Behavior Research Methods & Instrumentation
1978, Vol. 10 (6), 845-847

A simple procedure for accurately manipulating face-voice synchrony when dubbing speech onto videotape

JOHN MacDONALD, DOMINIC DWYER, JENNY FERRIS
and HARRY McGURK
University of Surrey, Guildford, Surrey, GU2 5XH, England

A system is described involving an audio-video tape recorder, a control device, and a delay timer that dubs speech sounds onto prerecorded lip movements on videotape. Accuracy of synchrony or measured asynchrony between the sound and lip movements is a few milliseconds.

We have been investigating the effect on speech perception of disruption of the normal face-voice synchrony that occurs in natural face-to-face speech (McGurk & MacDonald, 1976). The basic paradigm is one in which the subject is asked to watch and listen to a video film of a young woman speaking and is asked to repeat what she says. The stimuli are short prerecorded sequences in which the model utters simple syllables while fixating a television camera. Two types of disruption can be introduced: (1) time asynchrony, in which the sound is displaced relative to the lip movements, either forward or backward, and (2) phonetic asynchrony, where a new phonological syllable is dubbed onto a previously recorded set of lip movements for a different phoneme (e.g., the sound /ba/ dubbed onto the lip movements for /ga/). Temporal disruption may also be involved in phonetic asynchrony, since there is no a priori point of temporal synchrony between a syllable presented auditorily and a different set of lip movements. It is possible, however, to establish empirically a range of voice onset and lip movement onset times that are subjectively experienced as coincident.

The procedure described here provides a technique for dubbing speech onto the audio track of a videotape, which allows accurate placement of the new sound relative to either the lip movements or the previously recorded speech. The technique is conceptually very simple, requiring an inexpensive control unit and no modification of commercially available apparatus.

EQUIPMENT

The apparatus consists of a good-quality videotape recorder with audio dubbing facilities (National Panasonic Edit NV-3160), a good-quality audio tape recorder with remote start and stop facility (Tandberg 9000X), a variable delay timer (in this particular system an Electronics Developments three-field tachistoscope), a digital timer (Forth Instruments), and a specially designed control unit to interface between the videotape recorder and the tachistoscope. Figure 1 shows the basic systems diagram.

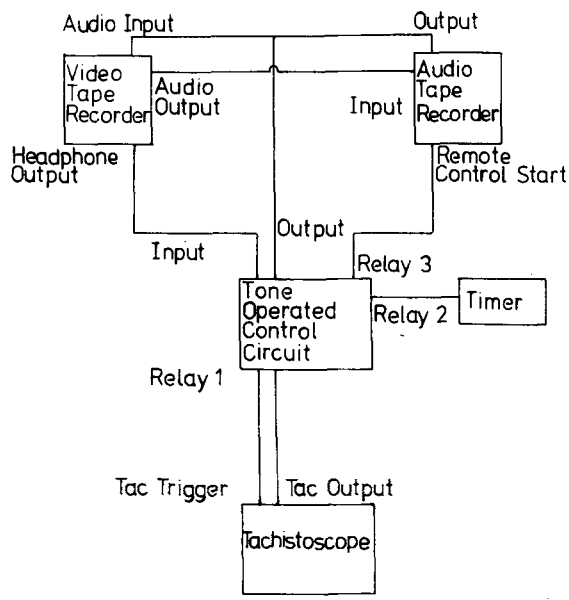

Figure 1. Systems block diagram for audio dubbing of videotape.

In order to construct the stimuli, it is necessary to prepare good-quality video and audio recordings of the syllables. This is achieved by having the recordings made in a professionally equipped television studio and by using a well-practiced model with good articulation. The result is good equalization of the duration and intensity of different syllables. The sound track is then either transferred to an audio tape recorder or recorded simultaneously but independently of the video recordings. A 1-kHz tone burst is recorded onto the audio track of the video recording approximately 5 sec before the onset of the recorded syllable. This is achieved by using the audio-dub facility of the video recorder and the manually operated tone generator situated in the control unit (see Figure 2).

This generator was constructed using a 556 dual timer. The left side of the 556 was wired in a monostable (one-shot) mode, to give a short positive pulse on its Pin 5 output, when Pin 6 is momentarily switched to earth. Switch 1 is a "push-to-make" switch that the

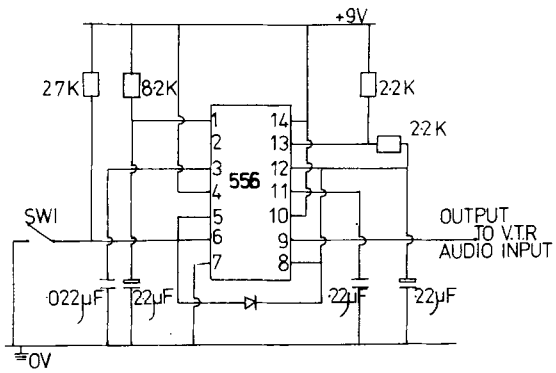

Figure 2. Circuit diagram of tone-burst generator.

operator depresses whenever the tone burst is required. This sends a pulse to trigger the right side of the 556, which is wired as a multivibrator with a frequency of 1 kHz. It is important to remember throughout this operation not to erase the original speech signal. The exact time between the onset of the tone and the onset of the speech signal (T_1) is measured. The control unit contains a tone-operated switch that can be set to activate the timer at the onset of the tone and to stop the timer at the onset of the speech signal. Although the relay introduces some delay at this point, it can be discounted, since the same delay will operate in the recorded sequence. Figure 3 shows the tone-operated relay circuit.

The 1-kHz tone-operated switch was constructed from a 741 operational amplifier wired to act as a sensitive "ac overvoltage" switch. A twin T filter network was incorporated in the negative feedback path of the 741, wired as a noninverting amplifier. The gain of the operational amplifier is adjusted using R, so that TR1 is switched "on" when the tone burst or speech signal is fed into input. When TR1 is switched "on," the two relays are closed. These two relays are used to trigger the tachistoscope timer and the digital timer. To obtain a narrow bandwidth, the input signal must be limited to less than 10 mV rms, since the bandwidth is proportional to the amplitude of the input signal of the circuit. Next, the tape on the tape recorder is marked and set up so that, on being activated via the remote start switch, it allows a delay of approximately 300 msec (T_2) before the onset of the speech signal. This allows the recorder to achieve steady state running before the speech starts and also allows the operator

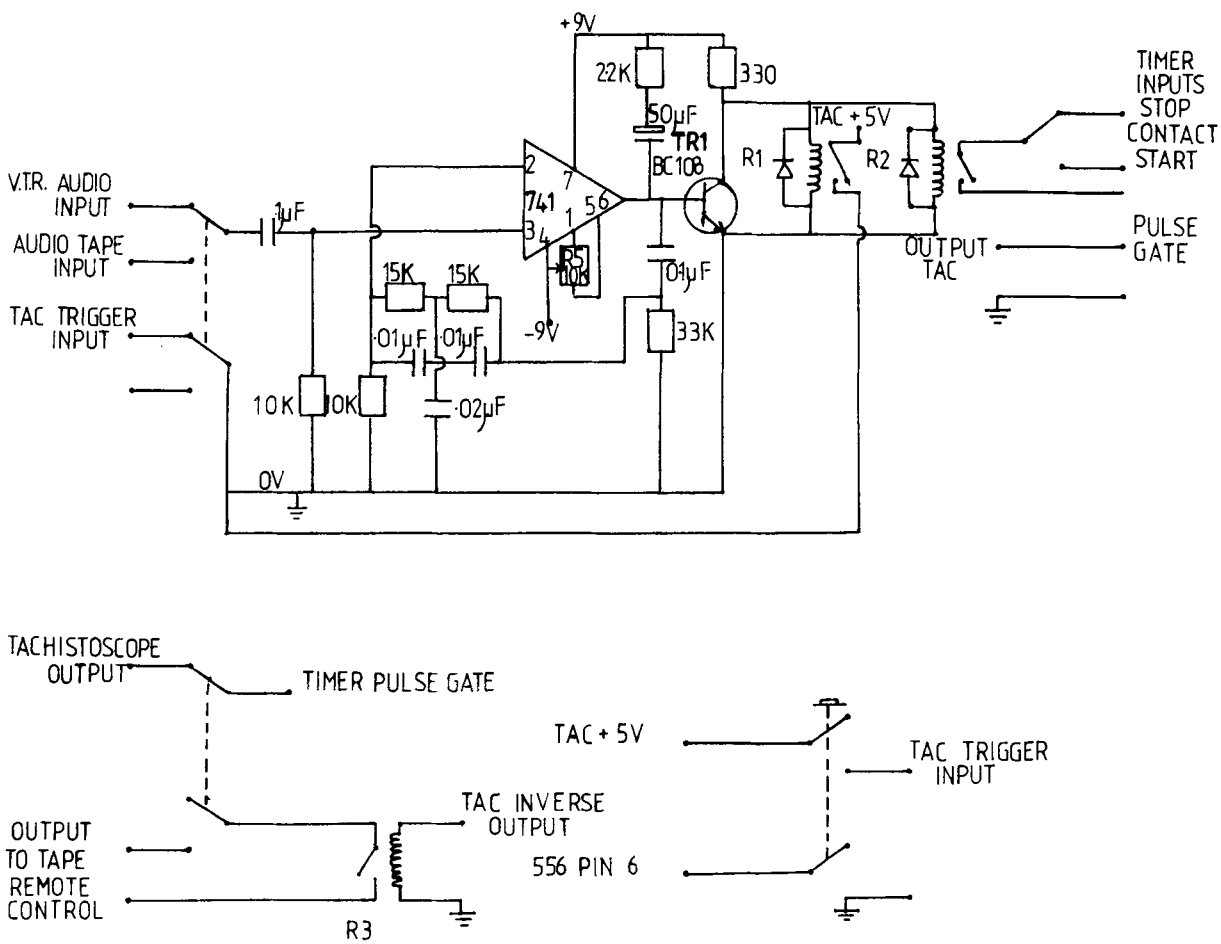

Figure 3. Circuit diagram of tone-operated relays.

to align the tape at exactly the same position, in case the desired offset is not achieved on the first run. The video recorder headphone output is connected to the input of the tone-operated switch (TOS) and the output of the TOS set to trigger the tachistoscope timer. Finally, the output from the tachistoscope timer is connected to the remote start input of the tape recorder. The tachistoscope timer is set up for T_3, which is T_1 plus or minus t msec, the offset required, dependent on whether the new audio signal has to lag or lead the position of the old audio signal, minus T_2, the delay introduced between the activation of the remote start on the tape recorder and the onset of the new audio signal. The video recorder is then started. The tone from the audio track fires the tachistoscope timer and the digital timer. At this point, one can either depress the audio-dub button on the video recorder and make the new recording or set the TOS to read from the tape recorder. After the elapsed time on the tachistoscope, the remote start on the tape recorder is activated and the digital timer stopped at the onset of the audio signal. The time can be checked to see if the required delay has been achieved. If not, then the tachistoscope timer can be adjusted accordingly and the procedure rerun.

ACCURACY

The accuracy of the procedure can be affected at a number of points in the chain of operation. Two have been mentioned: the delay introduced at the relays of the TOS, which should be constant over not only the initial timing stage and the rerecording stage, but also over repeated timings. The second delay is introduced at the remote start of the tape recorder. Over a repeated measurement, this should be approximately constant and can, in any case, be compensated for, by adjustment of the tachistoscope timer. The third problem that has not been addressed is that of tape stretch due to repeated forward and backward playing of the video recorder. When tests were conducted over a series of 15 repeated timings of the same recording (tone burst to speech), the variation in times was of the order of 5 msec. It appears that the problem of tape stretch over the period dealt with in this procedure, from original timing to recording the new sequence, is negligible, given also that the length of the tape involved is of the order of 10-15 sec duration.

The accuracy of the offsets that can be produced is very high. In one experiment, we required offsets from −300 msec through 0 to +300 msec, in steps of 20 msec, and found that offsets accurate to within 2-3 msec could be reliably produced after only a short period of familiarity with the procedure.

DISCUSSION

Throughout this description, it has been assumed that the researcher is dubbing the new sound with respect to an arbitrarily placed marker tone, which has a particular temporal relationship to the old sound. However, the procedure has a more general application in that, as described, the tone burst also has a known temporal relationship with the visual events on the videotape. In theory the method can be used for any visual sequence. The problem is to determine the exact time between the marker tone and the visual event.

The system's most immediate use is with audio-visual sequences that have their sounds in some direct non-arbitrary relationship to the visual events. Speech and lip movements have exactly these characteristics.

REFERENCE

McGurk, J., & MacDonald, J. Hearing lips and seeing voices. *Nature*, 1976, **264**, 746-748.

(Received for publication February 2, 1978;
revision accepted May 24, 1978.)