A Multiple Fundamental Frequency Tracking Algorithm

Özgür Izmirli **Semih Bilgen**

Electrical and Electronics Eng. Dept.

Middle East Technical University, Ankara, Turkey.

E-mail: *izmirli* or *bilgen* *@rorqual.cc.metu.edu.tr*

Abstract : *A fundamental frequency tracking algorithm is presented. A sliding window constant Q transform is used for spectral analysis of the input signal. Fundamentals are found for each window and then filtered in time to eliminate transient effects. The output of the tracker is displayed as notes.*

1. Introduction

Many methods have been proposed for determination of fundamental frequency of monophonic musical signals (see [Brown 1992]). Input signal to midi converters that operate in real time also have been reported [Quiros & Enriquez 1994], [Doval & Rodet 1994]. Still, acoustical input to midi converters can be commercially found. But nevertheless, the problem for polyphonic input remains unsolved. Moorer demonstrated the transcription of notes played by two guitars simultaneously with non-overlapping partials [Moorer 1977]. Source separation together with note identification in polyphonic music using spectral methods have also been reported [Chafe & Jaffe 1986]. The method proposed below, addresses the polyphonic case with the common exceptions. These exceptions are, as commonly encountered in window based analysis applications, the detection of notes with overlapping partials (octaves, octave + fifth etc.).

As far as window based techniques are concerned, reliable results are obtained using the power spectrum of a single window when the window is selected from a suitable portion of the input. Problems arise in detection of the fundamental during on-set or off-set transients and in cases where the previous notes are left to ring (e.g. melody being played on a guitar). In the latter case two or more fundamentals may occur simultaneously. The proposed algorithm detects possible fundamentals that appear within a single window. Elimination of spurious fundamentals due to transient conditions is performed by higher level filtering.

The algorithm is designed to perform in cases where a fundamental is present or not, thus fulfilling the requirement for the missing fundamental phenomenon. Other missing partials are also tolerated inherently by the algorithm. For generality, it is assumed that a single frequency component of the power spectrum is sufficient for it to be considered an individual fundamental. There also is an assumption on the overtone structure of the instruments having an upper limit to the number of partials they can have.

2. Spectrum Analysis

Spectrum calculation is carried out with a constant Q transform [Brown 1991] from the acoustical input that is sampled at 8 Khz. Each window of the input is multiplied by a Hanning window of appropriate length prior to transformation. In contrast to classical FFT techniques, constant Q transforms employ windows of variable size. The current implementation of the transform differs from the original formulation in that there is additional filtering related to frequency tracking and time shift in order to compensate for synchrony detection differences in on-set and off-set of group partials. The recoverable

frequency range is from C2 (65.41 Hz.) to half the sampling rate with 24 bins per octave. Resolution in frequency determination is increased by a method based on phase changes of the constant Q transform [Brown and Puckette 1993]. The time varying spectrum of the input signal is obtained by shifting windows at 7 millisecond intervals. The peaks of the power spectrum are extracted and a power threshold of 1 to 4 percent of the absolute attainable maximum is applied to filter out irrelevant information. We are thus left with a set of frequency components at each time step.

3. Multiple Fundamental Detection

The procedure for multiple fundamental detection is given below for a single window. Higher level filtering of the output generated by this procedure is explained in the next section.

The algorithm takes as input the set (**L**) of frequency components $\mathbf{f_k}$ corresponding to a single time step of the spectrum. **k** takes values from 1 to **N**, **N** being the number of frequency components detected at this step. The frequency components are recorded merely as frequencies disregarding the amplitudes.

The upper limit to the number of partials that are assumed to be associated with a fundamental is fixed at **M**. The number can be adjusted according to the overtone structure of the instruments being used. For example, Piszczalski and Galler [Piszczalski & Galler 1979] use 12 in their algorithm and we have used 8 in our implementation. The assumption to have an upper limit for the number of partials is reasonable for acoustical instruments.

Each frequency component in the set **L**, can be either committed or uncommitted. Initially all frequency components are uncommitted.

We define a function **Y** that is non-zero around a neighbourhood of a particular frequency **h**. The function has its maximum at **h** and is monotone decreasing to either side of it. **Y** can be selected to be a Gaussian [Goldstein 1973], [Doval & Rodet]. Below, a simpler form where **Y** is triangular is given:

$$\mathbf{Y(f,h)} = \begin{cases} \dfrac{\mathbf{f-h}}{\mathbf{h-f_{min}}} + 1 & \mathbf{f_{min} \le f < h} \\ \dfrac{\mathbf{f_{max}-f}}{\mathbf{f_{max}-h}} & \mathbf{h \le f \le f_{max}} \\ 0 & \textbf{otherwise} \end{cases}$$

For generality, $\mathbf{f_{min}}$ and $\mathbf{f_{max}}$ are chosen to be quarter tone in distance each from **h**. The function is used to give a quantitative measure of the strength of contribution in an existence region, with respect to frequency, of a probable fundamental. The function takes a maximum of 1 at the exact subharmonic frequency, **h**, and decreases linearly to 0 as **f** moves away from **h**.

Each subharmonic is further assigned a weight according to its number. Function **O(i)** gives the weight associated with subharmonic **i**. **O(i)** should take decreasing values with increasing **i**. In the current implementation the function shown below was chosen:

$$\mathbf{O(i)} = \frac{3+\mathbf{i}}{4\mathbf{i}}$$

For all uncommitted frequency components **W(f)** is calculated as:

$$\mathbf{W(f)} = \sum_{\mathbf{k}=1}^{\mathbf{N}} \sum_{\mathbf{i}=1}^{\mathbf{M}} \mathbf{Y(f, f_k / i)\ O(i)}$$

where **N** is the number of frequency components in the set **L**.

The frequency, **g** , that maximizes **W(g)** is a potential fundamental and is added to the output list of fundamentals at the current time step.

If any weighted subharmonic of the frequency component $\mathbf{f_k}$ contributes to **W(g)** with $\mathbf{Y(g, f_k / i)\, O(i)} > \varepsilon$, then the component is assigned to be committed. $\boldsymbol{\varepsilon}$ is chosen to be 0.001 in the current implementation.

Iteration continues by finding all fundamental frequency candidates that maximize **W(g)** using the remaining uncommitted components until all frequency components in **L** are committed.

4. Filtering

Candidate notes found for each time step (frame) are further filtered to eliminate spurious notes arising from synchronization differences in the detection of the fundamental and overtones. This is typical for impulse excited instruments where some weak partials disappear and strong partials prevail, thus giving rise to error in single window detection. Two concurrent mechanisms operate for on and off detection. A note is detected as on, when the number of frames in which the same note is detected, exceeds a threshold in a window of ON frames towards the future. The number ON and the threshold together, are determined according to the transient nature of the input instrument. A note is turned off whenever there are more empty frames, for the note under consideration, than the threshold in a window of OFF frames.

5. Results

Three examples are included to demonstrate the operation of the algorithm. The first example, given in Figure 1. is a wavetable generated clarinet tone. Each line above the reference line shows the presence of a detected note with respect to time. The line below the reference line shows the expected idealized starting times and durations for the played notes. The on and off window durations are held at 49 mseconds. Both thresholds are kept at 80 percent of ON and OFF frames for all examples. The well-defined and stable nature of the input results in prompt and correct detection.

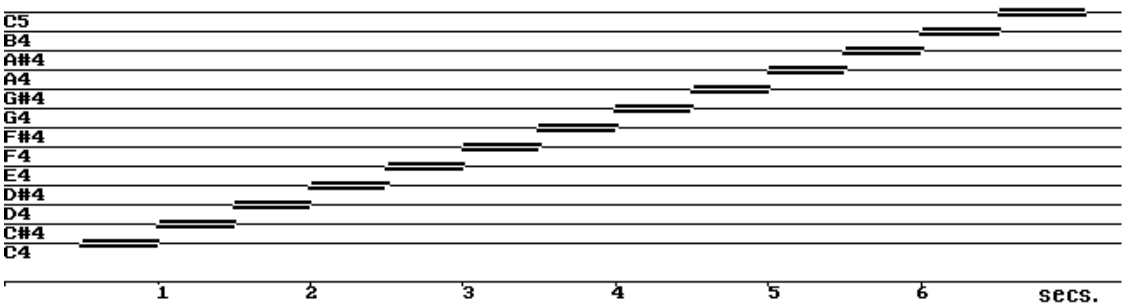

Figure 1. Notes detected from a synthetic clarinet tone. The lines above the reference line indicate the detected notes and those below are those that are expected.

The next example is a male voice singing an F major scale up to the fifth and down. The on window duration is 161 mseconds and the off window duration is 70 mseconds. Longer time is required for the voice to stabilize to a specific frequency. The results are displayed in Figure 2. Note that there are silences between detected notes where the unsettled fundamentals are filtered out.

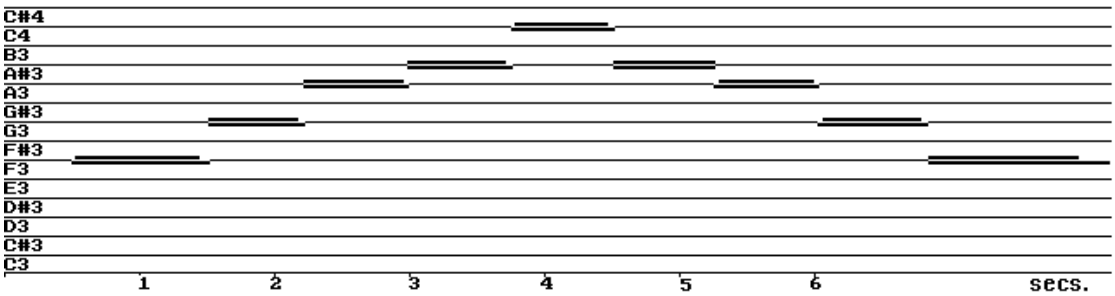

Figure 2. Notes detected from a male voice. The lines above the reference line indicate the detected notes and those below are the expected.

The last example is taken from a record, an initial part from Duo number 1 of 44 Duos (for violin) by B. Bartok. The results can be seen in Figure 3. Detection of multiple fundamentals can be observed clearly and compared to the expected notes adapted from the score. Naturally, octaves in a few places are not detected (e.g. the ending E5 and F#5 at 12 seconds). The window durations are the same as those used in the previous example. Starting times of expected notes have been determined by inspection of the time waveform and their durations have been kept as ideal durations as given in the score.

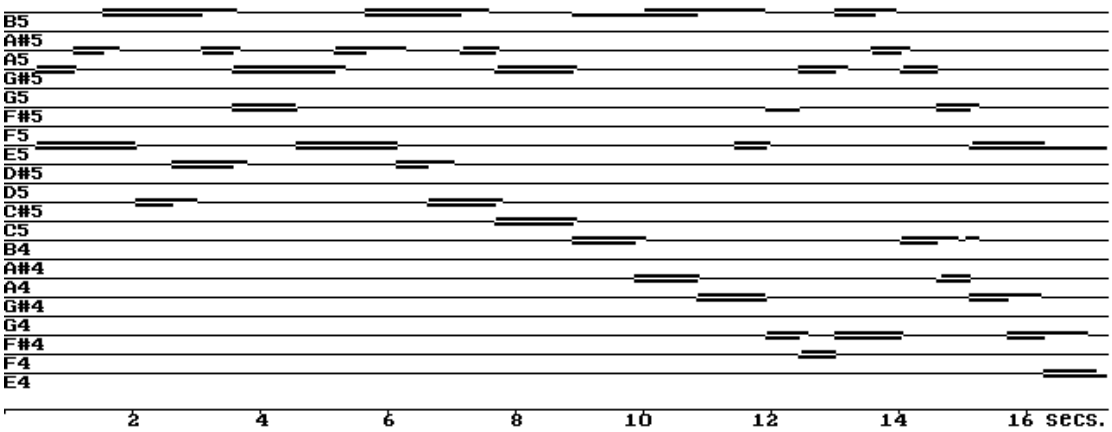

Figure 3. Notes detected from a recording of a violin duo. The lines above the reference line indicate the detected notes and those below are the expected.

6. Conclusion

The presented algorithm addresses the problem of multiple fundamental detection with the mentioned limitations. Although understood as monophonic during listening, pieces may have notes that hang into the time interval of other notes due to various reasons. Fundamental frequency detection algorithms that can handle this and similar cases will be more robust. The presented algorithm handles one or more fundamental frequencies at a time. Thus, polyphonic inputs with non-overlapping partials are completely resolved. This is an advantage over other methods which require strictly monophonic inputs.

For the algorithm, in its given form, there are no assumptions as to what the time-varying spectral characteristics of the input instrument are, or whether they change in time or not. The more strictly the time-varying spectral characteristics of the input instrument can be defined, the better the recognition will be. For example, if it is guaranteed that an instrument will have more than one frequency component for all of its notes, then the algorithm can be modified such that only those fundamentals with at least one overtone are detected. The algorithm can track faster passages as the spectral characteristics become well defined.

References

[Brown 1991] J.C. Brown, "Calculation of a Constant Q Spectral Transform", Journal of Acoustical Society of America, 89(1), pp. 425-434, 1991.

[Brown 1992] J.C. Brown, "Musical Fundamental Frequency Tracking Using a Pattern Recognition Method", Journal of Acoustical Society of America, 92(3), pp. 1394-1402, 1992.

[Brown & Puckette 1993] J.C. Brown and M.S. Puckette, "A High Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform", Journal of Acoustical Society of America, 94(2), pp. 662-667, 1993.

[Chafe & Jaffe 1986] C. Chafe and D. Jaffe, “Source Separation and Note Identification in Polyphonic Music”, Proc. ICASSP, Tokyo, 1986.

[Doval & Rodet 1991] B. Doval and X. Rodet, "Estimation of Fundamental Frequency of Musical Sound Signals", Proc. ICASSP, Vol. 5, pp. 3657-3660, 1991.

[Goldstein 1973] J. Goldstein, "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones", Journal of Acoustical Society of America, 54, pp. 1496-1516, 1973.

[[Moorer 1977] J. Moorer, “On the Transcription of Musical Sound by Computer”, Computer Music Journal, 1(4): pp. 32-38, 1977.

[Quiros & Enriquez 1994], F.J.C. Quiros and P.F. Enriquez, "Real-Time, Loose-Harmonic Matching Fundamental Frequency Estimation for Musical Signals", Proc. ICASSP, Vol. 2, pp. 221-224, 1994.