A Simple Hardware Pitch Extractor*

BERNARD A. HUTCHINS, JR., AND WALTER H. KU

Cornell University, School of Electrical Engineering, Ithaca, NY 14853, USA

The need exists for a simple, hardware, real-time pitch extractor for speech research, training, and similar applications requiring less than perfect accuracy. A simple combination of a band-pass filter and a fast amplitude detector formed from a tapped analog delay line provides useful and reliable pitch contours from a live speech input.

0 INTRODUCTION

While lists of publications on the subject of pitch extraction would run many pages long, most are concerned with computer software, and only a few of these become involved with real-time operation [1], [2]. Still fewer also consider hardware in any way [3]–[6], and those that involve hardware that is relatively simple are quite rare [7]–[9]. Such simple, hardware, real-time pitch extractors are useful in a number of applications, such as linguistic studies, language training, and speech therapy aids for the deaf. Depending on the accuracy of the extractor and the range of inputs to be processed, simple devices may also suffice for speech encoding systems and for electronic music applications. The pitch extractor, or pitch-to-voltage converter, considered here is a device which takes ordinary speech as an input and gives a voltage contour proportional to the pitch as the output. This contour is available for display on an oscilloscope face, for storage, or for transmission as needed.

As always, the distinction between pitch and frequency must be made, and this probably is easiest to understand if we consider pitch extraction in basic terms, appropriate to the actual device to be discussed. The pitched (or voiced) speech signal is produced in the throat by pulses of air passing through the vocal cords, and this excitation is filtered (convolved) by the vocal track. The result is that the actual speech signal is a mixture of excitation features and resonant features. The waveform actually evolves on a cycle-to-cycle basis, and it is virtually assured that there will be no continuously available single prominent peak in the waveform corresponding to the excitation. Thus a simple pitch extraction scheme is to preprocess the speech signal so that features corresponding to excitation are enhanced, while features resulting from resonance are reduced. Once the peak is isolated, its point of occurrence can be identified, and its repetition frequency can be reported with a much simpler frequency-to-voltage (f/v) converter.

1 THEORY OF OPERATION

1.1 Peak Amplitude Detector

In the pitch extractor or p/v converter described here, the isolation of a single strong feature is achieved as a two-step process. First a simple band-pass filter, chosen experimentally, serves to strengthen the fundamental frequency in a region where resonances (formants) tend to emphasize upper harmonics. The second element in the process is a very rapid peak amplitude detector formed from a tapped delay line which serves to detect the height of the maximum peaks as well as their point of occurrence.

The need for some sort of peak amplitude detector can be understood since it is necessary to know the maximum amplitude in order to detect the moment at which this maximum is achieved. Traditional amplitude detectors such as full-wave rectifiers and peak detectors are often unsatisfactory in this application because they require some sort of averaging time to determine the

* Manuscript received 1981 July 6.

 0004-7554/82/030135-05$00.75

amplitude level, and the actual parameters of speech can change significantly over the few pitch cycles it takes to get this average. The parallel of this situation with the energy–time uncertainty relationships of physics becomes apparent, and it is indeed interesting that human speech has evolved to approach this limit in a significant way.

The ideal peak amplitude detector would be a device that looked at exactly one cycle of a speech waveform (one pitch period), and searched this waveform instantaneously for a maximum. This is analogous to the process of determining the amplitude of a waveform by scanning it by eye as it is displayed on the face of an oscilloscope. In a practical case, sampled analog values of a speech waveform can be stored, and then the maximum of these stored values can be determined. In the present pitch extractor a fixed-delay tapped analog delay line is used to make available samples of a preprocessed speech waveform in this manner. Fig. 1 illustrates the basic idea where the preprocessed speech is first half-wave rectified, and where parallel diodes then select the maximum of the stored values. The delay time is set so that even at the lowest expected pitch, a full cycle will be represented on the delay line. While this means that there are several cycles on the line in the upper pitch regions (as in Fig. 1), this is not a serious problem because significantly higher pitches tend to have simpler waveforms, being above one or more of the format frequencies, and vary less on a cycle-to-cycle basis.

1.2 Block Diagram

Fig. 2 shows a block diagram of the pitch extractor. The topmost portion consists of four preprocessing sections. First there is a set of input amplifiers to accept various input signal levels. Second there is a gain adjustment section which works to equalize middle amplitudes, reducing amplitudes that are very low or very high. Third there is a filter, which is the single most important part of the whole extractor. It is a voltage-controlled band-pass filter, but is generally used as a fixed band-pass filter (typically with a center frequency of 130 Hz and a Q of 3 for a male talker). The fourth element is a half-wave rectifier which is used here mainly to take better advantage of the available dynamic range of the delay line that follows.

The tapped delay line in the middle of the diagram is the peak amplitude detector as described briefly above. A comparator below the delay line serves to detect the actual occurrence of a peak feature to within a threshold voltage. The bottom of the diagram shows the f/v converter and the output control logic which makes a voiced/unvoiced decision, and outputs the pitch for a voiced input.

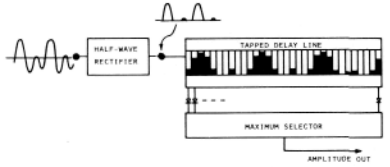

Fig. 1. Fast amplitude extractor based on tapped delay line.

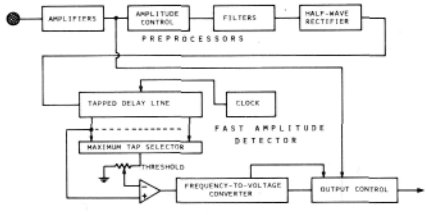

Fig. 2. Block diagram of pitch extractor.

2 CIRCUITRY OF PITCH EXTRACTOR

2.1 Preprocessing Unit

Fig. 3 shows the circuitry of the preprocessing unit. The input amplifiers at the top and the half-wave rectifier at the bottom are well-known circuits and need no further discussion. The filter is a standard transconductance-controlled state-variable filter [10] used as a bandpass with peak response independent of Q. Voltage control is used here mainly as a way of avoiding the use of a dual pot, but also because some users may wish to experiment with feedback arrangements to change the filter frequency slightly when an initial pitch readout is obtained. The automatic gain adjustment unit is neither a true automatic gain control nor a compressor. The actual output amplitude versus input amplitude curve for the unit is shown in Fig. 4, and it can be seen

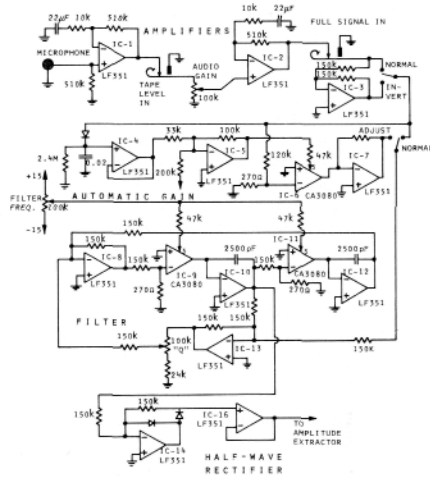

Fig. 3. Preprocessors of pitch extractor.

that it is the middle amplitudes that are favored, helping the overall unit to ignore low background noise, and to avoid clipping for high levels.

2.2 Peak Amplitude Detector Circuitry

The circuitry of the peak amplitude detector is shown in Fig. 5. A simple two-phase clock for the delay line is formed around timer IC-17 and flip-flop IC-18, and these clock the tapped delay line IC-19. IC-56 is the comparator (here actually a Schmitt trigger) indicated in Fig. 2. The remainder of the circuitry, IC-20 through IC-55, serves to accommodate the needs of the delay line, and could perhaps be simplified if a more suitable charge-transfer device for this application becomes available.

Fig. 4. Response of automatic gain adjustment section.

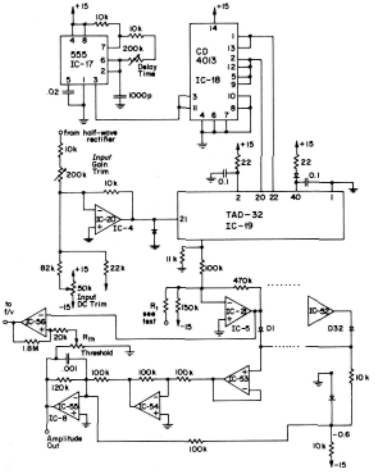

Fig. 5. Peak amplitude detector.

Any actual signal that is to be carried on the TAD-32 delay line must enter through pin 21 and be within a relatively small range of voltage, typically about +6 V with an ac excursion of about ±1 V allowed, with any excessive ac voltage being distorted or clipped. IC-4 serves to adapt the input signal to this requirement, *attenuating it (input gain trim) and adjusting the dc* level (input dc trim). Since the input signal here is unipolar (coming from a half-wave rectifier), it would be a waste of available dynamic range if dc zero were set at the center of the TAD-32 range. Therefore the input dc zero is set at the bottom of the range. Consequently the half-wave rectifier used is redundant since negative signals would be clipped off anyway, but it seems good practice to include it anyway to avoid all possible problems.

The signal, thus properly inputted to the TAD-32, appears at each of the taps with approximately the same dc and ac levels, roughly the same as the input ac and dc levels at pin 21. To properly complete the design, each of these taps must be amplified, level shifted, and buffered by its own operational amplifier, of which IC-21 is typical. The tap numbers and pin numbers are given in Table 1.

It is fairly important that all the taps be equally dc trimmed. This is difficult to do directly because of thermal drift of the dc level on the TAD-32 itself. While this drift is significant, the relative drift between taps is very small. In this application, once the taps are all set, any modest absolute drift is of little importance. The problem is thus one of holding the line steady while *relative trimming is done. One approach is to trim the* first tap to zero and then trim all the other taps relative to the first one by measuring the voltage difference between the taps with a digital multimeter. Another approach is to set the first tap near zero, and connect it to the (−) input of an extra operational amplifier, with the (+) input grounded and the output fed back to the input of the delay line. This extra operational amplifier makes the first tap a virtual ground, and the other taps can then be trimmed to zero dc, after which the extra operational amplifier can be removed.

Dc trimming is done through resistors of which R_1 is

Table 1. TAD-32 tap pincut.

TAP	PIN	TAP	PIN
1	23	2	18
3	24	4	17
5	25	6	16
7	26	8	15
9	27	10	14
11	28	12	13
13	29	14	12
15	30	16	11
17	31	18	10
19	32	20	9
21	33	22	8
23	34	24	7
25	35	26	6
27	36	28	5
29	37	30	4
31	38	32	3

typical. If done with trim pots connected between +15 and −15 V, R_1 should be about 1.5 MΩ. It may be nearly as easy, and much less expensive, to calculate a value of R_1 that will zero the dc, and then solder the free end to +15 or −15 V supplies as needed.

With these adjustments finished, IC-19 through IC-52 can be considered a good approximation to an ideal tapped delay line for the purposes of this application. The delay line then consists of 32 samples of the input waveform. Whichever of these samples (taps) is the largest will pass through its associated diode and back bias all the other diodes. This voltage represents the peak amplitude and is buffered by IC-53, and adjusted slightly by IC-54 and IC-55. A fraction of this voltage is set by threshold control R_{th}, and any tap can be chosen to be compared with this threshold voltage by IC-56. Thus the output of IC-56 changes ideally once per pitch period, and this is fed to the f/v converter. An amplitude readout is available from IC-55 if desired.

3 OUTPUT SECTION

3.1 Frequency-to-Voltage Converter

The frequency-to-voltage converter is shown in Fig. 6 with the upper portion being a period-to-voltage converter and the lower portion an analog divider. The period-to-voltage converter works by sampling, holding, and resetting a ramp voltage each time a pulse comes out of the peak amplitude detector. The frequency-to-voltage converter is described in more detail elsewhere [11] and has no measurable error in the region from 70 Hz through 500 Hz.

3.2 Voiced/Unvoiced Logic Control

The output control section is shown in Fig. 7. It is essentially an analog switch (IC-70) which passes the output of the frequency-to-voltage converter when a voiced decision is arrived at. The voiced/unvoiced logic is as follows. First a signal is determined to be unvoiced if the energy passing through a 4-kHz high-pass filter (IC-67) exceeds a certain threshold (as with fricative wide-band noise, for example). Second, a signal is "possibly voiced" if the ramp voltage of the frequency-to-voltage converter remains below a +5-V level (as for periodic triggering, or for random triggering due to transients). The logic voiced condition is then an AND function (IC-71) of a "not unvoiced" (IC-68) and a "possibly voiced" (IC-69), with these logic inputs defined as above.

Other voiced/unvoiced logic schemes can of course be considered [12] and will work. In fact, there seems to be a degree of latitude in this decision since linguistic researchers are not wholely comfortable with a voiced/unvoiced dichotamy in the first place (preferring that a transition phase between the two be included), and in the second place, they soon learn to mentally discard portions of the readout that they consider to be in error. For speech encoding, small errors in voiced/unvoiced determination have a larger effect on the naturalness of the speech than they do on its intelligibility.

4 OPERATION AND RESULTS

As with any new and experimental device, there is often a tendency to make as many parameters user controllable as possible, with the result that there are so many panel controls that the device is confusing. In the present extractor the indicated switches should be panel features, along with the "audio gain," "filter frequency" and "filter Q" controls from Fig. 3, the "delay time" and "peak amplitude threshold" controls from Fig. 5, and the "unvoiced threshold" control from Fig. 7. A certain amount of calibration and experimentation will be required. Some suggested initial settings of controls are given in Table 2.

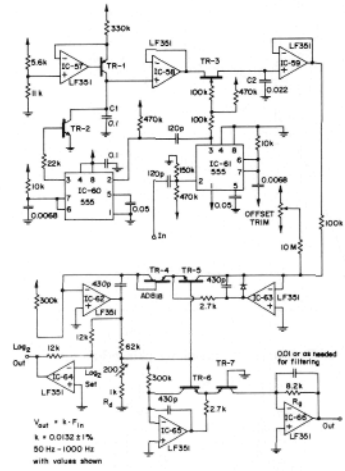

Fig. 6. Frequency-to-voltage converter.

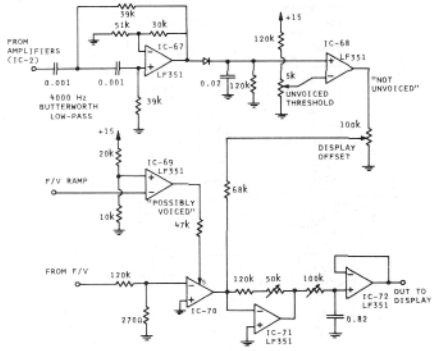

Fig. 7. Voiced/unvoiced logic output and control.

Table 2. Suggested initial settings.

Parameter	Male Talker	Female Talker
Polarity	No preference	No preference
Automatic gain	Auto	Auto
Audio gain	Medium	Medium
Filter frequency	130 Hz	150 Hz
Filter Q	3	1.5
Delay line time	Maximum	3/4 maximum
Peak amplitude threshold	80%	80%
Unvoiced threshold	2/3 maximum	1/3 maximum

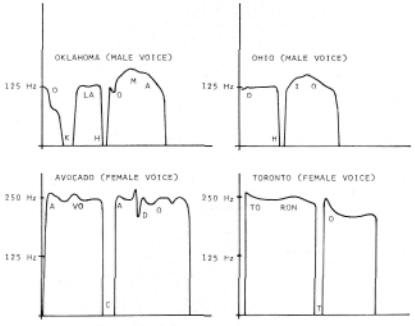

Fig. 8. Example pitch output curves.

A sample pitch contour readout is shown in Fig. 8. In general, fairly good results are obtained for a given user after about 5 min of parameter adjustment. Nearly all users will have to make at least a few adjustments. If the very best results are desired for a given word or phrase, it is best to record this and play it into the extractor using a tape loop. Then the exact effects of the varius controls can be studied under controlled conditions, and the optimum settings can be determined.

5 ACKNOWLEDGMENT

The authors wish to express their thanks to a number of persons at Cornell University who have aided in this project. These include S. Zwolinski and T. Nolan who worked in the School of Electrical Engineering, and J. Grimes, S. Hertz, D. Walter, and J. Gale of the Department of Linguistics and Modern Languages. This work was supported by Rome Air Development Center (RADC), Deputy for Electronic Technology, Hanscom Air Force Base, MA, under Contract F49620-77-C-0069 (Dr. William Ewing, Technical Monitor).

6 REFERENCES

[1] J. N. Maksym, "Real-Time Pitch Extraction by Adaptive Prediction of the Speech Waveform," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 149–154 (1973 June).

[2] S. Seneff, "Real-Time Harmonic Pitch Extractor," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-26, pp. 358–365 (1978 Aug.).

[3] M. M. Sondhi, "New Methods of Pitch Extraction," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, pp. 262–266 (1968 June).

[4] J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-Time Digital Hardware Pitch Extractor," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-24, pp. 2–8 (1976 Feb.).

[5] R. O. Hamm, "Fast Pitch Detection," presented at the 58th Convention of the Audio Engineering Society, New York, 1977 Nov. 4–7, preprint no. 1265.

[6] W. H. Tucker and R. H. T. Bates, "A Pitch Estimation Algorithm for Speech and Music," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-26, pp. 597–604 (1978 Dec.).

[7] B. A. Hutchins, "Pitch Extraction, Part 3: The Complete Experimental Device," *Electronotes*, vol. 7, pp. 3–11 (1975 July).

[8] I. Fritz, "Simple Pitch Extractor for Clarinet," *Electronotes*, vol. 9, pp. 3–7 (1977 Sept.).

[9] D. Wills, "Pitch Extractor for Guitar and Microphone," *Electronotes*, vol. 10, pp. 15–23 (1978 Apr.).

[10] B. A. Hutchins, *Musical Engineer's Handbook* (Electronotes, 1975), chap. 5d.

[11] B. A. Hutchins, "A Frequency-to-Voltage Converter," Electronotes Application Note 114, 1978 Dec. 11.

[12] S. G. Knorr, "Reliable Voiced/Unvoiced Decision," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-27, pp. 263–267 (1979 June).

The biographies of Messrs. Hutchins and Ku were published in the Jan./Feb. issue.