Ann. Bull. RILP, No. 9, 15-20 (1975)

SYNCHRONIZATION OF A FILM IMAGE WITH A SPEECH SIGNAL

S. Sekimoto, H. Imagawa and M. Sawashima

In frame-by-frame analysis of articulatory movements of speech organs on film, an accurate correspondence of each frame to the time course of the speech signal is essential. This paper is a brief description of our synchronization technique for this purpose. The technique involves two procedures: generation of a frame-by-frame signal from the cine-camera and placing a timing mark on both the film and the audiotape.

We use a 16 mm "BEAULIEU" attached to a fiberscope. In the BEAULIEU camera there is a one revolution per frame drive shaft installed on the side of the camera body. This shaft was originally meant to be connected to a sync. pulse generator, one of the accessories of this camera, which generates and feeds a special signal to a tape recorder for synchronizing the sound with the film at a regular frame rate, i. e. 24 frames per second.

With this commercially available set, however, we need a special type of tape recorder such as "NAGRA" that has a special recording head for the camera signal. Furthermore, the system is designed to synchronize the sound with the moving picture when the film is projected and is not used in matching the visual trace of the sound to the film frame, especially when the film is taken at a frame rate faster than 24 per second.

We then manufactured a home-made frame signal generator which is also connected to the one revolution per frame drive shaft. The signal is recorded on one channel of the two channel tape recorder with the voice being recorded on the other channel. Visual traces, oscillograms or sound spectrograms of both the frame signal and the voice are later made, in order to synchronize the time relationship between the two for any frame rate of filming.

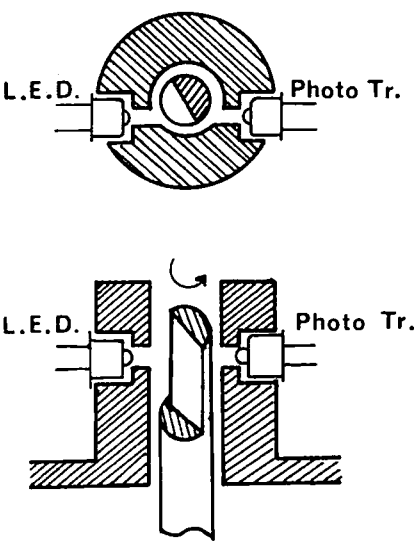

Fig. 1: The screen shaft in the cap with light-emitting diode (L. E. D.) and photo-transistor.

The mechanical part of the signal generator is a special shaped screen shaft which is attached to the drive shaft (Fig. 1). At the top of the screen shaft, approximately half of its diameter is cut off for the length of 2 cm. This shaft is covered by a cap in which a light-emitting diode (L.E.D.) and a photo-transistor are placed facing each other (Fig. 1). When we power L. E. D., the amount of the light reaching the photo-transistor is modulated by the rotation of the screen shaft which is driven by the one revolution per frame drive shaft. A block diagram of the signal generator is shown in the upper part of Fig. 2. L. E. D. is powered by a sine wave current with a frequency of 4 kHz or 8kHz. The frequency was selected for the purpose of superimposing sound spectrograms on voice signals. The output of the photo-transistor is high-pass filtered and sent through a switch (SW. 1) to a mixer and then to a tape recorder. Figure 3(a) presents the timing relationship between the frame signal and the camera shutter.

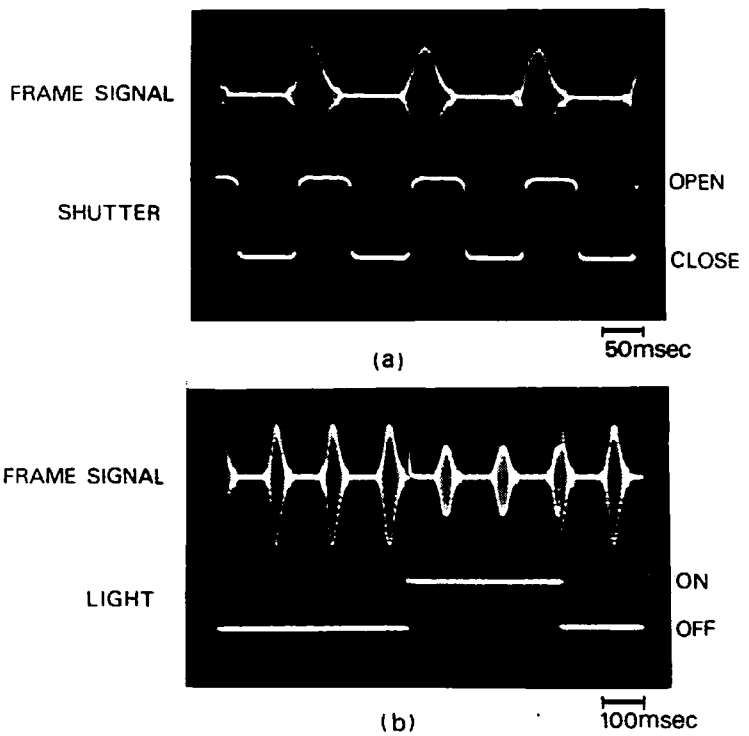

Fig. 3: Time correspondence between the frame signal and the camera shutter (a), and function of the timing mark switch (b).

From this we can see exactly where in the frame signal the timing of the open shutter, i. e. the film image, should be located. The frame rate in this figure is very slow, but the timing relationship is quite consistent for any fast frame rate.

The lower part of Fig. 2 shows a block diagram of the timing mark. Here, we use 6 L. E. D. 's which are placed in the adaptor connecting the fiberscope to the camera. L. E. D. 's in position are shown in Fig. 4. They

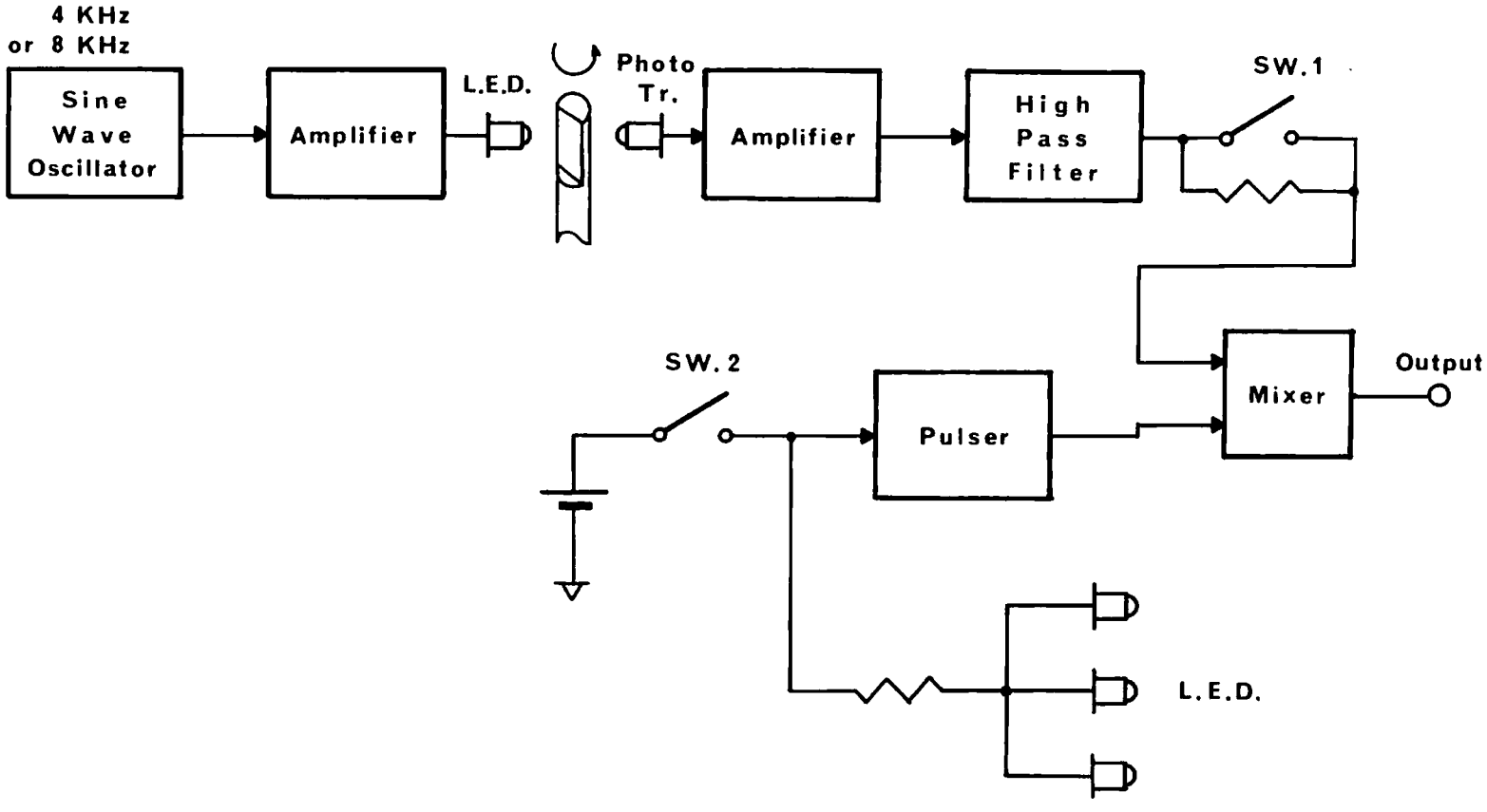

Fig. 2: Block diagram of the frame signal generator (upper line) and the timing mark generator.

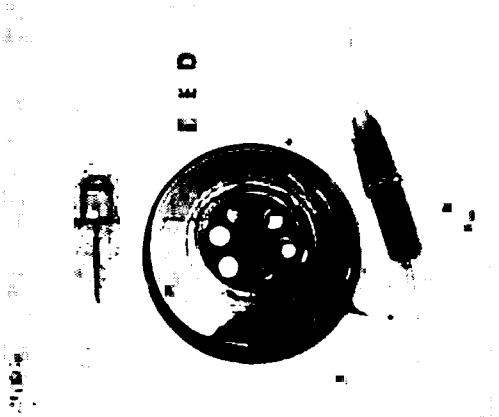

Fig. 4: L. E. D. 's for timing mark placed in the sine-adaptor.

are powered by a DC source in this case. When we press a switch (SW 2. in Fig. 2) the light is turned on with a very sharp rise time and the film is exposed to the light until the switch is turned off. Simultaneously with the on and off state of the switch, a pulser generates a pulse and sends it to the mixer and then to the tape recorder. Also, SW. 2 is electrically linked to SW. 1 so that pressing (on) SW. 2 makes SW. 1 turn off and reduces the amplitude of the frame signal. The function of SW. 2 is shown in Fig. 3(b). By the exposure of the film and by the pulse with a sudden change in the amplitude of the frame signal, we can easily and accurately achieve synchronization between a particular part of the film and the trace of an acoustic event. In the actual experimental run, we usually place one timing mark before and after each utterance. The timing mark is also utilized to place appropriate identification marks on both the film and the sound trace.

Figure 5 is a picture of the camera with the frame signal generator, cine-adaptor and the fiberscope in position. Using this system, at the frame rate of 50 per second (which is usual in our experiments) we can match each film frame to the time course of the speech signal with an accuracy of approximately 10 msec.

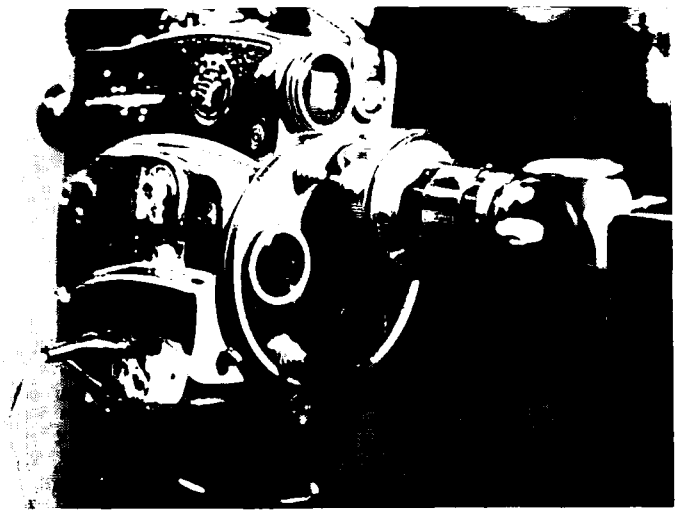

Fig. 5: A picture of the BEAULIEU with the frame signal generator, the cine-adaptor and the fiberscope in position.