Editing silent hesitation from speech recordings

A. I. HENDERSON and D. G. SMITH

University of Southampton, Highfield, Southampton, SO9 5HH, England

A simple method of increasing speech rate by shortening or removing silent hesitation from speech recordings is described. Two tape recorders and a voice key are used. The transmitting tape recorder is modified to have two playback heads in line. The onset and offset of speech signals fed from the first head, via a voice key, control the tape transport of the receiving tape recorder. The speech signals from the second playback head are fed directly to the line input of the receiving tape recorder. Time delays can be arranged so that the tape transport of the receiving machine is only in motion for the duration of the speech signals. The logic circuit of a suitable voice key with independently variable onset and offset delay times is also described.

This note describes a way of re-recording speech to remove all the silent breaks in the original recording that exceed some selected duration. This technique is an addition to those described by Foulke (1970) for controlling the rate of recorded speech. In practice, the maximum compression that may be achieved is limited by the quality of the original recording and the voice characteristics of the speaker. Using carefully spoken and well-recorded experimental tapes, all breaks longer than 150 msec have been removed. For recorded conversations, breaks longer than 250 msec have been removed. Breaks of this duration are in, or closely approach, the range of those occurring within words, e.g., to stop consonants.

THE TECHNIQUE

Two tape recorders are used, a transmitter and a receiver. The transmitter is a modified half-track stereophonic tape recorder with upstream and downstream playback heads set in line so as to monitor the same track. Figure 1 shows a timing diagram. The output from the upstream playback head (P.B.1.) is fed via a high-pass filter and a variable gain amplifier to a voice key. The voice key, which should have variable onset and offset timed delays, acts as the editor. Its output operates the remote-control switch of the receiving recorder, which is set in the record mode. Thus, the output from the upstream head on the transmitting tape recorder controls the operation of the tape transport of the receiving tape recorder, switching it on at the onset of speech signal and off at the end of a speech signal.

The output of the downstream playback head (P.B.2.) on the transmitter is fed directly to the line input of the receiving tape recorder. The tape transport of the receiver is in operation for the length of the speech signal plus the duration of the offset

time selected on the editor and the remote-control solenoid delay (see Fig. 1). The offset time controls the duration of the maximum break in the original recording, which is re-recorded by the receiving tape recorder. In order to avoid clipping the ends off words, the offset time must not be less than the tape traveling time between the two playback heads on the transmitter. This ensures that the tape transport on the receiver is still

operating and will record the end of the signal sent from the downstream playback head.

The operating time of the remote-control solenoid on the receiving tape recorder plus the time chosen for the onset delay on the editor must be slightly less than the traveling time between the playback heads on the transmitter. This ensures that the tape transport on the receiving tape recorder is up to full running speed in time to record the onset of the speech signals.

The onset delay on the editor may be thought of as a throwaway time. All signals of shorter duration than the selected time fail to switch the receiving tape recorder. Thus, it is protected from being switched by short bursts of noise, e.g., electrical disturbance. Suitable times for the onset delay would be in the range of 10-20 msec. As with all voice keys, which are threshold operating devices, a high-quality recording with a good signal-to-noise ratio and a minimum of extraneous noises is essential if the device is to trigger to all and only speech signals.

A simple system of this sort has the advantage of using standard tape recorders, one of which requires only

Fig. 1. Timing diagram showing the functions of the editor. The output from the upstream head on the transmitting recorder (P.B.1.) is fed to the editor. t_1 is the throwaway time; t_2 is the offset delay. The onset and offset of the editor switches the tape transport of the receiving recorder. The receiver is recording only during the ON time. X represents the lag due to the operating time of the remote control on the receiving recorder. By adjusting the offset delay more or less silent time on the original tape may be re-recorded. Delayed speech signals are fed directly to the line input of the receiving recorder from the downstream head (P.B.2.). The delay is achieved by adjusting the tape traveling distance between P.B.1. and P.B.2. on the transmitting recorder.

Fig. 2. Circuit of the editor: P.S., pulse shaper, INV., gate inverters, O.S., one-shot multivibrator; R.D., relay driver; the NAND operation is performed by a gate inverter. The terminals are: t, trigger input; q, output, g, gate.

minor modification. The transmitting tape recorders should be chosen for easy access to the heads. The recorder used was an old Revox, Model 36(F). Most Revox and Ferrograph stereorecorders of the Brennell STB1 or STB2 would be suitable. The manufacturers are: Revox Recorders, Willi Studer, Regensdorf, Zurich, Switzerland; The Ferrograph Co. Ltd., 84 Blackfriars Road, London, S.E.1., U.K.; Brennell Engineering Co. Ltd., 231-232, Liverpool Road, London, N.1., U.K. The receiving tape recorder should have a remote-control feature which has a sharp braking action and produces satisfactory recordings at tape speeds of 4.75 cm/sec (1.7/8 in./sec) and 2.375 cm/sec (15/16 in./sec). Used at these speeds, the braking action of the remote control shows a minimum of tape slip. An Uher 4000 Report-L was used as the receiver. Almost any reasonable quality recorder with a remote-control facility would serve. Information about Uher recorders may be had from Bosch Ltd., Rhodesway, Radlett Road, Watford, Herts., U.K.

A modification similar to that described by Huggins (1967) enables the tape traveling distance and the time delay between the two playback heads to be varied. This modification allows for the lag involved in switching on the receiver (occasioned by the remote-control operating time plus the onset delay selected on the editor) to be matched by the traveling time between the two playback heads on the transmitter.

The voice key used (Fig. 2) was made up from M.E.L. Series 10 circuit blocks. These are made by Mullard Ltd., Manor Royal, Crawley, Sussex, U.K. The blocks are intended for building digital control, small

computing, or measuring systems. The logic facilities within the blocks require only a minimum of external components. The circuits are based upon NPN germanium transistors and have a power requirement of $V_p (+12 \pm 5\%)$, $V_n (-12 \pm 5\%)$. The voice key uses five sorts of blocks from the Series 10 range; they are the pulse shaper, the timer, the one-shot multivibrator, the gate inverter, and the relay driver. Brief details of these blocks are as follows:

The pulse shaper (P.S.) comprises a Schmitt trigger and an inverting amplifier. It may be used to redefine signals of small amplitude or slow rise time, producing at its output the dc levels standardized for the other units in the Series 10 range. The timer unit can be used for delays, the duration of which depends upon the values of external capacitors and resistors. The unit has a recovery time of 15 microsec. There is an output gate facility (via an external diode). They also have a topping-up facility which is used on Timer 2 (Fig. 2), i.e., if during the delay the input is triggered a second time, the delay action will start all over again. The one-shot multivibrator (O.S.) has an intrinsic delay of 4 microsec which may be extended with an external capacitor. The gate inverter consists of a diode-resistor gate followed by an inverting amplifier stage. By suitable interconnections, gate inverters may perform all the basic logic functions. In Fig. 2, they are used either simply as inverters (INV.) or to perform a NOT-AND (NAND) function. The relay driver (R.D.) comprises a single input positive diode gate followed by a noninverting amplifier.

The operation of the voice key may be followed from Fig. 2. Speech

signals are fed through a low-impedance transformer, rectified, and then fed to a pulse shaper (P.S.). If the magnitude of the signal exceeds the tripping level of the unit, the signal is reshaped and fed to the trigger inputs (t) of the two timer units. Initially the gate of Timer 1 is open. As soon as Timer 1 is triggered, its gate is closed by its own output fed via the gate inverter (INV.1) for a preselected duration controlled by external capacitors and resistors (not shown). If the signal exceeds this duration (the throwaway time), a brief output pulse is generated before the gate of Timer 1 is closed again by its own output.

This output pulse from Timer 1 is fed to the one-shot multivibrator (O.S.), which extends the duration of the pulse and feeds it to the gate inverter which performs a NAND function. The duration of this pulse is sufficient to gate Timer 2 for long enough to allow low speech frequencies to trigger the timer. Once Timer 2 has triggered, its output controls the gate on Timer 1 for the duration of its output pulse. The duration of the output pulse can be varied by external capacitors and resistors (not shown). The duration selected is determined by the individual speech characteristics being processed and the requirements of the operator. This output is fed to the remote control of the receiving recorder via the relay driver. The tape transport is set in motion by the onset of the output pulse and stopped by its offset. If the timing has been arranged correctly, this should coincide with the speech signal fed into the line input of the receiving recorder.

If used in conjunction with a pen oscillograph, the voice key produces a simple on-off record of the signal input fed to it. In this way, the temporal structure of the speech samples used may be inspected. A more sophisticated technique for measuring and recording the output of the voice key on punch paper tape and processing it by computer has been developed. A report of its performance is in preparation.

REFERENCES

- FOULKE, E. Methods of controlling the word rate of recorded speech. *Journal of Communication*, 20, 1970, 305-314.
- HUGGINS, A. W. F. Accurate delays for auditory feedback experiments. *Quarterly Journal of Experimental Psychology*, 1967, XIX, 78-80.
- M.E.L. Equipment Co. Ltd. *Design with Series 10 circuit blocks*. Electronics & Control Division, M.E.L. Equipment Co. Ltd., Manor Royal, Crawley, Sussex, U.K., 1964.