FUNDAMENTAL FREQUENCY TRACKING

ARNE RISBERG

The earliest and still most frequently used method for automatic tracking the fundamental frequency of speech both for phonetic studies and in analysis-synthesis-telephony systems (vocoders) is the method of low-pass filtering (1, 2). Speech is then considered as a quasiperiodic signal with a fundamental frequency component and overtones. The fundamental component is filtered out by means of a low-pass filter and subjected to some form of frequency measurement.

The difficulty with this method is that the fundamental frequency varies very rapidly over about two octaves in continuous speech of a single speaker and that human speech covers a total frequency range of more than three octaves. This implies that no fixed value can be given to the cutoff frequency of the low-pass filter unless the signal to noise ratio is extremely good which can be achieved only in laboratories using special soundtreated rooms and good microphones. In practice, the signal to noise ratio cannot be expected to be good enough for the demands of this technique neither in phonetic research nor in analysis-synthesis-telephony systems and certainly not if the voice fundamental is very weak or absent in the signal available for analysis.

Two schemes which have been commonly suggested for overcoming this problem are based on fundamental enhancing circuits before the filter and adjustable filters. These two techniques can of course be used simultaneously. In the enhancing circuit comprising some form of non-linear network, the level of the fundamental frequency relative to the overtones, especially the second harmonic, is raised or the fundamental is regenerated from higher frequency regions if it is missing in the original signal.

As non-linear networks have been used half-wave and full-wave rectifiers, square-law rectifiers, rectified single side-band signals, etc. (1, 3, 4). None of the methods used up till now have been fully successful over the whole range of fundamental frequencies and wave shapes encountered in human speech but all of them give improvements in the fundamental to harmonic ratio during part of the speech, but at the same time they tend to destroy information during other parts (5).

The technique of using a variable cutoff frequency of the low-pass filter has been tried by many investigators. In fundamental frequency trackers used in phonetic research filters can easily be varied manually (6). The cutoff frequency is selected approximately equal to the mean value of the fundamental frequency of the talker

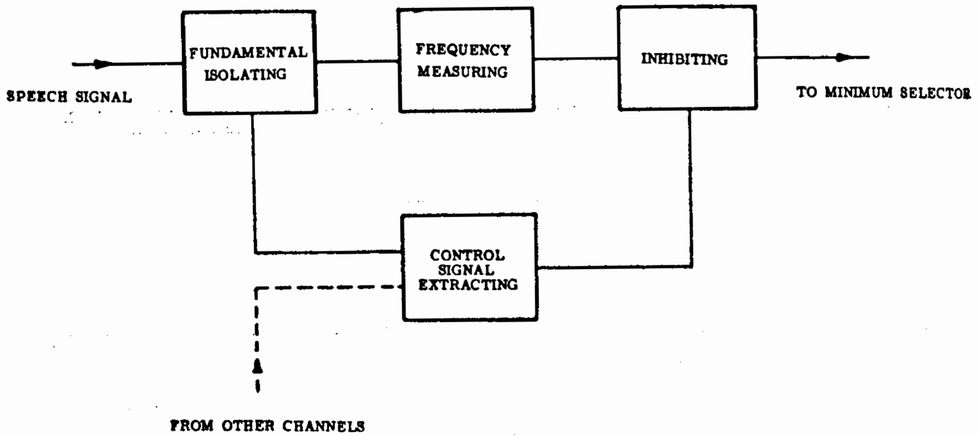

Fig. 1. Block diagram showing the principle of operation of one of the channels in the proposed multichannel fundamental frequency tracker.

whose speech is being investigated. This gives a relatively good result, especially if the signal to noise ratio is reasonably good and the range of the fundamental frequency is not too great. If the fundamental frequency range is great or if a manual selection of cutoff frequency cannot be accepted, as in vocoders, a technique using electronically controllable filters or switching between two or more filters with fixed cutoff frequency has been adopted (7, 8, 9). The results of these experiments have not been successful mostly due to transients generated by the controlling or switching of the filters.

At the Royal Institute of Technology almost all of these different techniques to improve the basic low-pass filter circuit have been tried and as a result of the experience gained an approach to the problem of fundamental frequency tracking has been proposed using two or more channels in parallel. Block diagram of one of these channels is seen in Figure 1. Each channel gives an output signal proportional to frequency and the lowest of these is selected in a minimum selector. As each channel is designed to work only for a limited range of fundamental frequencies it is possible that a frequency lower than the fundamental, e.g. hum, is measured and in each channel a circuit must be incorporated that prevents such signals from reaching the minimum selector. This circuit gets control voltage either only from the channel in question or also from other channels. Two examples of this type of circuits are shown in Figures 2 and 4. Examples of signals in the three channels of circuit in Figure 2 can be seen on Figure 3.

The advantage of this circuit compared with circuits using electronically tuned filters or switching between filters is that switching transients from the filters are avoided as all switching is made on DC signals representing the first frequency measure. By using several circuits in parallel maximum use can be made of each system and the disadvantage of a particular system can be compensated by circuits having complementary characteristics.

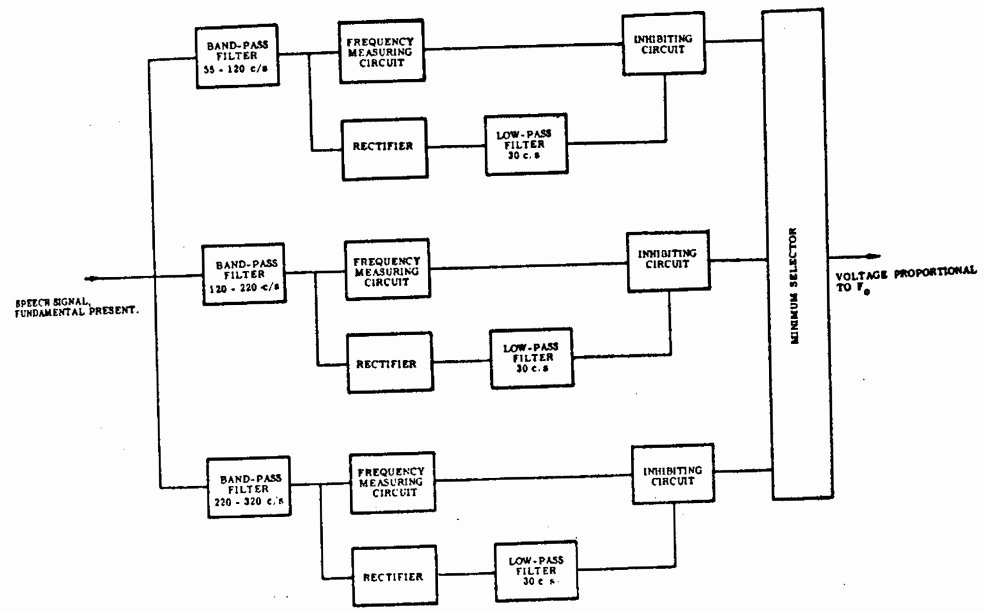

Fig. 2. Fundamental frequency tracker using three band-pass filter channels.

Our tests have been very promising but much work has still to be done to find the best combination of circuits.

For use in phonetic research a considerable simplification can be made by making this choice visually instead of using an automatic choice between the outputs. The outputs from two systems that are considered best for the voice being investigated are recorded on an oscillograph, for example the Mingograph, and the selection is made by inspecting the recordings. This is in most cases easy as most errors are due to missing fundamental or too strong second harmonic.

For such applications a simple equipment has been constructed comprising a low-pass filter with cutoff frequency adjustable in 1/3-octave steps from 70 to 530 c/s. The filter was designed for minimum overshoot and 18 dB/oct attenuation. In practice it was found to be advantageous to have more gradual cutoff characteristics and therefore an additional simple RC-network with cutoff frequencies 50 and 100 c/s was incorporated. To minimize low frequency disturbances when a fundamental enhancing circuit is used a high-pass filter with cutoff frequency 60 c/s can be switched in. The three filters are combined to give best result. The filters are followed by an infinite clipper, differentiating circuit and a sawtooth generator that is synchronized by the spike generated in the differentiating circuit (1). The height of the sawtooth wave-form is made approximately inversely proportional to the logarithm of the fundamental frequency from 50 c/s to 400 c/s. As fundamental enhancing circuit a full-wave or a half-wave rectification is used, full-wave rectification mostly for low fundamental frequencies and half-wave rectification for high fundamental frequencies.

The high-pass filter of 60 c/s, which is a necessary supplement to the rectification

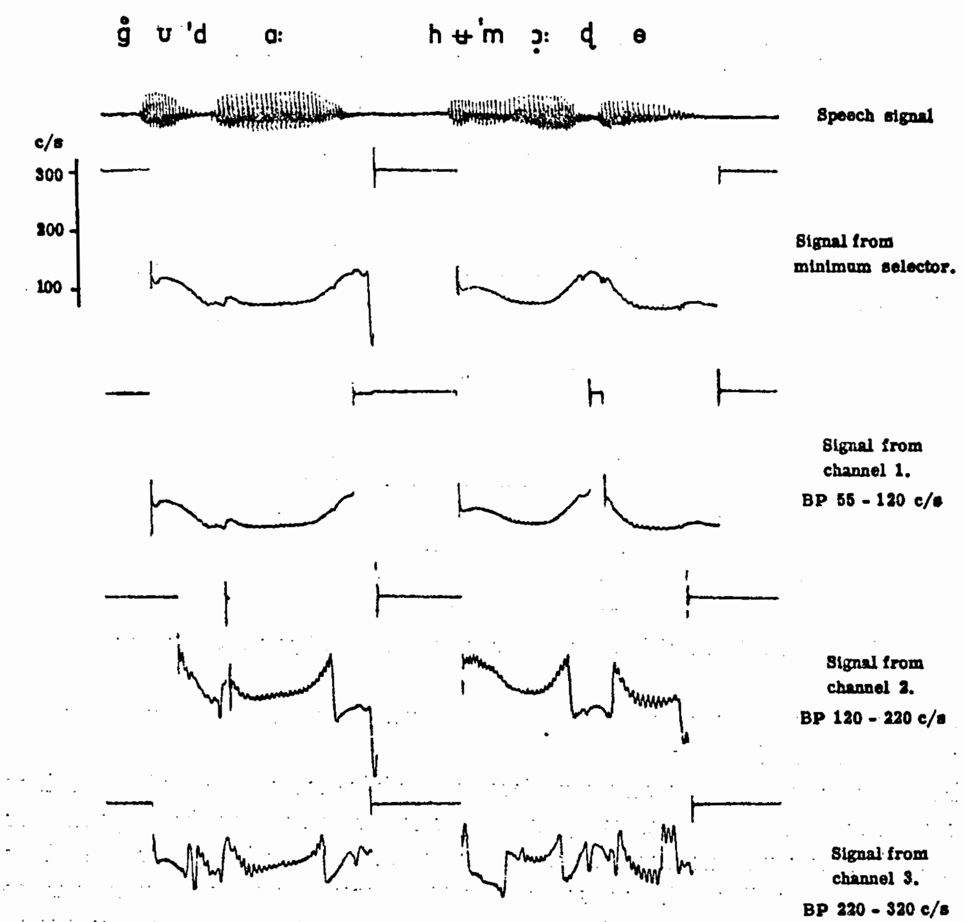

Fig. 3. Signals for a male speaker in each channel and signal from the minimum selector in the fundamental frequency tracker shown in Fig. 2. Note the inhibiting of the signal in channel one when the fundamental frequency rises and that the minimum selector then follows the signal in channel two.

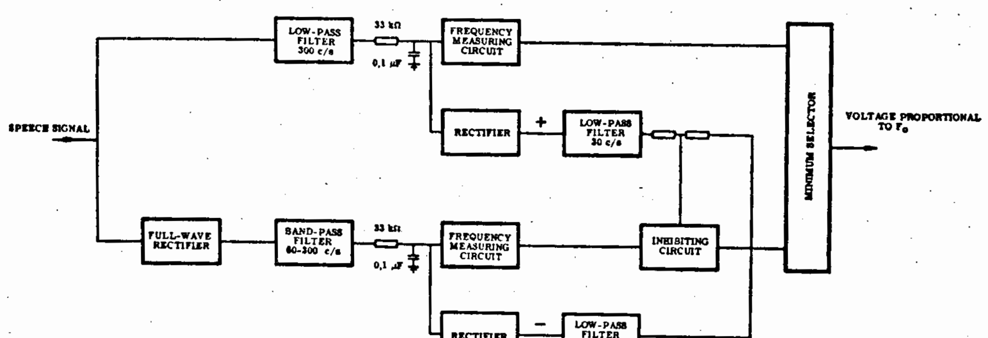

Fig. 4. Fundamental frequency tracker with a low-pass filter channel and one channel containing a full-wave rectifier. The channels complement each other, the low-pass channel being most effective for high fundamental frequencies and the channel with full-wave rectification most effective for low fundamental frequency.

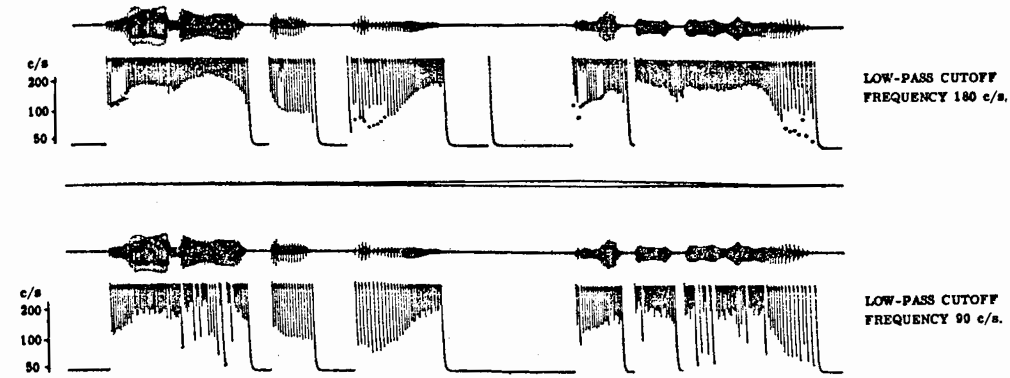

Fig. 5. Two recordings from the simple fundamental frequency tracker intended for phonetic research using two different low-pass filter cutoff frequencies. The fundamental frequencies for the parts where erroneous values have been obtained in the upper recording have been replaced by values from the lower recording.

stage, introduces unavoidable errors at those parts of the speech wave where the overall amplitude changes rapidly, e.g. in the onsets of voiced sounds.

Example of a recording using this instrument can be seen on Figure 5.

Speech Transmission Laboratory
Royal Institute of Technology
Stockholm

REFERENCES:

(1) Grützmacher, M., and Lottermoser, W., "Über ein Verfahren zur trägheitsfreien Aufzeichnung von Melodiekurven", *Akust. Z.*, 2, Sept. (1937), pp. 242–248.
(2) Dudley, H., "Remaking Speech", *J. Acoust. Soc. Am*, 11 (1939), pp. 165–177.
(3) Kriger, L. V., "Pitch Extraction for Speech Synthesis with Special Techniques for Use in Digitized Bandwidth Compression Systems", March 1959, AFCRC–TR–59–116.
(4) Lerner, R., "A Method of Speech Compression", M.I.T., Lincoln Laboratory, *Group Report* 36–41, Aug. 1959.
(5) Risberg, A., "Voice Fundamental Frequency-Tracking", R.I.T., Speech Transmission Laboratory, *Quarterly Progress and Status Report* No. 1/1961, April 15, 1961, pp. 3–6.
(6) Kallenbach, W., "Eine Weiterentwicklung des Tonhöhenschreibers mit Anwendungen bei phonetischen Untersuchungen", *Akust. Beihefte*, Heft 1 (1951), pp. 37–42.
(7) Halsey, R. J., and Swaffield, J., "Analysis-Synthesis Telephony, with Special Reference to the Vocoder", *J. of the Inst. of Electr. Eng.*, 1947, pp. 391–411.
(8) Edson, J. O., and Feldman, C. B. H., "Derivation of Vocoder Pitch Signals", US Pat. 2,906,955.
(9) Miller, R. L., "Determination of Pitch Frequency of Complex Wave", US Pat. 2,627,541.