Title	Transcription of polyphonic signals using fast filter bank(Accepted version)
Author(s)	Foo, Say Wei; Lee, Edwin Wei Thai
Citation	Foo, S. W., & Lee, E. W. T. (2002). Transcription of polyphonic signals using fast filter bank. IEEE International Symposium on Circuits and Systems 2002. (pp. 241-244).
Date	2002
URL	http://hdl.handle.net/10220/4616
Rights	IEEE International Symposium on Circuits and Systems. Journal can be found at http://www.ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=7897

TRANSCRIPTION OF POLYPHONIC SIGNALS USING FAST FILTER BANK

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

In this paper, a novel approach to the transcription of polyphonic signals is investigated. The method makes use of top-down analysis and bottom-up reconstruction of the time-frequency contents of the signals. Time-frequency contents of the signals are obtained using a bank of narrow band filters with very sharp transition band. The filters are designed using the Frequency Response Masking technique. The high selectivity of the filters has enhanced the top-down analysis and increased the accuracy of identification over the simple frequency transform techniques. The method was tested with musical notes generated from an acoustic piano. Results show that for single notes and chords up to four notes, perfect identification can be achieved.

1. INTRODUCTION

The transcription of musical notes normally begins with an analysis of the harmonic contents of the signals. Discrete Fourier transform and Constant-Q transform (CQT) [1],[2] are some of the methods used to extract the spectral contents of musical notes. However, these methods have some drawbacks. It is at times difficult to identify spectral components, especially harmonics that are low in amplitude or close together. This poses problems for accurate analysis of multiple notes played at the same time especially notes of very short duration.

In this paper, a new method to enhance the accuracy of musical notes identification is proposed. The method makes use of a top-down analysis and bottom up reconstruction algorithm [3] and time-frequency contents obtained from Fast Filter Bank (FFB) generated using the Frequency Response Masking technique [4].

The top down analysis uses a harmonics-tracking algorithm to identify the set of possible notes that are played. The bottom up reconstruction algorithm is a process that combines the spectral contents of the possible notes identified and compares the reconstructed spectral profile with the actual spectral contents. The spectral profiles of the notes may be obtained for a particular instrument or for the type of instrument in general on a prior basis.

The spectral contents for the analysis are obtained using FFB. FFB has narrow transition width and flat pass band response. These characteristics are useful in separating harmonics with low amplitude and/or close together.

The remaining part of the paper is organized as follows. In Section 2, the spectral profiles of notes are presented. This is followed by description of the top-down analysis and bottom-up reconstruction method in Sections 3 and 4 respectively. The design of the Fast Filter Bank is given in Section 5 and its merit in the separation of harmonics is illustrated in Section 6. Results of experiments are summarized in Section 7 and concluding remarks presented in Section 8.

2. SPECTRAL PROFILES OF NOTES

First, the ratios of the amplitudes of the first four harmonics and the fundamental frequency of all notes of the musical instrument are determined. Mathematically, the following ratios are computed.

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

where A_i and A_j are the amplitudes of the ith and the jth harmonics respectively and A_o 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,

$$\overline{r_{ij}} = \frac{1}{N}\sum_{k=0}^{N-1} r_{ij}(k) \qquad (2)$$

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. TOP DOWN ANALYSIS

For recognition, the signal is first filtered at 10 kHz and sampled at 22050 samples per second and coded with 8 bits per sample.

The top down analysis consists of a harmonics-tracking algorithm. The purpose of the analysis is to identify all the prominent spectral components and hence the possible notes that may be present. 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.

4. BOTTOM UP RECONSTRUCTION

The bottom up reconstruction is based on a set of the expected profiles of single notes. The expected profiles of all the individual notes were obtained based on Equation 2 and stored in a database.

Based on the set of provisional notes obtained from the top down analysis, 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 top down analysis 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.

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. The bottom up reconstruction takes that into account. 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 of the test signal will be taken as the actual notes that are played.

5. FAST FILTER BANK

The basic way of obtaining the spectrum of a signal is using FFT. The disadvantages of the sliding FFT as a filter bank are its poor pass band response and high side lobes in its stop bands. The peak of the first side-lobe is about -13dB, regardless of the FFT length. Increasing the FFT length will result in a narrower main lobe width but has little effect on the side-lobes magnitudes.

On the other hand, 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.

Another beauty of FFB is that it is able to achieve these without much increase in computational load. The computation complexity of the sliding FFT is 1 complex multiplication per channel per sample while that of a 4096-channel FFB is 1.012 complex multiplications per channel per sample [6].

The impulse responses of prototype filters of the 4096-channel FFB designed for the proposed system are shown in Table I.

Table I: Impulse Response of Prototype Filters

n	$h^1_a(n)$	$h^2_a(n)$	$h^3_a(n)$	$h^4_a(n)$	$h^5_a(n)$	$h^6_a(n)$
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	$h^7_a(n)$	$h^8_a(n)$	$h^9_a(n)$	$h^{10}_a(n)$	$h^{11}_a(n)$	$h^{12}_a(n)$
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. 1(a) with an expanded view of the main lobe shown in Fig.1(b).

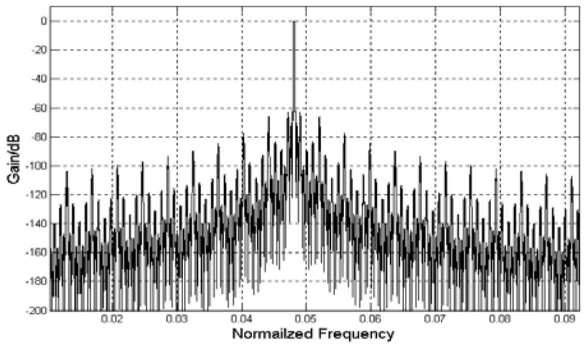
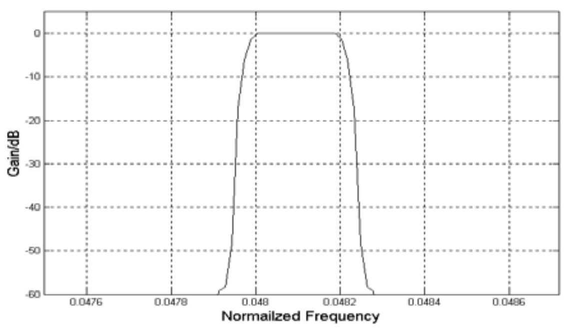


Fig.1 (a)



(b)

Fig.1: Frequency response of Channel 197 of a 4096 channel FFB (a) showing main lobe and side lobes, (b) expanded view of the main lobe.

6.SEPARATION OF HARMONICS

For multiple notes played at the same time, especially in cases where the harmonics of notes are close together, the FFB is very useful in identifying the relevant harmonics. As an example to illustrate the additional information provided by FFB, 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.2.

From Fig.2, it is observed that the third harmonic of Note C4 and the second harmonic of Note E4 have been fused and could not be resolved. With the application of the FFB, the two harmonics mentioned can be resolved as outputs of Channel 186 and Channel 197 as shown 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.

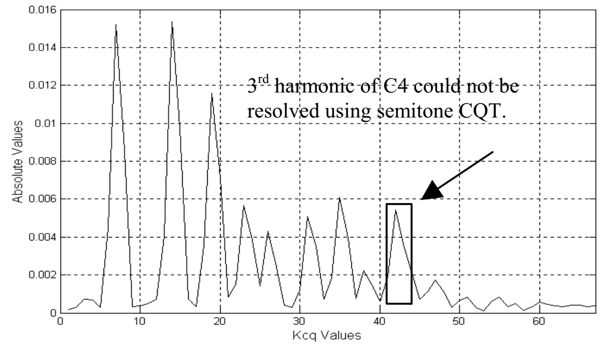


Fig.2: The Constant Q Transform of Notes C3, G3, C4 and E4 played simultaneously

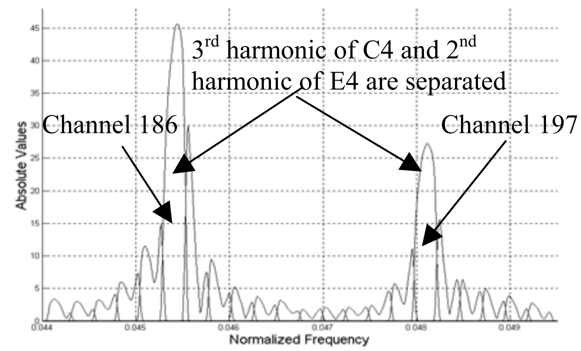


Fig.3: Output of the Fast Filter Bank

From Fig.4, the attack (A), delay (D), sustain (S) and release (R) (ADSR) profiles of the output signals of Channels 186 and 197 in the time domain that characterize musical notes [7] can be identified. This serves to confirm the presence of the notes.

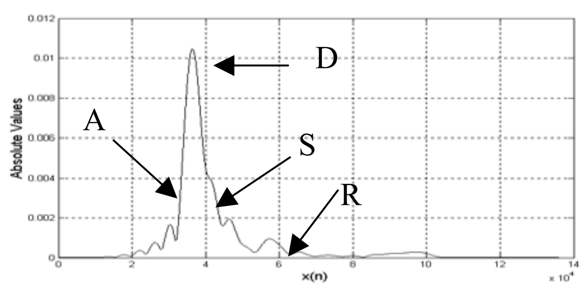


Fig.4(a)

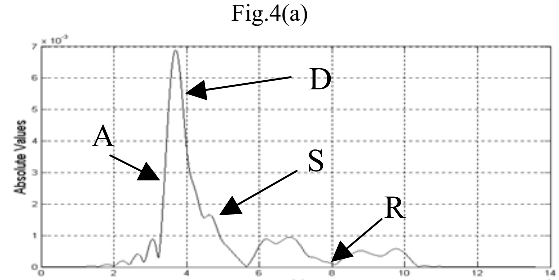


Fig.4(b)

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

As a matter of comparison, the output of Channel 194 that does not contain the fundamental or harmonics of any musical note is shown in Fig.5. It can be seen that the amplitude of the output is small and the ADSR profile is absent.

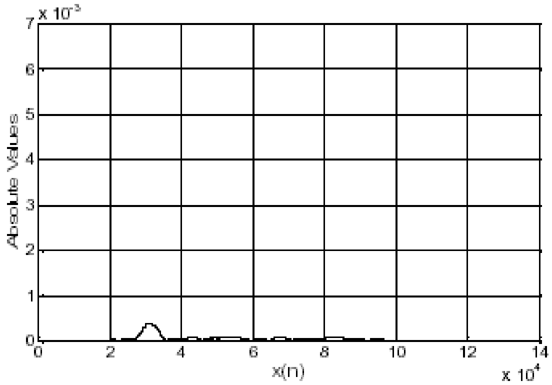


Fig.5: Output of Channel 194

7. RESULTS AND DISCUSSIONS

The proposed system 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 " Bagattelle In A Minor " were used. Notes that were taken form 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 musical notes tested, without and with the use of FFB are shown in Tables II and III respectively.

Table II: 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 III: 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 showed that the application of FFB has increased the recognition accuracy by 5%. With FFB, perfect recognition was achieved for chords with four notes and below. An overall accuracy of about 98% was achieved.

8. CONCLUSION

A new method of transcribing musical notes is presented. The method makes use of 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. The fast filter bank, which has high selectivity of the frequency bands, is used to extract the spectral components present in the signal. By analyzing the output of the filter bank, notes are positively identified and any false note is weeded out.

Results show that perfect identification can be achieved for chords with four notes and below. The overall algorithm is able to separate chords consisting up to 6 notes that were played simultaneously. The application of Fast Filter Bank has significantly improved the accuracy of identification especially for chords with multiple notes.

9. 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] 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

[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] 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

[6] 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

[7] James H. McClellan, Ronal W. Schafer "DSP FIRST A MULTIMEDIA APPROACH" Prentice-Hall 1998, pp 435-455.