Periodicity Analysis using a Harmonic Matching Method and Bandwise Processing

Nicolas Durand, Emilia Gómez

Music Technology Group, Pompeu Fabra University

{nicolas.durand, emilia.gomez}@iua.upf.es, http://www.iua.upf.es/mtg

Abstract

In this paper, two methods for fundamental frequency estimation are compared in order to study the advantages of using a bandwise processing for periodicity analysis. The first one is a harmonic matching algorithm that tries to match the peaks of the magnitude spectrum to a harmonic series. The second one splits the signal in separate frequency bands and computes an estimate for each band. Finally, the result is combined to obtain a global estimate.

1 Introduction

Pitch detection has always been an important field of research in the scope of speech and audio processing. There are a hundred of different methods that have been proposed and that work well for different types of sounds in different conditions.

The goal of this paper is to study the advantages of using a bandwise processing in a fundamental frequency estimator.

To do so, two algorithms for fundamental frequency estimation will be compared. The first one is a harmonic matching method that deals with a single band, and the second one processes separately different frequency bands.

2 Harmonic Matching Method

The Two-Way mismatch algorithm is a method that tries to find the harmonic series that best corresponds to the spectral peaks. This algorithm is presented at [5] and has been adapted to the SMS context (see [1]).

Once the peaks of the magnitude spectrum are identified, they can be compared to the predicted harmonics for each of the possible candidate note frequencies, and a measure to fit can be developed. A particular fitness measure is described in [5] as a Two-Way Mismatch procedure.

For each fundamental frequency candidate, mismatches between the harmonics generated and the measured partials frequencies are averaged over a fixed subset of the available partials. A weighting scheme is used to make the procedure robust to the presence of noise or absence of certain partials in the spectral data. The discrepancy between the measured and predicted sequences of harmonic partials is referred as the *mismatch error*. The harmonics and partials would "live up" for fundamental frequencies that are one or more octaves above and below the actual fundamental; thus even in the ideal case, some ambiguity occurs. In real situations, where noise and measurement uncertainty are present, the mismatch error will never be exactly zero.

The solution presented is to employ two mismatch error calculations. The first one is based on the frequency difference between each partial in the measured sequence and its nearest neighbor in the predicted sequence (see figure 1). The second is based on the mismatch between each harmonic in the predicted sequence and its nearest partial neighbor in the measured sequence.

This two-way mismatch helps avoid octave errors by applying a penalty for partials that are present in the measured data but are not predicted, and also for partials whose presence in the measured data is predicted but do not actually appear in the measured sequence. The TWM procedure has also the benefit that the effect of any spurious components or partial missing from the measurement can be counteracted by the presence of uncorrupted partials in the same frame.
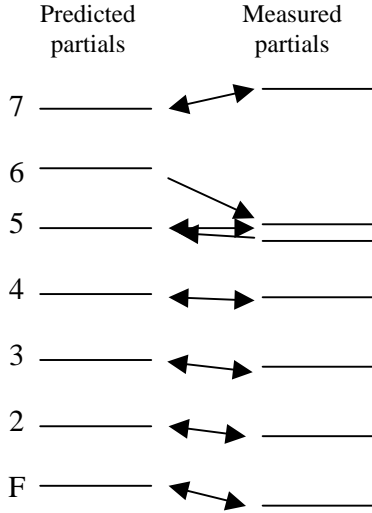
Figure 1: TWM procedure

The two error measurements are computed as following:

- Predicted-to-measured mismatch error

$$Err_{p \to m} = \sum_{n=1}^{N} E_w(\Delta f_n, f_n, a_n, A_{max}) \quad (1)$$

$$= \sum_{n=1}^{N} \Delta f_n \cdot (f_n)^{-p} + (\frac{a_n}{A_{max}}) \times [q\Delta f_n \cdot (f_n)^{-p} - r]$$

where a_n, f_n correspond to the amplitude and frequency of the predicted partial number n, A_{max} is the maximum amplitude, and Δf_n is the difference between the frequency of the predicted partial and its closest measured partial.

- Measured-to-predicted mismatch error

$$Err_{m \to p} = \sum_{k=1}^{K} E_w(\Delta f_k, f_k, a_k, A_{max}) \quad (2)$$

$$= \sum_{k=1}^{K} \Delta f_k \cdot (f_k)^{-p} + (\frac{a_k}{A_{max}}) \times [q\Delta f_k \cdot (f_k)^{-p} - r]$$

where a_k, f_k correspond to the amplitude and frequency of the measured partial number k, A_{max} is the maximum amplitude, and Δf_k is the difference between the frequency of the measured partial and its closest predicted partial.

The total error for the predicted fundamental frequency is then given by a combination of both errors:

$$Err_{total} = Err_{p \to m} / N + r \cdot Err_{m \to p} / K \quad (3)$$

The different parameters of the algorithm are set empirically.

This is the method used in the context of SMS (see [1]) including some improvements, as having pitch dependent analysis window, a selection of spectral peaks to be used, and an optimisation in the search for fundamental frequency candidates. In this algorithm the whole spectrum is processed at the same time, having as an input of the algorithm the collection of detected peaks from the magnitude spectrum.

3 Bandwise processing algorithm

Klapuri [3] proposed an algorithm for periodicity analysis that calculates independent fundamental frequencies estimates at separate frequency bands. Then, these values are combined to yield a global estimate. This solves several problems, one of which is inharmonicity. In inharmonic sounds, as stretched strings, the higher harmonics may deviate from their expected spectral positions, and even the intervals between them are not constant. However, according to the equation (4), we can assume the spectral intervals to be piece-wise constant at narrow enough bands, and increasing function of the center of the considered band.

$$f_n = nf_0\sqrt{1 + bn^2} \quad (4)$$

where β is the inharmonicity factor, which value $\beta \in [0, 0.0008]$.

Thus we utilize spectral intervals to calculate pitch likelihoods at separate frequency bands, and then combine the results in a manner that takes the inharmonicity into account. Another advantage of bandwise processing is that it provides robustness in the case of badly corrupted signals, where only a fragment of the whole frequency range is good enough to be used.

A single fast Fourier transform is needed, after which local regions of the spectrum are separately processed. Before the bandwise processing, the spectrum is equalized in order to remove both additive and convolutive noise simultaneously as explained at [3] and seen at equation (5). This method is based on the RASTA spectral processing [2].

First, a transformation is applied to the magnitude spectrum. This transformation makes additive noise go through a linear-like transformation while the harmonic spectrum go through a log-like transform.

Then, a moving average is subtracted in order to eliminate convolutive noise.

$$X_e(k) = \log[1 + J * X(k)] - X_{av}(k) \quad (5)$$

The equalized spectrum is processed in 18 logarithmically distributed bands that extend from 50Hz to 6000Hz. Each band comprises a 2/3-octave wide region of the spectrum that is subject to weighting with a triangular window. Overlap between adjacent bands is 50%, which makes them sum unity when the windowing gets into account. Fundamental frequency prominence vectors are calculated at each band as explained at [4] according to the following equation:

$$L_B(n) = \max_{m \in M} \left\{ W(H) \sum_{h=0}^{H-1} X_e(k_B + m + hn) \right\}$$
$$m \in M = \{0, 1, ..., n-1\} \quad (6)$$
$$H = [(K_B - m)/n]$$
$$W(H) = 0.75/H + 0.25$$

Finally, the likelihood values are combined getting into account that fundamental frequency can increase as a function of the band center frequency for string instruments. Some improvements were made to provide robustness in interference, where pitch is observable only at a limited band, and to adapt the algorithm to signals containing a mixture of several harmonic sounds.

This filter bench is much alike to the human hearing system's filtering properties. Ear acts like an analyzer composed of a set of "continuous" band pass filters. The bandwidth of a noise affects the loudness of the sound, allowing the definition of a critical bandwidth, function of the band center. Critical bandwidth is usually between 1/6 and 1/3 octave.

We use here a discrete bench of 18 triangular filters whose efficient bandwidth is 1/3 octave, covering the 50Hz – 6400 Hz range (see figure 2). Furthermore, we normalize the filters in energy, so that the filter bench remains coherent when applied to a power spectrum.
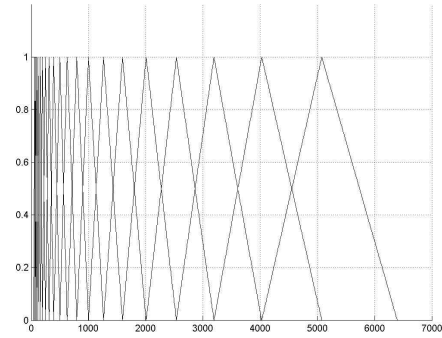

Figure 2: Bark bands triangular filter bench

4 Algorithm comparison

For isolated notes, we look at the output of both algorithms in order to decide if the correct pitch has been detected. As the fundamental frequency estimation is performed frame by frame, the algorithm performance at transitions tracking is not relevant.

For polyphonic sounds, we can in the same way judge if the predominant pitch has been correctly found.

We dealt with sounds of different natures:

• Quasi-harmonic sounds as wind instruments (saxophone, trumpet, clarinet, etc).

• Sounds whose harmonics are not equally spaced, presenting a small inharmonicity factor as string instruments (piano, guitar).

• Sounds with a strong inharmonicity as bells notes.

• Noisy sounds, in order to analyze the behavior against noise.

• Spectrums where a frequency band is filtered in order to measure the bandwise robustness.

• Polyphonic sounds: in this case, we try to extract the predominant fundamental frequency, i.e., the frequency that presents the clearest harmonics. The advantages of using a band-wise processing are also evaluated when dealing with polyphonic sounds, because the predominant frequency may be clear only in a small frequency band.

In order to study the behavior for a wide frequency range, sounds of low and high pitch have been used.

5 Results

For harmonic sounds, both algorithms performances are similar, as can be seen at the following figure.
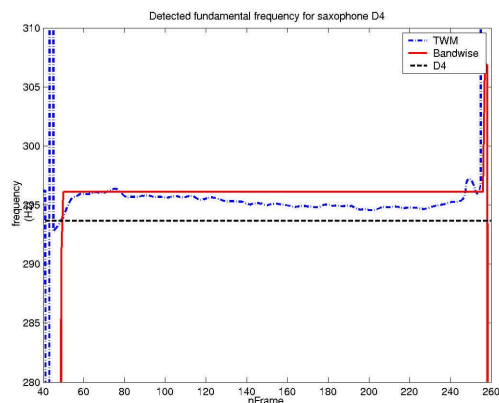

Figure 3: Detected fundamental frequency for saxophone D4 Note

The main difference between both methods is the frequency resolution. The TWM algorithm gets a better spectral resolution by interpolation of the magnitude spectrum when detecting spectral peaks. The spectral resolution could be also decreased using zero-padding.

For sounds whose harmonics are not equally spaced, some differences are found. First, the bandwise algorithm gets into account inharmonicity, which avoids a number of false pitch estimations, particularly if a part of the spectrum is damaged (overnoised or erased). And the algorithm estimates the inharmonicity factor, which is useful for multipitch detection.
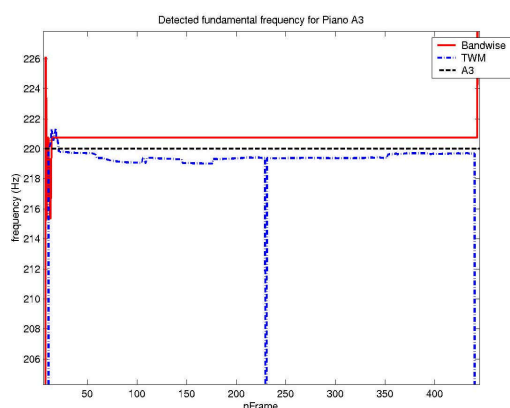

Figure 4: Detected fundamental frequency for piano A3

For noisy sounds, it appears that the bandwise algorithm is much more efficient, but this is mainly thanks to RASTA preprocessing which efficiently cleans the spectrums. Fundamental frequency is thus tracked nearly as long as the isolated note is clearly hearable in the original signal.
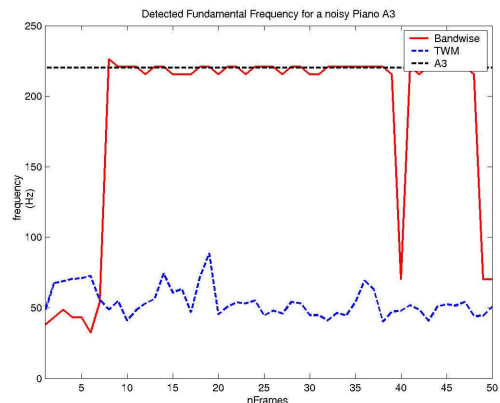

Figure 5: Detected fundamental frequency for noisy piano A3 note

Polyphonic sounds allow a comparison of the algorithms performance in another kind of noisy environment (mix containing many instruments). In this kind of environment, the TWM algorithm proved weaker than the bandwise processing.

Furthermore, bandwise processing aims towards multipitch estimation (MPE) explained at [4]. For each frame, the general model we use extracts one fundamental frequency and its associated inharmonicity coefficient at a time. This coefficient contains information about the locations of harmonics, making the building of a one-note spectrum more reliable. Before subtracting it to the equalized spectrum, we apply to the harmonics heights a smoothing ("smooth+min") in order to leave a part of the partials coinciding with other notes'.

After subtraction, the pitch detection algorithm may look for a new pitch in the same frame. The efficiency in MPE depends mainly on this one-note spectrum subtraction process. Therefore, it proved working well with mixtures of isolated notes whose harmonics are clear enough (violin, for example).

6 Conclusions and perspectives

The advantages of using bandwise processing for periodicity analysis have been tested. The main differences between both algorithms performances can be observed when the harmonicity is found in a particular frequency band, as for example filtered and polyphonic sounds.

It has also been proved that the equalization performed by the bandwise processing algorithm make this method more robust to the influence of noise.

As a perspective, we could try to test if the equalization is also valid as a general preprocessing method for the TWM algorithm, and we could think of applying some kind of post-processing to eliminate isolated errors and abrupt transitions between consecutive frames.

Another possibility is to apply a harmonic matching method to separated frequency bands, instead of computing frequency likelihoods or prominence vectors for each single frequency. This would imply an optimization of the computation charge of the algorithm.

Further developments may be done to improve multipitch estimation. In fact, we observed that, although it is a good modelisation, the use of the inharmonicity factor is not always precise enough to locate rightly the harmonics of a note. We actually tried to cross the detection methods. As RASTA preprocessing leads us to working on "denoised" spectrums, it is efficient to pick the peaks in the spectrum (like in TWM algorithm) and match them with the predicted sequence of harmonics to obtain an efficient reconstitution of a one-note spectrum before subtraction.

7 Acknowledgments

References

[1] Cano, P, "Fundamental Frequency Estimation in the SMS Analysis", DAFX,1998.

[2] Hermansky, H, Morgan, N, and Hirsch, H. G, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing.", ICASSP,1993.

[3] Klapuri, A, "Qualitative and quantitative aspects in the design of periodicity estimation algorithms", EUSIPCO,2000.

[4] Klapuri, A, Virtanen, T, and Holm, J. M, "Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals", DAFX,2000.

[5] Maher, R. C and Beauchamp, J. W, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure", Journal of the Acoustic Society of America, Vol. 95, page 2254-2263, 1993.

[6] CUIDADO IST project, http://www.cuidado.mu