PERIODIC SIGNAL MODELING FOR THE OCTAVE PROBLEM IN MUSIC TRANSCRIPTION

*Antony Schutz, Dirk Slock**

EURECOM
Mobile Communication Department
2229 Route des Crêtes BP 193, 06904 Sophia Antipolis Cedex, France
firstname.lastname@eurecom.fr

ABSTRACT

Precise automatic music transcription requires accurate modeling and identification of the spectral content of the audio signal. Whereas a deterministic model in terms of modulated periodic signals allows to distinguish different notes, the presence of multiple notes separated by octaves poses a big problem since they share the same periodicity, and hence completely overlapping spectral content. In this paper we propose the introduction of a spectral model to allow distinction of such mixtures of spectral content at various octaves. Cyclic correlations are estimated at its pitch and decomposed into even and odd parts, corresponding to even and odd harmonics.

Index Terms— Music transcription, Audio Processing, Pitch Detection, Periodic signal extraction

1. INTRODUCTION

Fundamental frequency (f_0) estimation of a periodic signal has been dealt with extensively in the literature. Many methods devoted to this estimation try to extract this information by using a function of time or frequency (ACF [1],[2], AMDF [3], [4], cepstrum [5], spectrum [6],[7] and High Resolution method [8]). However, audio signals are rarely monophonic and several fundamental frequency can be present at the same time. In the research of speech processing [4] and in the context of musical signal analysis (automatic transcription for example), ([9],[10]) multipitch estimation is an important topic. The spectral interference of the overtones of simultaneous notes has been analyzed by various methods, some aiming at detecting a periodicity in the signal [11], in its spectrum [6], or by using a combination of both spectral and temporal

methods [12], [13]. Other research are based on of a bayesian framework [14] or in a perceptually compliant context [13]. For treating periodic signals, the state of the art was limited to the estimation of pure periodic signals with periodicity equal to an integer number of samples [15, 16]. In these references, the authors propose a Maximum Likelihood approach to analyze pure periodic signals. The decomposition of audio signals into periodic features was reconsidered in [17], and was applied for periodic source separation. In [18] the authors have proposed to merge the periodic signal analysis and sinusoidal modeling in order to give more flexibility to the periodic signal analysis and impose more structure on sinusoidal modeling. They have considered periodic signals with non-integer period, global amplitude variation and time warping. Temporal or spectral methods tend to make sub-octave or octave errors respectively and more again when multiple octaves of the same note are present, since they share the same periodicity and hence completely overlapping spectral content. If a note and its octave are played together the even harmonics of the note should been increased by the harmonics of its octave.

Here we depart from the theory of the method based on cyclic correlation analysis, extending it by using the even and odd part of the periodic signature of the signal. In section 3 we apply the method as a pitch determination algorithm on both synthetic and acoustic signal. Then, in section 4 we use it for solving the octave ambiguity problem and compare it to a more sophisticated spectal method and, finally, we conclude the work in section 5.

2. PROPOSED METHOD

2.1. Method

Generally audio signals are defined as a sum of sinusoids with time varying parameters and an additional noise. For an instrumental or a speech signal, the signal is also harmonic with fundamental frequency equal to f_0.

$$s(t) = x(t) + n(t), \tag{1}$$

$$x(t) = \sum_{n=0}^{N-1} A_n(t) \cos(2\pi \frac{f_n(t)}{f_s} + \phi_n(t)) \tag{2}$$

$$f_n(t) = n \ f_0(t) \tag{3}$$

As defined in [19] the periodic signal can be expressed by its *generalized ACF*, which is cyclic and without any phases.

$$r_k^P = r_k * \delta_{k,o,P}, \ \delta_{k,n,P} = \sum_{i=-\infty}^{+\infty} \delta_{k,n+iP} \tag{4}$$

where $*$ denotes the convolution operator; and δ the Kroenecker delta. Its spectral expression is given by:

$$S^P(f) = S(f) \frac{1}{P} \delta_{\frac{1}{P}}(f), \ \delta_{f_0}(f) = \sum_{k=-\infty}^{+\infty} \delta(f - kf_0) \tag{5}$$

If we define $S(f)$ as:

$$S(f) = \sum_{k=0}^{P-1} r_k^P \ e^{-j2\pi fk}, with \ r_k^P \ = r_{P-k}^P \tag{6}$$

$$\tag{7}$$

The spectral envelope of a such periodic signal can be written as:

$$S(f) = r_0 + 2 \sum_{k=1}^{\frac{P}{2}-1} r_k \ \cos(2\pi fk) + \ r_{\frac{P}{2}} \ \cos(2\pi f \frac{P}{2}) \tag{8}$$

We can define the even and odd parts of the cyclic correlation:

$$r_k^P = r_k^{P,e} + r_k^{P,o}, \tag{9}$$

$$r_k^{P,e} = \frac{1}{2} \ (r_k^P \ + r_{k+\frac{P}{2}}^P), \tag{10}$$

$$r_k^{P,o} = \frac{1}{2} \ (r_k^P \ - r_{k+\frac{P}{2}}^P), \tag{11}$$

$$r_{k+\frac{P}{2}}^P = r_{\frac{P}{2}-k}^P \tag{12}$$

The influence on the spectrum is expressed as follow:

$$S(f) = S^e(f) + S^o(f), \tag{13}$$

$$S^e(f) = S(f)[\frac{1}{2}(\frac{1 + e^{-j2\pi f \frac{P}{2}}}{2} + \frac{1 + e^{j2\pi f \frac{P}{2}}}{2})], \tag{14}$$

$$S^e(f) = S(f)(\frac{1}{2} + \frac{1}{2} \cos(2 pi f \frac{P}{2})) = S(f) \ F^e(f), \tag{15}$$

$$S^o(f) = S(f)[\frac{1}{2}(\frac{1 - e^{-j2\pi f \frac{P}{2}}}{2} + \frac{1 - e^{j2\pi f \frac{P}{2}}}{2})], \tag{16}$$

$$S^o(f) = S(f)(\frac{1}{2} - \frac{1}{2} \cos(2 pi f \frac{P}{2})) = S(f) \ F^o(f) \tag{17}$$

Fig. 1 show the frequency selection of the even and odd parts. As the Fourier Transform is done with P points, with
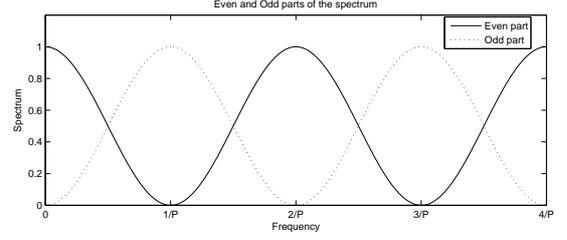

Fig. 1. Even and odd parts of the spectrum.

P the period of the signal, each point of the spectrum is a peak of the periodic signal and the Spectrum represent the spectral envelope. If we define the fundamental frequency as the first harmonic, the even part cancels the odd harmonics and leaves the even harmonics unchanged and vice-versa for the odd part.

2.2. Definition of the periodic signature

The signal is first resampled to a power of two samples for avoiding problem when the even and odd part are computed and for having an integer period.Then the signal is cuted into frames of length P, the periodic signature is expressed by its *generalized ACF* :

$$R^P = IDFT(|DFT(X^P)|^p) \tag{18}$$

where R^P and X^P are two matrices for which each column represent a period of the signal and its cyclic representation respectively:

$$X^P = [x_1 \ldots x_m] \tag{19}$$

$$x_m = [s_{(1+(m-1)P)} \cdots s_{(mP)}]^T \tag{20}$$

Where T denote the transpose operator, m is the number of period in the analysed signal and x is a signal vector containing P samples.

As the harmonics of an audio signal are time varying and non perfectly harmonic, we need to have a robust estimate of the periodic signal. This signature is estimated as the principal vector of the eigen value decomposition of R^P. We define u, the periodic signature, as the first column of $U = SVD(R^P)$.

Then the odd and even parts of the signature are computed:

$$u_k^{P,e} = \frac{1}{2} \ (u_k^P \ + u_{k+\frac{P}{2}}^P), \tag{21}$$

$$u_k^{P,o} = \frac{1}{2} \ (u_k^P \ - u_{k+\frac{P}{2}}^P), \tag{22}$$

$$\tag{23}$$

3. APPLICATION TO PITCH DETECTION

3.1. Discussion

For estimating the pitch of the signal we reduce the set of fundamental frequencies to the first twelve frequencies of the first octave from a midi correspondance. For all of this set we perform the algorithm describe before and choose as candidate the one which maximize an energy criterium. Since the periodic signature is normalized in energy we will work with its even part, but the even part also represents the octave of the pitch so we change the set of candidates to the previous octave. Working with the lower octave candidates didn't reduces the set of octaves to the first one. When a candidate is choosed, we compute the energy of its *Even To Odd Parts Ratio* (*EOR*), if it's more than a threshold we decide that its true octave is the next one and we continue on the next octave by keeping as periodic signature the even part.

Since the energy of the periodic signature is normalised to one, the energy of the Even and Odd Part are bounded to 0.5, the choosed threshold is compared to the Even to Odd Parts Ratio and set to 10.

3.2. Simulation

For this simulation we have generated light inharmonic signals, in fact all the parameters are randomly generated. The Inharmonicity coefficient is set to $B = 10^{-5}$, so the frequencies follows as a rule $f_n = n\, f_0 \sqrt{1 + B\, n^2}$. The amplitudes and phases are uniformly distributed from [0;1] and [0;2π] respectively. The amplitudes are also decreasing with the index and the sum of the amplitudes is normalized to 1. We have choose the tessitura of the guitar for our analyse so the set of midi code is [40;88].

Fig. 2 show the result of the analysis, as expected the notes are correctly interpreted on the octave zero, and their true octaves are correctly found. The second possible candidat is also show for each notes, as we can see for the first and a half octave it has a semitone difference but for the next octave it's a perfect fourth difference (5 semitones upper).

3.3. Application to a true signal

For this analysis we have record all the first 37 notes of the guitar (midicode 40 to 76) on a acoustic guitar. The notes are played with a guitar pick and the guitar was plugged and link to an external soundcard. The analysis is made on the first $250ms$ of the signal (including the attack). Note that the guitar was not perfectly tuned (impossible) and the used candidate are determined again by the midi reference frequency.

Fig. 2 show the result of the analysis for the guitar, the result is not perfect but we can see that if a note is not well detected its octave is false and the note found is the perfect fourth of the played note, the second candidat of the previous
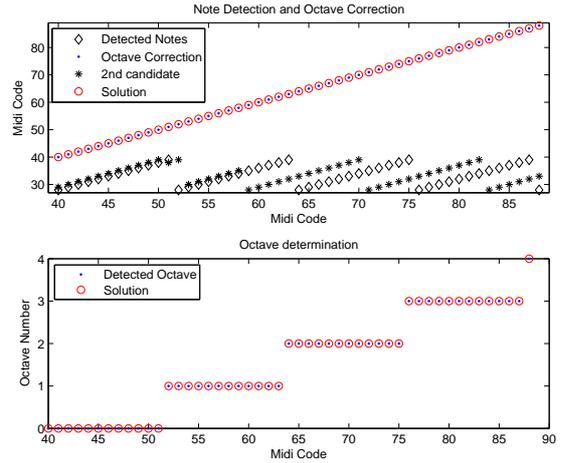

Fig. 2. Pitch detection and Octave Selection for a synthetic signal.
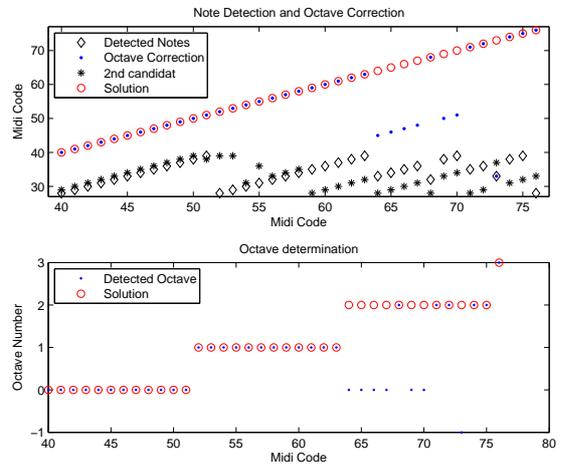

Fig. 3. Pitch detection and Octave Selection for guitar.

analysis, in this case the true note become the second choice. Note that the perfect fourth share some harmonics in the even part but don't share its fundamental frequency.

4. APPLICATION TO THE OCTAVE PROBLEM

In this section we analyse the octave problem. The octave problem appears when a note an its octave are played together. They share the same periodicity and the even harmonics of the played notes are amplified by the harmonics of the Octave. For the analysis we assume that the fundamental frequencies are known. In spectral analysis there is, at least, two way for estimating even and odd frequencies. The first one consist on finding all the peaks in the spectrum, by peak picking, and by paying attention to don't miss some of them otherwise an odd harmonic can become an even harmonic and vice-versa, an-

other point is the inharmonicity of the signal. For finding the peaks we have to adjust, from one peak to the next one, the distance and searching a local maximum around it. The second method is equivalent to the proposed method, it consist on computing the spectra of the matrix X^P, define before, and taking the average trough the time dimension, it's a Welch's periodogram, then the even harmonics are the even samples of the spectrum.

4.1. Note plus its Octave

Here a note is played with and without its octave, recorded in the same condition as before with an acoustic guitar. We compare the results of the proposed method with the first spectral technics (with peak picking). The second spectral method explain before give very similar result than the proposed one (temporal) so we just show our proposed method. The results (Fig. 4) are poor for the two methods due to the coloration of the spectra.
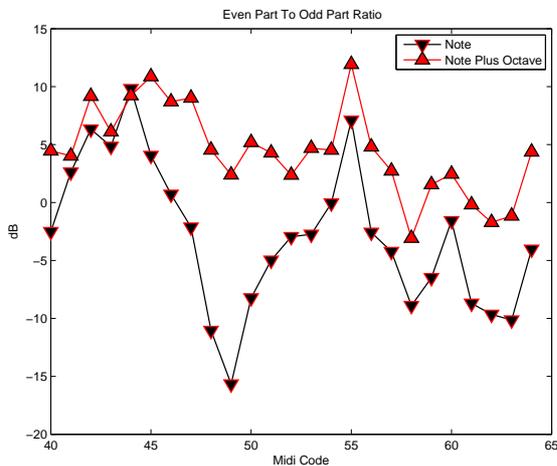

Fig. 4. Octave problem, a note with its octave.

We have decided to add in our framework another one preprocessing, for the rest of the simulation we will work in the prediction error of the signal. The signal is modeled as an autoregressive model of order ten, the prediction error is the residual. And we defined that a note can't be interpreted as its octave but a note with its octave can be interpreted as the note alone.

The results (Fig. 5) are better for the two methods. The dashed line is the upper value of the notes alone, in the two cases we make one error.

4.2. Note plus its first two octaves

In this part the notes are compared to the case where the first two octaves are present simultaneously. The analysis is performed at the fundamental frequency ($f0$), at twice and triple
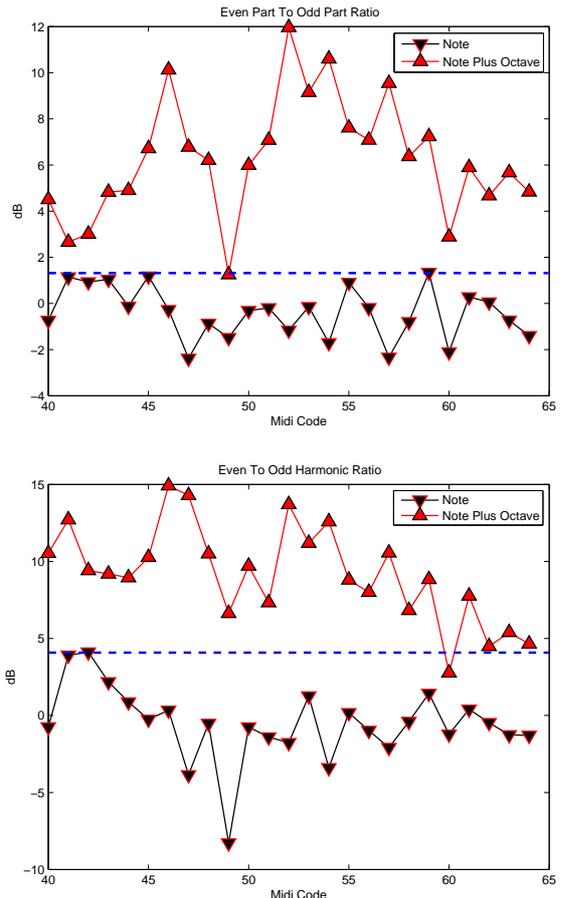

Fig. 5. Octave problem in the prediction error, a note with its octave with the temporal method (top) and the spectral method (bottom).

of the frequency. For a visibility problem we don't show the result for the notes alone and for an evident reason the analysis is done on the first octave (midi code 40 to 52).

The results in Fig. 6 are also good for the two methods. The analysis at the fundamental frequency find the next octave, at the first octave we found the $2nd$ octave and after there is nothing.

4.3. Note plus its second octave

Now we compare the two methods for the case of a note with its second octave (an octave is missing). The second octave influence one harmonic over four from the fourth harmonic, so the result of the analysis sould be slightly similar to the previous analysis. Fig. 7 shows the result, we know which octave is the last one but nothing between the note and the octave, the only possibility for solving this problem is to estimate the envelope of the individual component of the signal.
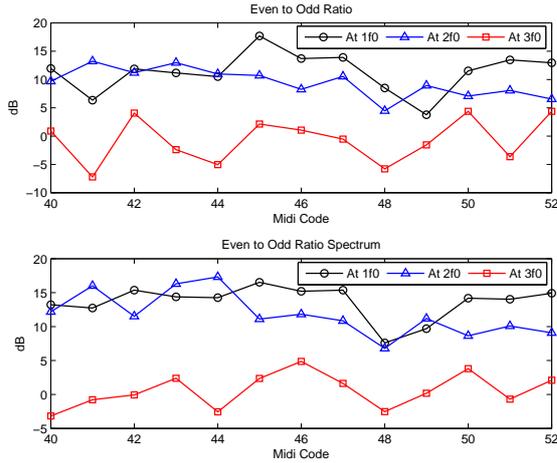
Fig. 6. Octave problem in the prediction error, a note with its first and second octaves. Temporal method (top) and Spectral method (bottom).

Fig. 7. Octave problem in the prediction error, a note with its second octave. Temporal method (top) and Spectral method (bottom).

4.4. Parameters used

The records were performed with a sampling frequency of $44100\ Hz$ with a normal acoustic guitar, the sound card use is a Firebox from Presonus. The period of each analysis is resampled to 512 which allow a significant number of decomposition for the Even and Odd decomposition. The parameter p of the $generalized\ ACF$ is set to 1. The order of the predictor used for the prediction error is 10 and the time duration of each analysis is $250\ ms$.

5. CONCLUSION AND FUTURE WORK

A novel pitch determination algorithm is proposed using the separation of the Even and Odd parts of a cyclic signature of the signal. The ratio of the even and odd parts can determine the octave of the note. Simulations on synthetic and true signal show the potential of the proposed method, which can be improve by adding some constraints on the pitch candidat. A temporal vision for the estimation of the present octave in the signal is proposed, the results are compared to a more optimised method reach similar results. Although the intermediate octave problem is not solved we will extend our algorithm by including the estimation of the spectral envelope.

6. REFERENCES

[1] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 25, pp. 24–33, 1977.

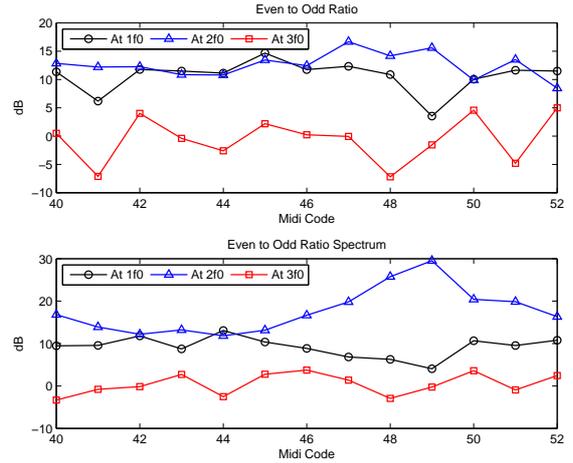[2] R. Meddis and M. J. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periph- ery. i: Pitch identification," *JASA*, vol. 89, pp. 2866– 2882, 1991.

[3] A. C. R. F. M. Ross, H. Shaffer and H. Manley, "Av- erage magnitude difference function pitch extractor," *IEEE Trans. on Acoustics, Speech, and Signal Process- ing*, vol. 22, p. 353362, 1974.

[4] A. de Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *JASA*, vol. 111, p. 19171930, 2002.

[5] A. M. Noll, "Cepstrum pitch determination," *JASA*, vol. 41, pp. 293–309, 1967.

[6] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. on Speech and Audio Processing*, vol. 11, pp. 804–816, 2003.

[7] A. Schutz and D. Slock, "Modele sinusoidale : Estima- tion de la qualite de jeu d'un musicien, detection de cer- tains effets d'interpretation," *Gretsi*, 2007.

[8] B. D. R. Badeau and G. Richard, "High-resolution spec- tral analysis of mixtures of complex exponentials mod- ulated by polynomials," *IEEE Trans. on Signal Process- ing*, 2006.

[9] M. Ryynnen and A. Klapuri, "Polyphonic music tran- scription using note event modeling," *in Proc. of WAS- PAA*, pp. 319–322, 2005.

[10] M. Marolt, "A connectionist approach to automatic tran- scription of polyphonic piano music," *IEEE Trans. on Multimedia*, vol. 6, pp. 439–449, 2005.

[11] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. on Speech and Audio Processing*, vol. 8, p. 708716, 2000.

[12] G. Peeters, "Music pitch representation by periodicity measures based on combined temporal and spectral representations," *in Proc. of ICASSP*, vol. 5, pp. 53–56, 2006.

[13] A. Klapuri, "A perceptually motivated multiple-f0 estimation method," *in Proc. of WASPAA*, pp. 291–294, 2005.

[14] S. G. M. Davy and J. Idie, "Bayesian analysis of polyphonic western tonal music," *JASA,*, vol. 119, p. 24982517, 2006.

[15] D. Muresan and T. Parks, "Orthogonal, exactly periodic supspace decomposition," *IEEE Trans. on Signal Processing,*, vol. 51, 2003.

[16] J. C. J.D. Wise and T. Parks, "Maximum likelihood pitch estimation," *IEEE Trans. on Acoustics, Speech, and Signal Processing,*, vol. 51, pp. 418–421, 1976.

[17] A. de Cheveign and M. Slama, "Acoustic scene analysis based on power decomposition," *In Proc. of IEEE Int. Conf. on Acoustic, Speech, and Signal Processing,*, 2006.

[18] M. Triki and D. Slock, "Periodic signal extraction with global amplitude and phase modulation for music signal decomposition," *In Proc. of IEEE Int. Conf. on Acoustic,Speech, and Signal Processing (ICASSP)*, 2005.

[19] A. Klapuri, "Multipitch analysis of polyphonic music and speech signals using an auditory model," *IEEE Trans.on Speech and Audio Processing*, vol. 16, pp. 255–266, 2008.