Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 569–573
(S. Karger, Basel/New York 1965).

Communication Sciences Laboratoy, University of Minnesota, Minneapolis

Laryngeal Analog Synthesis of Glottograms

By Ronald W. Wendahl, Minneapolis, Minn.

The subject of my paper is the use of an electrical analog of the larynx to study vocal productions.

Before we go into the operation of the analog, let me state some of the simplifying assumptions that must be made and the limitations created by these simplifying assumptions. First, one may look at a typical area analysis of the operation of the vocal folds. In figure 1 openings and closings of the folds are depicted on the ordinate, as a function of time on the abscissa. If one were interested only in areas of opening as a function of time he could plot in a single dimension as a function of time and he would have the configuration shown in figure 1-B. It would not be a great violation to translate directly to the waveform shown in figure 1-C if one were interested in preserving frequency information and were willing to accept the differences one might obtain in the harmonic spectrum.

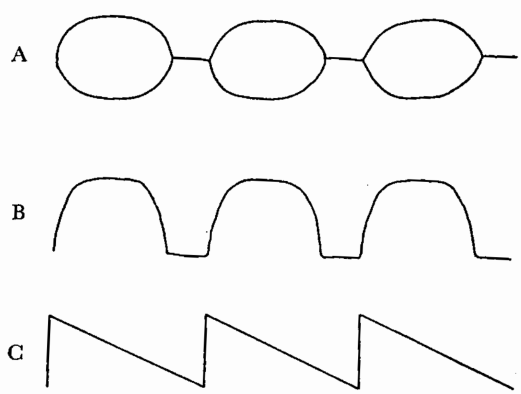

Fig. 1

In figure 2 one may see that frequency information may be preserved in several ways. Figure 2-A shows the fold area openings as a function of time complete with the closed as well as the open phases of the vocal fold actions. If one desired to preserve only frequency information he might use a waveform as is shown in figure 2-B. However, there is good reason to believe that one might have to consider the closed phase in attempting to synthesize human vocal productions and therefore he might be forced to use a waveform as is shown in figure 2-C.

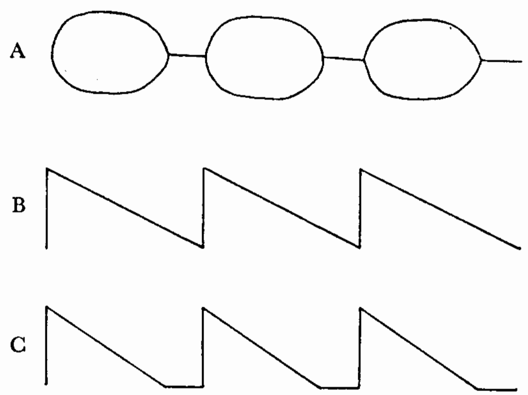

Fig. 2

Figure 3 shows a not atypical mode of vibration of the vocal folds. The second laryngeal opening in each case, is of lesser amplitude and shorter in time than the first. One may wish to be able to program amplitude as well as frequency information as in the case of figure 3-A, or he may wish to ignore amplitude information unless he suspects that the amplitude differences between adjacent wave fronts are sufficient to create a pitch change. In that case he may wish to simulate figure 3-A with figure 3-C.

In the tape recordings you will hear, all simplifying assumptions were made. The closed phase of the folds was ignored and amplitude differences between successive wave fronts also were ignored. The reason for this procedure was purely that this recording was considered to be a first approximation to the synthesis of harsh voice quality.

Before playing the tape recordings made from LADIC, may I go briefly into the operations of the device. Figure 4 shows a simplified block diagram of LADIC. The output waveform is gener-

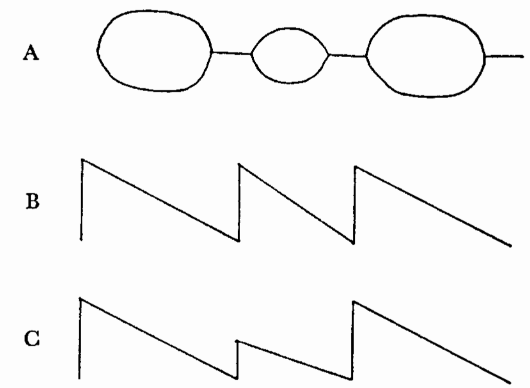

Fig. 3

ated by a master oscillator which is controlled in frequency and amplitude by a small special-purpose computer. The master oscillator is an unijunction transistor in a modified 'relaxation' circuit. This arrangement produces two waveforms; one, a slightly nonlinear sawtooth, and the other a narrow clock (synchronizing) pulse. The memory unit is binary which controls the sequence of the frequency changes. Switching from one frequency to another, or from one amplitude to another is accomplished through the use of saturated transistor gates that operate with a speed of approximately

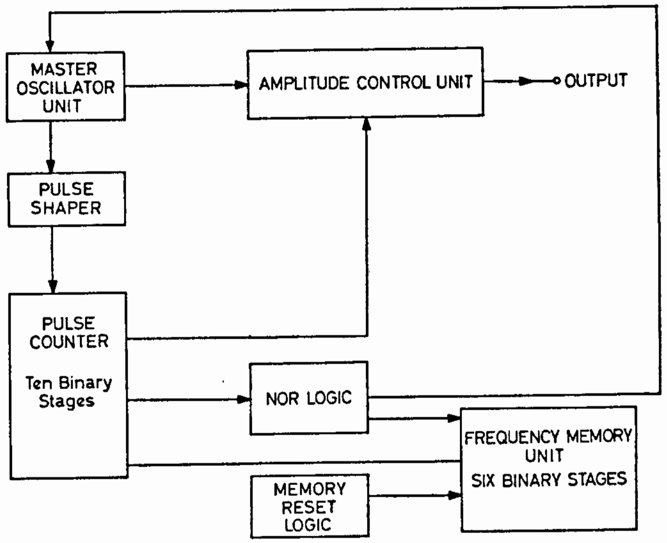

Fig. 4

500 nanoseconds. The pulse counter is also binary with the desired number of cycles of each frequency being programmed by means of toggle switches. The present counter capacity for a given frequency is 2,027 cycles. When a programmed count has been reached, NOR logic circuits cause any or all of the following events to occur in accordance with the present program.

a) The memory may be advanced to the next instruction. This causes the next sequence to begin.

b) The memory decode logic may shift the master oscillator frequency to any desired value between 50 and 1,000 cps.

c) The output amplitude may be switched to any predetermined value from zero to maximum.

d) The pulse counter may be set to zero count.

e) The count decode switches are scanned and the proper count for a new frequency is automatically selected.

The preceding sequence may be repeated, without transients, for a maximum of 63 distinct frequencies. Depending upon the purpose of the experiment, the output of LADIC either is fed directly into a tape recorder for data storage or is fed through vowel shaping filters. When it is deemed necessary for the stimuli to sound humanoid, vowel shaping circuits are used.

The procedures used to obtain the source data for programming LADIC are frequency measures obtained from phonelegrams, from ultra-high speed cinematography, and from transillumination. Each of the techniques offers advantages and has limitations. The phonelegraphic technique has the limitation of the transposition of times from points often obscure and the limited accuracy of the dividers and ruler used. A further drawback of the phonelegraphic technique of obtaining frequency information is that the open-phase time of the vocal fold vibration is indistinguishable from the closed time.

The camera technique offers the advantage of being able to separate the open from the closed phases of the vibratory cycle; however, it has the disadvantage that it is a discrete and not a continuous operation even at such speeds as 7,400 pictures per second. Hence, an error around 135 cps of plus or minus 3 cps is not unexpected.

The transillumination technique offers the advantages of being able to distinguish the open and closed phases of vibratory patterns and that it is a continuous reproduction but offers the ambiguity of divider measures.

Several subjects were chosen having deviant voice qualities. Cinematographic, photocell, and light-writing oscillographic measures were made of their phonations. From the various measures LADIC was programmed and fed to vowel filters to simulate the original phonations.

Tape recordings will present the human phonations and their synthesized counterparts. In many of the tape recordings a remarkably similar quality between the two types of stimuli can be heard. In some recordings the reverse is true. The rationale for both the successes and the failures will be described.

Author's address: Dr. Ronald W. Wendahl, Communication Sciences Laboratory, University of Minnesota *Minneapolis, Minn.* (USA).