BUILDING CHROMA FEATURE: COMPARISONS BETWEEN SHORT-TIME FOURIER TRANSFORM AND CHROMA CONVOLUTION METHOD

Jose Pedro de Santana Neto
University of Brasilia - FGA
1jpsneto@gmail.com

Henrique Gomes de Moura
University of Brasilia - FGA
hgmoura@yahoo.com

Fernando William Cruz
University of Brasilia - FGA
fwcruz@unb.br

ABSTRACT

This paper is presented as an alternative and efficient method to the construction of the chroma feature or Pitch Class Profile (PCP). Chroma Convolution Method (CCM), which is an operation essentially in the time domain, embodies peculiar characteristics that optimize the distinctions of musical tones and timber in musical instruments. For the demonstration of the efficacy of this method, comparative experiments with the traditional method, DFT or STFT, were made. The experiments have shown that CCM is the most efficient in the identification of musical tones; moreover, it holds the capacity of distinguishing musical tones of different musical instruments, when they are also played at the same time.

1. INTRODUCTION

The process of extraction of chroma feature is a steady matter in solutions of automatic music transcription. Basically, this process consists in obtaining the spectrum of frequency in audio signals, and also calculating the level of energy of each one of the 12 tones that are part of the temperament scale. Traditionally, the Fourier transform is used for this extraction since about 1999 with Fujishima [9], and, since then, a great effort is being made in order to optimize the Pitch Class Profile (PCP), built by the discrete Fourier transform (DFT).

Many studies approach the matter of construction of chrome feature in music. These studies have built solutions which are based on DFT and STFT (Short Time Fourier Transform) to the identification of melodies: [19], [1], [2], [10], [21], [8] and [14], this last one appling optimizations, by using particle filters. Also, there are some chroma feature studies which focus on the automatic transcription of chords via DFT: [12], [15], [11], [20], [5] [17], [7], [3], [4] and [13].

The aforementioned studies focus on adapted version of the DFT, has there are not many alternatives of tones characterization, besides the DFT. There are again many DFT limitations, quoted in papers like Harte's [11]. The study of Mauch [18] for instance, focus on the application of NNLS to the increasement of identification of tones, and paper [23] focus on the mathematical model to the representation of tones distributions and its frequencies. However, there still is the need of alternative methods in relation to DFT and the construction of chroma feature.

It is shown on this paper an alternative method for chroma feature caracterization and melodies identification, in time domain. The Chroma Convolution Method (CCM) embodies peculiar characteristics that optimize the distinctions of musical tones and timber in musical instruments. The CCM can replace the DFT or STFT to extract polyphonic sounds with more accuracy, as well as identifying tones of different musical instruments.

This paper has been organized in the following way: Section 2 describes the use, general aspects and limitations of DFT; Section 3 presents the conceptualization of the suggested method, characteristics and the process of construction of chroma feature using CCM; Section 4 presents two comparative experiments, of DFT and CCM methods; Section 5 delimits discussions about the results of CCM's efficacy; Section 6 presents conclusions and further papers.

2. DISCRETE FOURIER TRANSFORM (DFT)

Chroma feature or pitch class profile (PCP) have been exclusively used as an front-end to the recognition of chords or extraction of melodies of recorded audio. Fujishima [9] developed a system of automatic transcription of chords in real-time, where he derives a chroma feature in 12 dimensions from the DFT audio signal. Since then, the Discrete Fourier Tranform (DFT or FFT) has been used to the construction of chromatic features the audio. The function of this transform is translating pieces of information which are in time domain to frequency domain in a way to project, in orthonormal basis, the value of each sine component present in the signal in question. This projection results of the sum of inner sines (complex exponentials) products, by the signal [22].

Since this transform only offers information in terms of frequency, the need of adapting it to the visualization of the variation of frequency through time has taken place.

This adaption is called short time fourier transform (STFT) [6]. In a nutshell, this process of windowing consists of dividing the signal in parts to be analyzed individually in the frequency domain, which means that in each minute-time that refers to each sound clip it is possible to obtain frequency analysis.

However, this technique presents some limitations:

- the appearance of aliasing that can make the identification of actually played tones difficult;
- problems of spectral leakage, generated from signal truncation, possibly causing frequencies that make the spectrum visualization more difficult;
- problems in the determination of the window length, as the sampling parameter can limit the frequency band to be analyzed.

It can be observed that the presented limitations are, mostly, originated from the digital processing of signals necessary to the constitution of frequency spectrum. Being so, one might research on alternatives for this problem of chroma feature construction in the frequency domain, in order to preserve the original characteristics of audio signals.

3. CHROMA CONVOLUTION METHOD (CCM)

The chroma convolution method (CCM) aims to project the sound passage of the sign on each one of the tones sought. It can be observed that when the sign to be processed is projected on the sign of a musical note, through convolution, the sound of the first tone is amplified and the others are suppressed.

Let's take a note formed from a monochromatic sine. When this sign is convoluted with processed audio sign it is possible to extract as a result the level of energy related to that frequency. This energy is measured through the following Eqn (1) equation:

$$E = \sum [f(x) * g(x)]^2 \tag{1}$$

In this context, the chroma features were built from the following procedures:

1. the audio was divided to be analyzed in signals with pre-determined duration;
2. monochromatic sine functions were used to build musical tones in time domain, corresponding to the chromatic musical scale;
3. each part of the audio signal was convoluted with the musical tones;
4. the convolution energy was extracted from the equation Eqn (1);
5. Finally, each energy was summed up with the respective eights, originating the chroma feature, presented in a schema on Figure 1.

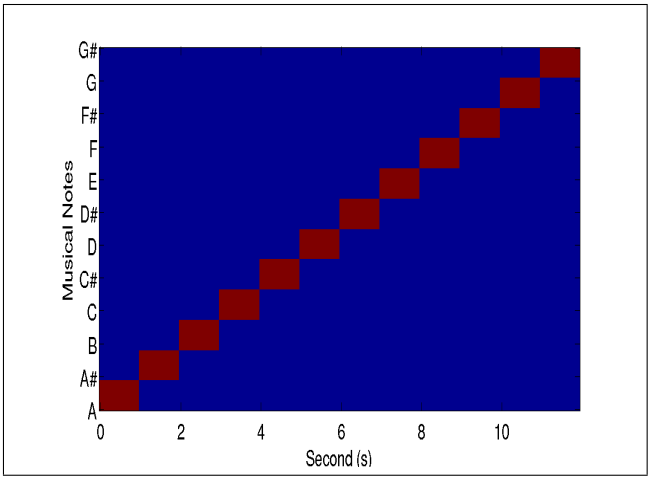

Figure 1. Schematic representation of the chroma feature. On this chart, a full chromatic scale starting from A up to G#.

Figure 2 represents a schematic point of view of the proceeding described. By the end of the process, the chroma features can be visualized from an spectrogram of 12 dimensions.

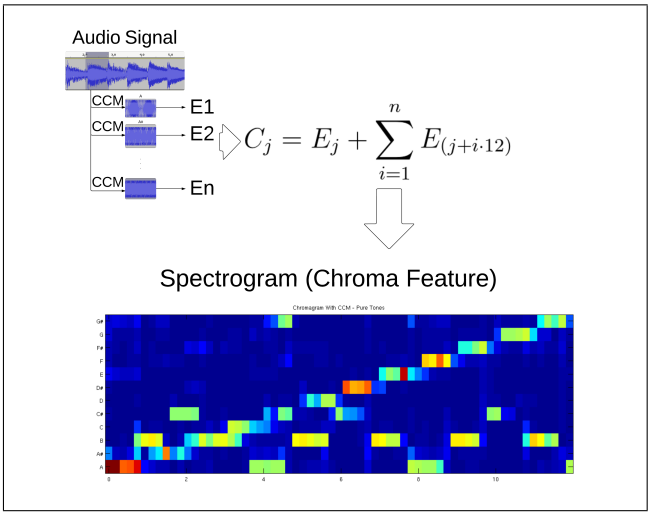

Figure 2. Schematic of process to build chroma feature using CCM.

4. EXPERIMENTS AND RESULTS

In order to proove the efficacy of CCM in relation to the traditional method STFT [16], in what concerns the identification of tones played from their chroma features, 2 experiments were made, using MATLAB as programming language. Recorded audio signals of actual musical instruments were used as data entry [1] .

4.1 Experiment 1: Identification of Musical Notes

The first experiment is about the identification of musical tones in a piano melody. Those tones were played accord-

[1] Code and files available in https://github.com/josepedro/ismir_article.

ing to Figure 3.

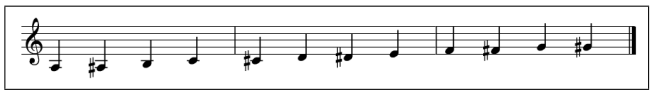

Figure 3. Recorded notes played on piano.

Convolution with 96 pure tones of different frequency was used to generate the chroma feature in CCM results. The windows of both proposals were set in the length of 0.120 seconds and the samplings rate of the audio was of 44.1 kHz.

4.1.1 STFT Results

In the following, chroma feature found using the traditional method of STFT in Figure 4. It can be observed that the chromatic scale was identified, yet with some discrepancies in relation to the exact moment that the tones were played. One can notice, for example, that one tone E was identified in the beginning as, in reality, it should be identified in about 7 seconds.

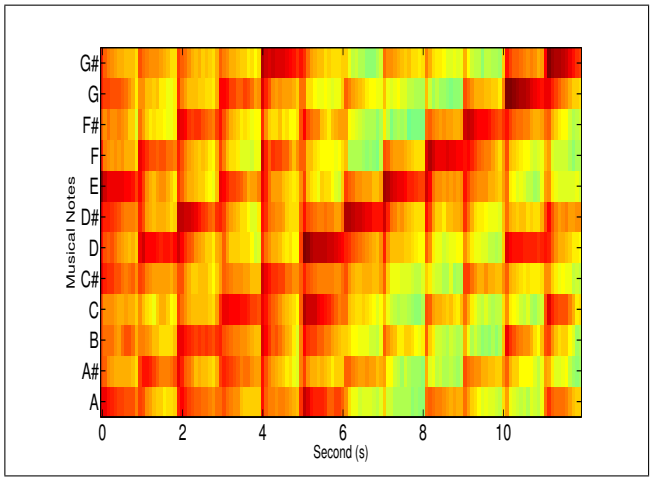

Figure 4. Obtained chroma feature by using the STFT method.

An efficiency analysis of the tones recorded in audio with the sequence of tones identified in the chroma feature was also made. The identified tones were selected thorough the highest intensity in each minute-time. Table 1 represents the tones which were played and also the ones which were identified in the chroma feature. There were 9 out of 14 tones playing in the correct moment, so, 64.3% of fidelity to the original melody.

4.1.2 CCM Results

Chroma feature found using CCM on Figure 5. It can be observed that the melody represented itself more clearly. A comparative table of the played and identified tones was made through the extraction of the tone of the chroma feature with more intensity in each minute-time. Keeping in mind the results of Table 2, 13 out of 14 tones were playing in the correct moment, so, 92.9% fidelity to the original melody.

Played Notes	Identified Notes
A	E
A#	D
A#	D#
B	B
B#	E
C	D
C#	G#
D	D
D#	D#
E	E
F	F
F#	F#
G	G
G#	G#

Table 1. Comparison between played and identified notes by the STFT method.

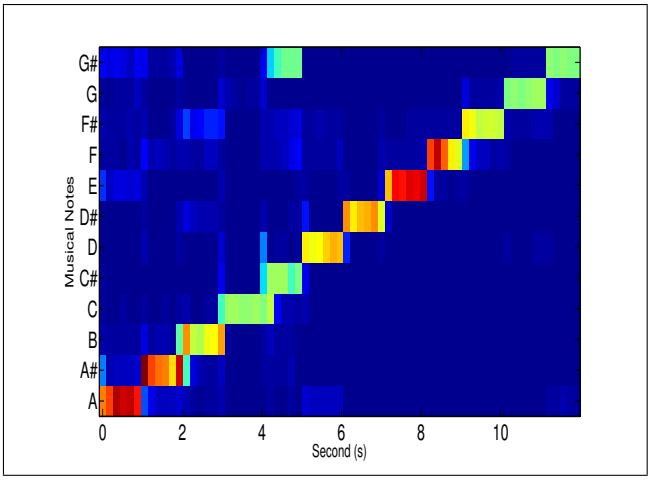

Figure 5. Obtained chroma feature by using the Chroma Convolution method.

Played Notes	Identified Notes
A	A
A#	A#
B	B
C	C
C#	C#
C#	G#
C#	C#
D	D
D#	D#
E	E
F	F
F#	F#
G	G
G#	G#

Table 2. Comparison between played and identified notes by the Chroma Convolution Method.

4.2 Experiment 2: Identification of Musical Notes of Different Instruments

The second experiment is about the identification of tones of different instruments. For this to happen, the following melodies of the piano and the acoustic guitar were played together, as represented on Figure 6.

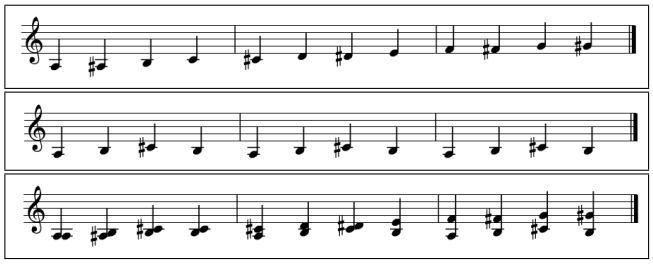

Figure 6. The first sheet music is only piano, the second sheet music is only acoustic guitar and the third sheet music is the both together.

In CCM results, convolution of 12 tones of the acoustic guitar and 12 tones of the piano of different frequencies was used, generating two chroma feature respectively, one for each instrument. Both proposal windows were set in the length of 0.120 seconds, and the samplings rate of the audio was of 44.1 kHz.

4.2.1 STFT Results

In the following chroma feature was found using the traditional method of STFT on Figure 7. It can be observed on this picture that the chromatic scale of the piano was partially identified. The acoustic guitar tones are practically unnoticeable.

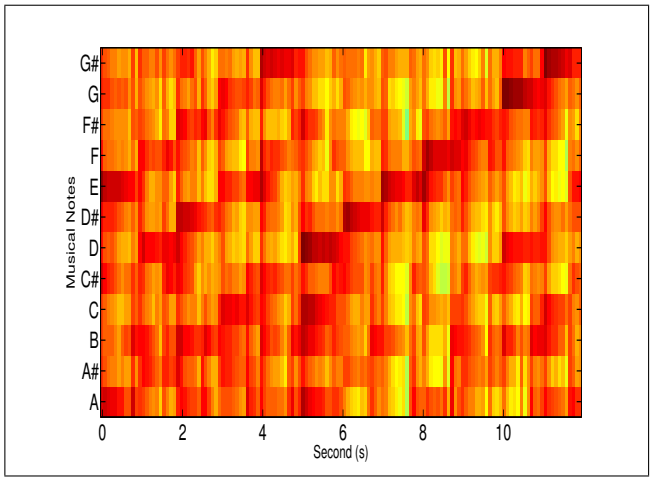

Figure 7. Chroma feature of audio using STFT.

It was also made an efficiency analysis on the recorded tones of the two instruments, with the sequence of identified tones in the chroma feature, using highest intensity in each minute-time. In the following the comparative table of Table 3.

First of all it is important to highlight that it is not possible to obtain a chroma feature for each instrument using the STFT method. This way, using an extraction of the tone of the chroma feature with more intensity in each minute-time, it was possible to obtain a value of the global efficiency, about 38.7%, so, in each moment of time one of the two tones played by the instruments, piano and acoustic guitar, should be identified. The results of this analysis are on Table 3.

Piano	Acoustic Guitar	Identified Notes
E	A	A
A	A	A
B	A	A
E	A	A
D	A#	B
F	A#	B
D	A#	B
D#	B	C#
F#	B	C#
B	B	C#
F#	B	C#
C	C	B
A	C	B
E	C	B
G#	C#	A
B	C#	A
D	D	B
D#	D#	C#
B	D#	C#
E	E	B
F	F	A
F#	F#	B
F	F#	B
C#	F#	B
F#	F#	B
G	G	C#
B	G	C#
G	G	C#
G#	G#	B
A	G#	B
E	G#	B

Table 3. Comparison between played notes of piano, acoustic guitar and identified by the STFT method.

4.2.2 CCM Results

In the following chroma feature found using the traditional method using CCM of piano tones on Figure 8.

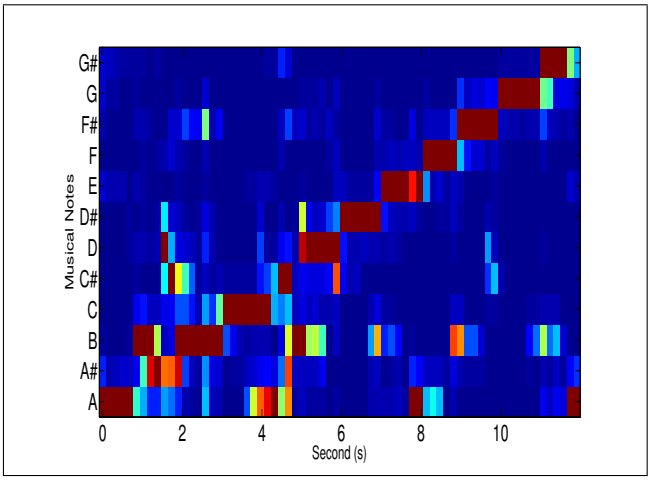

Figure 8. Chroma feature of audio using CCM of piano tones.

It can be observed that on Figure 8 the played melody of the piano was identified more clearly. It is also visi-

ble that some tones were executed by the acoustic guitar. A comparative table of the tones played and identified was made, also through the extraction of the tone of the chroma feature with more intensity in each minute-time. According to the results of Table 4, there were 13 out of 18 tones of the piano in correct minute-time, representing 72.2% fidelity to the original melody. The guitar, on the other hand, resulted in 5 out of 18 tones in correct minute-time, so, 27.7%. This result shows that in this experiment the CCM amplified the executed tones, in the case of the piano, and suppressed the tones played by the acoustic guitar.

Piano	Acoustic Guitar	Identified Notes
A	A	A
A	B	B
A#	B	A#
A#	B	C#
A#	C#	A#
B	B	B
C	A	C
C#	A	A
C#	B	C#
D	B	B
D	C#	D
D#	B	D#
E	A	E
F	B	F
F#	C#	F#
G	B	G
G#	B	G#
G#	B	A

Table 4. Comparison between played notes of piano, acoustic guitar and identified notes by the CCM with piano tones.

In the following chroma feature found using the CCM of tones of the acoustic guitar on Figure 9. It can be observed on the picture that the melody executed by the acoustic guitar was also identified more clearly. Some tones of the chromatic scale executed by the piano are visible as well, although suppressed if in contrast with the result presented on Figure 8. A comparative table of the tones which were played and identified was also made through the extraction of the tone of the chroma feature with more intensity in each minute-time. As it is seen on Table 5, 9 out of the 19 tones executed by the acoustic guitar were identified correctly, so, 47.4% fidelity to the original melody. Talking about the piano, 2 out of 19 tones were identified correctly, representing 10.5%.

5. DISCUSSION

The results reached by CCM are more satisfactory than the traditional method STFT, although a largest computational cost is attained to CCM calculations. The time of processing of the STFT method was approximately 0.1 seconds against 13.2 seconds of CCM in the first experiment. It has to be highlighted that the high computational cost of

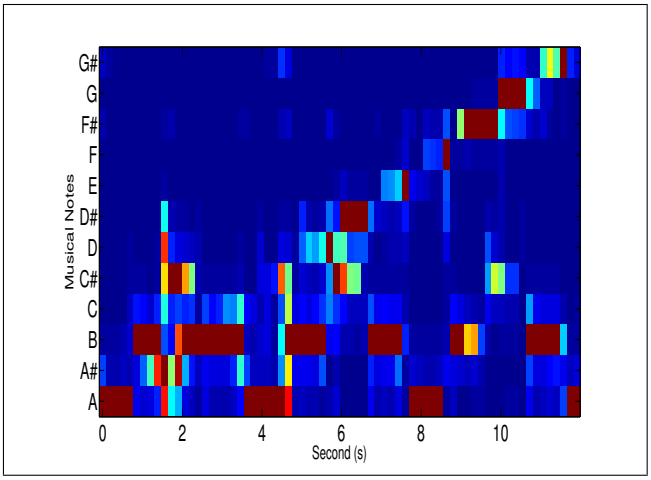

Figure 9. Chroma feature of audio using CCM of acoustic guitar tones.

Piano	Acoustic Guitar	Identified Notes
A	A	A
A#	B	B
A#	B	A#
B	C#	C#
C	B	B
C#	A	A
D	B	C#
D#	C#	B
D#	C#	D
D#	C#	C#
E	B	D#
F	A	B
F	A	E
F#	B	A
F#	B	B
G	C#	F#
G	C#	C#
G#	B	G
G#	B	B

Table 5. Comparison between played notes of piano, acoustic guitar and identified notes by the CCM with acoustic guitar tones.

CCM does not represent a limitation in practical terms, as the method can be executed, without great problems, by a personal computer.

In the first experiment, the CCM presented an improvement in its efficiency of about 28.6% in relation to the traditional method. In a way, the CCM values the energy of the tones that have some correlation to the reference signal, what favor identification. Another important point is the fact that the CCM suppresses the signal information which is not correlated to the reference signals used in the convolution.

Moreover, there is the possibility, offered by the CCM, of focusing the attention on an specific instrument, through the use of its timber, without any additional costs to the

algorithm. Reference tones recorded by the piano were better identified in the second experiment, as the acoustic guitar tones were suppressed. The inverted result was observed when the same audio signal was submitted to the reference signals produced by the acoustic guitar. The order of magnitude found to the identification of the privileged tones for the reference signals in relation to the suppressed tones was about 40.7%. The traditional method STFT does not offer the possibility, unless the optimizations start being practiced in order to consider the timber of the instruments in a similar way.

6. CONCLUSIONS AND FUTURE WORKS

This paper presented a new way of constructing the chroma feature or PCP, from the use of a new method, here denominated chroma convolution method (CCM). The obtained results were confronted with the traditional algorithm STFT, showing more efficiency to the CCM when identifying musical tones.

Another advantaged found to the CCM in relation to the traditional method STFT is about the detection of polyphonic melodies, produced by more than one instrument. The CCM operates with reference signals for the tones to be identified, what can be obtained from actual instruments, which naturally leads to the inclusion of information of timber of identification.

This paper presented a standard and basic way of CCM use, without any procedure for noise reduction or any improvement over the tone identification. In the same way that the traditional method STFT has been developed for more sophisticated applications, like the NNLS [18] and the particle filters [14], the CCM method can be improved, aiming better results.

It is worth mentioning that the CCM and the STFT methods were compared in their basic applications, so, there is a lot to be done in the CCM strategy in what concerns the identification of tones and polyphonic melodies, considering the great amount of papers already produced by the traditional solutions in the frequency domain.

7. REFERENCES

[1] Majid A Al-Taee, Mohammad S Al-Rawi, and Fadi M Al-Ghawanmeh. Time-frequency analysis of the arabian flute (nay) tone applied to automatic music transcription. In *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, pages 891–894. IEEE, 2008.

[2] Isabel Barbancho, Cristina de la Bandera, Ana M Barbancho, and Lorenzo J Tardon. Transcription and expressiveness detection system for violin music. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 189–192. IEEE, 2009.

[3] Nicolas Boulanger-Lewandowski, Yoshua Bengio, and Pascal Vincent. Audio chord recognition with recurrent neural networks. In *ISMIR*, pages 335–340, 2013.

[4] Ruofeng Chen, Weibin Shen, Ajay Srinivasamurthy, and Parag Chordia. Chord recognition using duration-explicit hidden markov models. In *ISMIR*, pages 445–450. Citeseer, 2012.

[5] Taemin Cho, Ron J Weiss, and Juan Pablo Bello. Exploring common variations in state of the art chord recognition systems. In *Proceedings of the Sound and Music Computing Conference (SMC)*, pages 1–8. Citeseer, 2010.

[6] Leon Cohen. *Time-frequency analysis*, volume 1406. Prentice Hall PTR Englewood Cliffs, NJ:, 1995.

[7] Bas de Haas, José Pedro Magalhães, and Frans Wiering. Improving audio chord transcription by exploiting harmonic and metric knowledge. In *ISMIR*, pages 295–300. Citeseer, 2012.

[8] Jana Eggink and Guy J Brown. Extracting melody lines from complex audio. In *ISMIR*, 2004.

[9] Takuya Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In *Proc. ICMC*, volume 1999, pages 464–467, 1999.

[10] Emilia Gómez and Perfecto Herrera. Automatic extraction of tonal metadata from polyphonic audio recordings. In *Proceedings of 25th International AES Conference, London*, 2004.

[11] Christopher Harte. *Towards automatic extraction of harmony information from music signals*. PhD thesis, Department of Electronic Engineering, Queen Mary, University of London, 2010.

[12] Christopher Harte and Mark Sandler. Automatic chord recognition using quantised chroma and harmonic change segmentation. *MIREX Annual Music Information Rerieval eX-change. Available at http://www. music-ir. org/mirex/abstracts/2009/harte\ _mirex09. pdf*, 2009.

[13] Alex Hrybyk. *Combined audio and video analysis for guitar chord identification*. PhD thesis, Drexel University, 2010.

[14] Seokhwan Jo and Chang D Yoo. Melody extraction from polyphonic audio based on particle filter. In *ISMIR*, pages 357–362. Citeseer, 2010.

[15] Maksim Khadkevich and Maurizio Omologo. Time-frequency reassigned features for automatic chord recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 181–184. IEEE, 2011.

[16] Columbia University LabROSA. Chroma feature analysis and synthesis, 2015.

[17] Kyogu Lee. Automatic chord recognition from audio using enhanced pitch class profile. In *Proc. of the International Computer Music Conference*, 2006.

[18] Matthias Mauch and Simon Dixon. Approximate note transcription for the improved identification of difficult chords. In *ISMIR*, pages 135–140, 2010.

[19] Youhei Muto and Toshiyuki Tanaka. Transcription system for music by two instruments. In *Signal Processing, 2002 6th International Conference on*, volume 2, pages 1676–1679. IEEE, 2002.

[20] Geoffroy Peeters. Chroma-based estimation of musical key from audio-signal analysis. In *ISMIR*, pages 115–120, 2006.

[21] Zheng Tang and Dawn AA Black. Melody extraction from polyphonic audio of western opera: A method based on detection of the singers formant.

[22] Parishwad P Vaidyanathan. *Multirate systems and filter banks*. Pearson Education India, 1993.

[23] Gregory H Wakefield. Mathematical representation of joint time-chroma distributions. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pages 637–645. International Society for Optics and Photonics, 1999.