COMPUTER PITCH RECOGNITION:
A NEW APPROACH

Paul E. Dworsak
North Texas State University

Jane Piper Clendinning
Yale University

Presented at the International Computer Music Conference, Rochester, New York, October 9, 1983

173
Abstract

This paper describes a low-cost pitch detector that can be interfaced with a microcomputer. This device and its associated software provide a rapid estimate of the pitch of a human voice or of acoustical instruments that are generating a single pitch. The pitch detector was designed for use in computer-assisted instruction, but in the future it might also be used in music analysis, automated notation of melodies and ethnomusicological research.

The pitch estimation algorithm employed in this application reduces a raw input waveform to a small amount of data. This algorithm uses analog circuitry to detect the zero crossings of the second derivative of the input waveform. These points correspond to the inflection points of the fundamental and the harmonics. The original waveform is then integrated between detected inflection points. Software simulations of the algorithm indicate that it yields correct or nearly correct solutions even in the case when the fundamental is missing.

The circuit described here attempts to emulate some of the operations that the ear performs in recognizing pitch. The circuitry involved is minimal, inexpensive and may be used in a microcomputer environment. Data is reduced significantly by the circuitry so that computation time required by the support software is a small fraction of what it would be if the waveform were processed only digitally. This circuit promises to find many applications in music research in the near future.
COMPUTER PITCH RECOGNITION: A NEW APPROACH

This paper describes the development and operation of a new approach to computerized analysis and identification of pitches. This approach is not merely an adaptation of previous approaches to the problem of computerized pitch analysis. Instead, it uses a mathematically simple and sensible method to emulate what is known of the ear's activity during the hearing process. The hardware implementation of the algorithm described uses inexpensive components and can be interfaced with existing microcomputer systems.

Using Derivatives in Pitch Analysis

As the field of robotics grows, more and more research is being focused on how computers might emulate the sensory capabilities of humans. Some interesting similarities exist in the operation of the different senses. Hildreth points out that the eye-brain system's ability to detect edges depends on the ability of certain retinal cells to respond to changes in light intensity (Hildreth, 1981). In one dimension, the edge point is the inflection point in the mathematical function that describes the intensity of light over space. Hudspeth indicates that, in a similar way, the hair cells of the inner ear react to sound pressure changes (Hudspeth, 1983). The pressure changes occur at the inflection points in the
mathematical function that describes the intensity of sound over time.

Edges and pressure charges can be found in functions of light and sound intensity by calculating their second derivatives. The derivative of a function, of course, is a second function that describes how rapidly the first function changes over time (or space). Figure 1 shows the first derivative of a sine wave. Note that the original function (dotted line) is neither increasing nor decreasing at its highest and lowest points. Note also that at these points (the maxima and minima), the value of the first derivative (solid line) is zero, meaning that no change is occurring. Figure 2 shows the second derivative of the same sine wave. The second derivative has zero points at the inflection points of the original wave. Detecting the zero crossings of the second derivative of a function provides a simple way of finding the inflection points in the original function.

The algorithm proposed in this paper locates the zero crossings of the second derivative of a sound wave to determine the points in the wave at which the ear's hair cells would respond. Figure 3 shows the second derivative of a wave with three harmonics. Note that in such sounds containing a number of harmonics, the number of zero crossings found in any period of the wave may be twice the number of the highest harmonic in the wave. Of course, the goal of a pitch extraction algorithm is to detect a sound's fundamental, just as the ear-brain system successfully distinguishes pitch and timbre.
In vision, the eye-brain system is faced with a similar problem: recognizing shapes when edges of different shapes overlap and when the edges are of different intensities. Hildreth postulates a parallel processing system that can filter intensities to the desired degree so that shapes can be sorted by the eye-brain system. A similar parallel processing system may exist in the ear-brain system, but the algorithm being proposed makes no attempt to model it directly. Instead, following the assumption that the ear-brain system responds more vigorously to greater pressure changes than it does to smaller ones, this algorithm calculates and saves the value of the integral between inflection points in the wave. These integral values, which indirectly represent patterns of intensity in the wave, are sampled and held at their value until a new inflection point is detected in the wave.

In the system under discussion, the values of the integrals are digitized and saved in computer memory. The time delay between integral values also is saved. The integral patterns then are autocorrelated to determine the time interval of repetition in the wave. This technique reduces data significantly, but it does not discard data used by the ear in determining pitch. This algorithm is capable of estimating the pitch of a variety of types of stimuli. However, it does not account for nonlinear aspects of the ear mechanism. These nonlinearities may cause a perceived pitch to differ somewhat from its calculated value.
The model is realized in hardware that combines linear and digital integrated circuits (see Figure 4). Operational amplifiers are used to calculate the first and second derivatives of the input signal. At the same time, the input is integrated until a zero crossing of the second derivative is detected. When this occurs, the output of the integrator is sampled and held, and the integrator then is reset to zero so that the next calculation of the area under the curve may begin.

The sampled output of the integrator is passed to an analog to digital converter, which generates a binary equivalent for the area under the part of the curve being studied at that time. A digital timer concurrently keeps track of the number of microseconds between zero crossings. Both the value of the integral and the amount of time between zero crossings are read by a microcomputer system and are stored in an array. An autocorrelation of the digitized areas in this array determines when the signal begins to repeat itself. Because the array is small, the autocorrelation can be done in software without any significant delay. The autocorrelation also could be done in hardware for true continuous, real-time pitch estimation.

**Circuit Performance**

Figures 5, 6, and 7 show how the circuit responds to tones produced by a female vocalist. The patterns of integral values clearly correspond to the patterns seen in the original waveforms. To further test the circuit, examples of tone
complexes that generate residue pitches also were studied. Schouten (1962) points out that when a subject hears a complex of sine tones that are integral multiples of a missing fundamental frequency, the subject will hear the missing fundamental to be the pitch of the tone, and he will perceive the pitch to have a sharp timbre. If, for example, a subject hears a complex of sine tones consisting of the frequencies 800, 1000, and 1200 Hz, he will perceive its pitch to be 200 Hz. Now, if the tone complex is altered to contain the frequencies 850, 1050, and 1230 Hz, the subject would be expected to hear a pitch at 50 Hz because this is the frequency of which the sine tones are integral multiples. Instead, the subject estimates the pitch to be about 206 Hz. Shifting the frequencies of the component tones of the complex by equal frequency increments shifts the perceived pitch by a smaller, but noticeable, amount in the same direction.

This phenomenon cannot be explained by spectrum analysis methods of pitch detection. However, as Patterson (1973) points out, fine structure theories of pitch estimation also fail to explain the phenomenon. These theories propose that details of the sound wave are used by the ear-brain system to determine pitch. Patterson demonstrates that when the fine structure of such a tone complex is altered by varying the relative phases of its components, the pitch perceived by subjects is not affected. Wightman (1973) attempts to solve this dilemma by postulating that the ear-brain system estimates pitch by an operation equivalent to autocorrelation.
Using autocorrelation, his pitch estimation algorithm correctly predicts the pitches of the shifted-frequency residues.

Figures 8a and 9a show residue-producing sine tone complexes of 800, 1000, 1200, and 1400 Hz, and of 850, 1050, 1250, and 1450 Hz respectively. Figures 8b and 9b show the corresponding integral patterns generated by the circuit. Because the components of these complexes are near or above the upper frequency limit of the current circuit, noticeable phase distortion is evident in the integral patterns obtained. However, in spite of this, a clear repetition rate of 200 Hz is seen in Figure 8b. The repetition pattern in Figure 9b is less obvious. As a result, a software simulation of the circuit was used to generate the integral values for these tone complexes, and these integral values then were autocorrelated. Figures 10 and 11 show the autocorrelation functions for the integrals derived from the waveforms in Figures 8a and 9a respectively. The autocorrelation period in Figure 11 is skewed to the left slightly in relation to the period seen in Figure 10. This skewing suggests that the pitch of the complex containing the frequencies 850, 1050, 1250, and 1450 Hz will be 210 Hz, which is close to the 206 Hz reported in the literature.

Although this circuit has limitations, it successfully detects the pitch of sung tones with frequencies between 150 and 1100 Hz. It also has worked well in detecting the pitch of some speech segments. It therefore can be expected to detect the pitches of instrumental tones and it promises to be an important tool for educators and researchers in the near future.
Figure 1. The first derivative of $y = \sin x$.

Figure 2. The second derivative of $y = \sin x$. 

181
Figure 3. The second derivative of a wave with three harmonics.
Figure 4. Hardware implementation of the algorithm.
Figure 5a. Female voice singing (L)a, 185 Hz.

Figure 5b. Integral patterns at 185 Hz.
Figure 6a. Female voice singing (L)a, 440 Hz.

Figure 6b. Integral patterns at 440 Hz.
Figure 7a. Female voice singing (La), 1045 Hz.

Figure 7b. Integral patterns at 1045 Hz.
Figure 8a. Tone complex consisting of the frequencies 800, 1000, 1200, and 1400 Hz.

Figure 8b. Integral patterns of the above tone complex.
Figure 9a. Tone complex consisting of the frequencies 850, 1050, 1250, and 1450 Hz.

Figure 9b. Integral patterns of the above tone complex.
Figure 11. Autocorrelation of integral patterns in a tone complex consisting of the frequencies 850, 1050, 1250, and 1450 Hz.
REFERENCES


