each speech signal to determine the total error in the measurement of a number of sections of speech in any complete passage. This error can be subtracted from the total articulation time to correct for the delay errors introduced by the switch.

The control voltage (Figure 1e) at the collector of Q7 is also used to operate Q9 (Figure 3) as a transistor gate. We use this gate to turn on a very accurate 1000-Hz signal that is fed into the "clock input" terminal shown (only when speech signals are fed into the switch input). The resulting gated signal at the "gated clock output" terminal is fed into a digital counter that counts the waves of the gated clock signal. Note that the counter counts the 1000-Hz clock signal only when there is a speech signal into the switch input. The result is that the speech signal activates the electronic switch, which in turn allows the 1000-Hz clock signal to pass to the gated output. The clock output may be read directly in milliseconds. The accuracy of the measurement depends

only on the precision of the switch and the 1000-Hz clock oscillator that supplies the timing signals. The circuit can be connected to measure both pause and speech times. The gate voltage may be used to control a computer that can be used to process the results.

CURRENT USE

This speech timer has been used to measure articulation and phonation times as follows: Articulation time is measured by using the timer to measure the total time a speech signal is presented from an ordinary microphone. Phonation time is measured by supplying a signal to the switch from a throat microphone specially designed to pick up only phonation. Figure 4 shows the analysis of a hypothetical passage of speech in which (1) total time represents 100% of the passage, (2) articulation time represents 80% of the passage, (3) phonation time represents 70% of the passage, (4) voiceless speech time represents 10%, and (5) pause time represents 20%.

Phonation time, or the amount of time spent in producing voiced speech signals, is determined by means of the special throat microphone and speech timer, as indicated above. Articulation time, or the presence of any acoustic signal associated with the passage, is measured by using an ordinary acoustic microphone in conjunction with the speech timer. The signals from either microphone, of course, may be stored on tape before inputting to a speech timer. To accurately determine the total speaking time, the total passage is first recorded on audio tape and separated from the rest of the reel by means of leader tape. A 1000-Hz tone is then recorded over the original passage or on a separate

Figure 4. Durations in speech.

channel. The speech timer measures the time of the 1000-Hz tone and, hence, of the passage, since both occupy the same amount of time.

Voiceless articulation time, which includes any acoustic signal that is not accompanied by the presence of vocal fold vibration, can be derived by subtracting the phonation time from the articulation time. Pause time, or the amount of time spent in natural interruptions or hesitations that might have semantic significance, can be derived by subtracting articulation time from the total speaking time. The effect of the implosive phase of the voiceless stop plosive, such as would occur between the /n/ and /t/ in the utterance "on top," can be eliminated by incorporating an appropriate amount of variable delay.

VALIDITY

To determine its validity, the system was used to make measurements on a locally recorded standard test tape. The results were compared with measurements made on the same tape with two traditional durational measurement procedures, sound spectrography and graphic level recordings.

The standard test tape was constructed as follows: a very accurate 1000-Hz tone from a precision oscillator was recorded on tape, divided into eight random segments, and butt-spliced with leader tape to provide a series of rapid on/off signals. This tape was then played directly to a frequency counter; tape speed was monitored for variations by comparing the recorded signal with a signal from the precision oscillator by means of Lissajous figures. Each portion of the signal was measured in milliseconds. The resulting times, accurate to within 1 msec, were used as a common reference for comparison measurements by means of (1) the electronic speech timer, (2) a Brüel & Kjaer Type 2305 graphic level recorder, and (3) a Kay Elemetrics Sonagraph Model 6061 A sound spectrograph.

Comparative measures indicated that the electronic speech timer and the sound spectrograph are comparable in accuracy. The error rates of the sound spectrograph (.21%) and the electronic speech timer (.13%) are relatively low. Both devices are consistently more accurate than the graphic level recorder (3.01% error rate), the accuracy of which is limited by the relatively slow writing speed of the pen mechanism. The value of the electronic speech timer is immediately apparent when, in addition to its superior accuracy, the relative ease and convenience with which the measurements can be made are taken into account.

RELIABILITY

To determine the reliability of the electronic speech timer, two subjects were tape recorded, and repeated measures of their articulation times were made. To

verify minimal fluctuations, variations in the tape recorder transport speed were again monitored by Lissajous figure against a highly accurate frequency standard.

Ten repeated measures of the articulation time for the male subject yielded a mean of 14.192 sec, with a range of .051 and a standard deviation of .015. The number of pauses in the male subject's reading sample was a mean of 72.6, with a range of 7 and a standard deviation of 2.95. The mean for the 10 repeated measures of the female subject's sample was 11.778 sec, with a range of .054 and a standard deviation of .017. Her number of pauses was a mean of 78, with a range of 5 and a standard deviation of 2.11.

Articulation time was selected for the determination of reliability because, of all the previously mentioned characteristics of the speech signal, it is more likely to be vulnerable to spurious readings in that it comprises greater shifts in intensity, in frequency and in the on/off characteristics of voicing than do measures of phonation time. Again, variations of less than 60 msec on repeated measures could be considered to be relatively insignificant in view of the long samples of speech that can be processed rapidly with this automatic method.

APPLICATIONS

To date, the electronic speech timer has been used to gather normative data for a wide variety of subjects. The speech duration characteristics of both black and white male and female college students have been quantified and compared during conversational speaking and reading tasks (Payne, 1981; Walker, 1979). These data have been compared with measures of aging black and white males and females to determine the stability of these characteristics throughout the life cycle (Staley, 1980; Walker, Hardiman, Hedrick, & Holbrook, 1981).

In addition to providing normative data, the electronic speech timer has been used to measure clinical changes in speech durational characteristics as a result of speech and voice pathologies such as stuttering, functional and organic voice disorders, and the speech production characteristics of hearing-impaired individuals (Holbrook, Dawirs, & Walker, Note 1; Holbrook, Dawirs, Walker, & Mosley, Note 2). Such measures not only provide comparative data for disordered, as opposed to normal, speech production but also provide for measures of clinical progress during the course of management. Pilot data have also been gathered to determine the value of the electronic speech timer as a measure of improved speaking durations as a function of theatrical training.

The system can also be applied to the measurement of nonspeech acoustic signals in cases in which rapid, long-term measurements are desirable. For instance, the amplification of muscle noise through the acoustic amplification system would provide normative data not

only on muscle physiology, but also on changes in such activity as a result of disease or injury. Studies of changes in heart rate as a function of physiological or emotional stress can also be studied. Long-term measures of industrial noise duration can also be provided.

REFERENCE NOTES

1. Holbrook, A., Dawirs, H. N., & Walker, V. G. *Clinical and research applications of an analog convertor for the study of duration and intensity in speech*. Paper presented at the meeting of the American Speech and Hearing Association, Houston, November 1976.

2. Holbrook, A., Dawirs, H. N., Walker, V. G., & Mosley, J. *Clinical applications of a duration encoder for the study of duration and intensity in speech*. Paper presented at the meeting of the Florida Language, Speech and Hearing Association, Ft. Walton Beach, Florida, April 1977.

REFERENCES

- DONAHOO, K., NETTLETON, N., & BRADSHAW, J. A system for accurately measuring articulation durations which is not reset by any included silent periods. *Behavior Research Methods & Instrumentation*, 1973, 5, 407-409.
- HANLEY, T. C., & PETERS, R. The speech and hearing laboratory. In L. E. Travis (Ed.), *Handbook of speech pathology and audiology*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

- HARGREAVES, W. A., & STARKWEATHER, J. Collection of temporal data with the duration tabulator. *Journal of the Experimental Analysis of Behavior*, 1959, 2, 179-183.
- JENSEN, P. J., RUDER, K. F., & HARRINGTON, W. D. Pause adjustment mechanism and measurement system (PAMMS). *Behavior Research Methods & Instrumentation*, 1972, 4, 304-312.
- PAYNE, J. A. *A study of speaking and reading durations of young black adults*. Unpublished doctoral dissertation, Florida State University, 1981.
- STALEY, A. A. *Speech durations of older white females in speaking and reading*. Unpublished master's thesis, Florida State University, 1980.
- STARKWEATHER, J. A. A speech rate meter for vocal behavior analysis. *Journal of Experimental Analysis of Behavior*, 1960, 3, 111-114.
- STEER, M. D., & HANLEY, T. D. Instruments of diagnostic, therapy, and research. In L. E. Travis (Ed.), *Handbook of speech pathology*. New York: Appleton-Century-Crofts, 1957.
- WALKER, V. G. *Speech durations of young adults during speaking and reading*. Unpublished doctoral dissertation, Florida State University, 1979.
- WALKER, V. G., HARDIMAN, C. J., HEDRICK, D. L., & HOLBROOK, A. Speech and language characteristics of an aging population. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 6). New York: Academic Press, 1981.

(Manuscript received December 7, 1982;
revision accepted for publication February 22, 1983.)