Recognition of piano notes with the aid of FRM filters

Foo, Say Wei; Lee, Edwin Wei Thai

2004

Foo, S. W., & Lee, W. T. Recognition of piano notes with the aid of FRM filters. Proceedings of the 3rd International Conference on Sciences of Electronic, Technology of Information and Telecommunications. (pp.409-413). Tunisia: IEEE.

<https://hdl.handle.net/10356/90831>

<https://doi.org/10.1109/ISCCSP.2004.1296315>

International Conference on Sciences of Electronic, Technology of Information and Telecommunications © 2004 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. <http://www.ieee.org/portal/site>.

RECOGNITION OF PIANO NOTES WITH THE AID OF FRM FILTERS

Say Wei Foo

School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore 639798

Edwin, Wei Thai Lee

DSO National Laboratories
20 Science Park Drive
Singapore 118230

ABSTRACT

An approach to recognition of piano notes using the Analysis-by-Synthesis technique is described in this paper. It is assumed that several notes may be played simultaneously and only a single signal source is available. The analysis is carried out in the frequency domain using Constant-Q Transform. In cases where the harmonics of notes are very close to one another, Frequency Response Masking (FRM) filters with suitably chosen selectivity and bandwidth are used to resolve the ambiguity. The method was tested with recording from an acoustic piano played by professional pianists. Results show that the system is able to resolve up to six simultaneous notes with high accuracy.

1. INTRODUCTION

In the jargon of signal processing, piano notes are quasi-periodic. The frequency spectra of piano notes show that for a single note, there is a strong fundamental frequency and up to six prominent harmonics. A single note is readily identified from the fundamental frequency and the harmonics from a short-time frequency spectrum of the signal. However, for signal consisting of several notes played together, the task becomes more complex as there are overlaps of the harmonic contents of the notes.

In the western musical scale, the frequency spacing between adjacent notes is an approximate 6% increase from one note to the next higher one. Mathematically

$$f_{n+1} = 2^{1/12} f_n \quad (1)$$

where f_n and f_{n+1} are the fundamental frequencies of the n^{th} and $(n+1)^{\text{th}}$ notes respectively. The fundamental frequencies of musical pitches are thus not linearly distributed. Hence, in spectral analysis, for efficient computation, instead of using a normal Fourier Transform, a constant resolution, say about 3% of the fundamental frequency over the frequency range of interest may be adopted. The Constant-Q Transform (CQT) [1] is essentially a variation of the Short-Time Fourier Transform (STFT). The constant resolution is achieved by having a variable window size for the STFT.

Since the required resolution is directly proportional to the fundamental frequency, the ratio of frequency to resolution, Q is a constant. For a resolution of 3% of the fundamental frequency, Q is given by

$$Q = \frac{f}{0.03f} = 33.3 \quad (2)$$

A value of 34 is normally used for Q .

The spectral coefficients for STFT, which uses windows of fixed length, are obtained using

$$X[k] = \sum_{n=0}^{N-1} W[n] x[n] e^{-j2\pi kn/N} \quad (3)$$

where $X[k]$ is the k^{th} spectral component of the STFT, $x[n]$ is the n^{th} sample of the signal, $W[n]$ is the window function of length N and $2\pi k/N$ is the digital frequency.

By replacing N in (3) by $N[k_{cq}]$, i.e., windows of variable length, we obtain the spectral components of the CQT as follows.

$$X[k_{cq}] = \sum_{n=0}^{N[k_{cq}]-1} W[k_{cq}, n] x[n] e^{-j2\pi Qn/N[k_{cq}]} \quad (4)$$

where $X[k_{cq}]$ is the k_{cq}^{th} spectral component of the CQT, $x[n]$ is the n^{th} sample of the signal, $W[k_{cq}, n]$ is the window function of length $N[k_{cq}]$ and $2\pi k/N[k_{cq}]$ is the digital frequency. For our analysis, a Hamming window expressed mathematically in (5) below is adopted as the window function.

$$W[k_{cq}, n] = \alpha + (1 - \alpha) \cos\left(\frac{2\pi n}{N[k_{cq}]}\right) \quad (5)$$

where $\alpha = 25/46$ and $0 \leq n \leq N[k_{cq}] - 1$.

An efficient algorithm for implementing CQT may be found in Brown [2].

To illustrate the problem of spectral overlap of harmonics, a piano chord that consists of Notes C3, G3, C4 and E4 is used. The Constant-Q Transform of a segment of the signal, using a semitone resolution, is shown in Fig.1.

The fundamental frequency and harmonics of the four notes are listed in Table I. The positions of the spectral components in the Constant-Q Transform domain are given in Table II.

Fig.1: The Constant Q Transform of Notes C3, G3, C4 and E4 played simultaneously

Table I: Major spectral components of the four notes

Not e	Fundamenta l frequency	1 st harmoni c	2 nd harmoni c	3 rd harmoni c
C3	130.8Hz	261.6Hz	392.4Hz	523.2Hz
G3	196.0Hz	392.0Hz	588.0Hz	784.0Hz
C4	261.6Hz	523.2Hz	784.8Hz	1046.4Hz
E4	329.6Hz	659.2Hz	988.8Hz	1318.4Hz

Table II. Positions k_{eq} in CQT domain

Not e	Fundamenta l frequency	1 st harmoni c	2 nd harmoni c	3 rd harmoni c
C3	6	18	25	34
G3	13	25	32	41
C4	18	30	37	46
E4	22	34	41	50

It can be readily seen from Fig.1 and also from the tables that the third harmonic of Note C4 (1046.4Hz) and the second harmonic of Note E4 (988.8Hz) are fused.

In this paper, a method for recognition of piano notes is described. The method makes use of Analysis-by-Synthesis approach [3] with the aid of Fast Filter Bank (FFB) [4]. For this method, a database of the spectral profiles of all piano notes is first constructed. Details of spectral profiles are given in Section 2. The Analysis-by-Synthesis method is described in Section 3. The design of the Fast Filter Bank is given in Section 4. The resolving power of FRM filters is illustrated in Section 5. Results of experiments are summarized in Section 6 and concluding remarks presented in Section 7.

2. SPECTRAL PROFILES OF NOTES

For the proposed method, a database of the ratios of the amplitudes of the first four harmonics and the fundamental frequency is first established. Mathematically, the following ratios are computed.

$$r_{ij} = \frac{A_i}{A_j} \quad i = 0, 1, 2, 3 \quad j = i + 1 \quad (6)$$

where A_i and A_j are the amplitudes of the i th and the j th harmonics respectively and A_0 denotes the amplitude of the fundamental frequency.

Many samples of the same note are used and the average ratio for each note is determined. Mathematically,

$$\bar{r}_{ij} = \frac{1}{N} \sum_{k=0}^{N-1} r_{ij}(k) \quad (7)$$

where k in the formula denotes the sample number of the N samples of identical notes used. A database of the average values of the ratios of all the notes is then created.

3. ANALYSIS-BY-SYNTHESIS TECHNIQUE

The signal is first filtered at 10 kHz and sampled at 22050 samples per second and coded with 8 bits per sample. Constant-Q Transform is performed on each frame of signal. Analysis of the harmonic contents of the frames is carried out. The first peak of the spectrum is taken as the fundamental of the first note of the chord and its first four harmonics are tracked.

The note is provisionally accepted as a possible note based on the relative position of the harmonics and other possible notes. The next spectral component that has not been accounted for is then tracked. If the corresponding set of harmonics can be found, it is again provisionally accepted as a possible note. If the spectral component appears at an octave of previous provisional notes, higher harmonics are then examined to ascertain if a set of fundamental and harmonics of sufficient amplitude can be found to justify it as a note at a higher octave of one provisionally determined.

Having completed the analysis process, synthesis of spectrum based on the provisional notes identified is then carried out. The spectrum of the notes is reconstructed based on the ratios of the expected spectral profiles of notes from the database. The amplitude of the lowest note obtained from the analysis process is used as the reference level for the first note to be reconstructed. The amplitudes of the first four harmonics are determined based on the average ratios of the notes as from the database. The spectral components of all provisionally identified notes are regenerated in the same way.

The resultant spectrum, which is the sum of the expected spectral profiles, is compared with the spectrum of the signal under test. Any provisional note that is wrongly identified or any note not identified will show up in the comparison. Further analysis may be necessary to

sort out any major difference including the application of Frequency Response Filters.

At times, two different sets of notes with overlapping harmonics may be in anti-phase with each other in the time domain. This results in the subtraction instead of addition of the harmonics. This is taken care of in the synthesis process. When reconstructing the overall profile of the chord, the harmonics of the spectral profiles of the individual notes are either added or subtracted to obtain the various spectral profiles of the combined notes. The optimum profile is then chosen from all the possible sets of profile reconstructed.

The set of notes that gives the closest resemblance to the spectral profile of the test signal will be taken as the actual notes that are played.

4. FAST FILTER BANK

The Fast Filter Bank (FFB) proposed by YC Lim [5], is able to produce a bank of filters with very narrow main lobe, good pass-band response and sharp cutoff by using a high order prototype filter. For a 4096-channel FFB, the bandwidth of the filters is about 5.4 Hz and the first side-lobe is about -60 dB. These characteristics are useful in separating harmonics with low amplitude and/or close together.

The impulse response of prototype filters for the 4096-channel FFB are shown in Table III.

Table III: Impulse Response of Prototype Filters

n	$h_a^1(n)$	$h_a^2(n)$	$h_a^3(n)$	$h_a^4(n)$	$h_a^5(n)$	$h_a^6(n)$
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1, -1	0.62750	0.62090	0.57380	0.56540	0.50130	0.50030
3, -3	-0.18620	-0.16880	-0.07530	-0.06540		
5, -5	0.08780	0.06590				
7, -7	-0.04260	-0.02290				
9, -9	0.01860	0.00550				
11, -11	-0.00670					

n	$h_a^7(n)$	$h_a^8(n)$	$h_a^9(n)$	$h_a^{10}(n)$	$h_a^{11}(n)$	$h_a^{12}(n)$
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1, -1	0.50010	0.50000	0.50000	0.50000	0.50000	0.50000

The frequency response of Channel 197 is shown in Fig. 2(a) with an expanded view of the main lobe shown in Fig.2(b).

Fig.2 (a)

(b)

Fig.2: Frequency response of Channel 197 of a 4096 channel FFB (a) showing main lobe and side lobes, (b) expanded view of the main lobe.

5. RESOLVING POWER OF FRM FILTERS

With the application of the FFB, the two harmonics mentioned in Section 1 and shown in Fig.1 can be resolved as outputs of Channel 186 and Channel 197. This is illustrated in the spectral domain in Fig.3. Channel 186 is the band containing the 2nd harmonic of Note E4 and Channel 197 is the band containing the 3rd harmonic of C4.

Fig.3: Output of the Fast Filter Bank

The filtered signals from the two channels are depicted in the time domain in Fig.4(a) and Fig.4(b).

From Fig.4, the attack (A), delay (D), sustain (S) and release (R) (ADSR) profiles of the output signals of Channels 186 and 197 that characterize musical notes [6]

can be identified. This serves to confirm that the two channels contain harmonics of the notes.

Fig.4(a)

Fig.4(b)

Fig.4: Outputs showing amplitude variation of signal with time.
Fig.4(a) Channel 186, Fig.4(b) Channel 197.

6. EXPERIMENTAL RESULTS

The method was tested with notes from an acoustic piano played by professional pianists. A set of single semi-quaver notes with very short duration of about 0.08s from "Partita V" by J.S. Bach was selected and tested. An accuracy of 100% was achieved.

For polyphonic signals, multiple notes taken from the piano pieces "In Church", "Partita V" and "Bagatelle In A Minor" were used. Notes that were taken from these pieces of music consist of 2 chords of six notes unison, 2 chords of 5 notes unison, 21 chords of 4 notes unison and 3 chords of 3 notes unison.

The results of the sets of musical notes tested, without and with the use of FFB are shown in Table IV and V respectively.

Table IV: Results of test without FFB

	Present	Recognized	Omission	False Notes
6 members chord	12 notes	10 notes	2 notes	NIL
5 members chord	10 notes	7 notes	3 notes	NIL
4 members chord	84 notes	81 notes	3 Notes	NIL
3 members chord	9 notes	ALL	NIL	NIL

Table V: Results of test with FFB

	Present	Recognized	Omission	False Notes
6 members chord	12 notes	ALL	NIL	NIL
5 members chord	10 notes	8 notes	2 notes	NIL
4 members chord	84 notes	ALL	NIL	NIL

3 members chord	9 chords	ALL	NIL	NIL
-----------------	----------	-----	-----	-----

Results show that of the 115 notes tested, 107 notes are correctly recognized using Analysis-by-Synthesis technique alone that is an accuracy of 93%. Chords with 3 notes can readily be identified.

With the aid of FFB, additional 6 notes are recognized, perfect recognition is achieved for chords with four notes and below. An overall accuracy of about 98% can be achieved using Analysis-by-Synthesis technique with the aid of FFB, an increase in the recognition accuracy by 5%.

7. CONCLUSION

It is shown that the system that makes use of Analysis-by-Synthesis (top down analysis to assess the possible notes in the musical signal followed by a bottom up reconstruction algorithm based on the spectral profiles of the notes identified) is a viable method for the recognition of notes from an acoustic piano. The performance of the method is further enhanced using FRM filters to resolve very closely spaced harmonics of musical notes. FRM filters have very high selectivity of the frequency bands, are most suitable to resolve any difference between the reconstructed signal and the original signal.

Experimental results show that the Analysis-by-Synthesis approach together with the application of FRM filters is able to recognize up to six simultaneously played notes with very high accuracy.

Although the Constant-Q Transform can be combined with FFB to achieve a Constant-Q Transform with high selectivity, it may be computationally expensive to use such Transform entirely as the basis for analysis. As experimental results indicate, the accuracy of the Analysis-by-Synthesis approach without FFB is already high. Only in cases where ambiguity arises and high selectivity filters are required, FFB will be applied. In this way, the overall computational load is reduced.

8. REFERENCES

- [1] Brown, Juidth C., "Calculations of a Constant Q Spectral Transform", J. Acoust. Soc. Am89, pp425-434 (1990)
- [2] Brown, Juidth C., "An Efficient Algorithm for the Calculation of a Constant Q Transform", J. Acoust. Soc. Am 95, pp2698-2710 (1992)
- [3] Y.C. Lim, "Frequency-Response Masking Approach For The Synthesis Of Sharp Linear Phase Digital Filter" IEEE Transactions On Circuit And Systems, Vol. Cas-33, No. 4, pp357-364, April 1986
- [4] Y.C. Lim, B. Farhang-Boroujeny, "Fast Filter Bank (FFB)", IEEE Transactions On Circuit And Systems-II Analogue And Digital Signal Processing, Vol. 39, No. 5, pp316-318, May 1992
- [5] B. Farhang-Boroujeny and Y.C. Lim "A comment on the Computational Complexity of Sliding FFT", IEEE

Transactions On Circuits And Systems-II Analogue And Digital Signal Processing, Vol. 39, No. 12, pp875-876 Dec 1992

- [6] James H. McClellan, Ronal W. Schafer "DSP FIRST A MULTIMEDIA APPROACH" Prentice-Hall 1998, pp 435-455.
- [7] Edwin Lee and Say Wei Foo, "An Innovative Approach To Transcription of Polyphonic Signals", 3rd IEEE International Conference on Information, Communication and Signal Processing (ICICS 2001), October 2001,P0093