

PAPER TEMPLATE FOR ISMIR 2015

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

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

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

ABSTRACT

Neste trabalho é apresentado um método alternativo e eficiente para a construção de chroma feature ou Pitch Class Profile (PCP). O Chroma Convolution Method (CCM), que é uma operação essencialmente no domínio temporal, possui características particulares que otimizam a distinção de notas e timbres de instrumentos musicais. Para a demonstração da eficácia desse método, foram feitos dois experimentos comparativos com o método tradicional DFT ou STFT. Os experimentos mostraram que o CCM é mais eficaz em identificação das notas, como também, possui a capacidade de distinção de notas de diferentes instrumentos musicais tocados ao mesmo tempo.

1. INTRODUCTION

O processo de extração de características cromáticas é um problema constante em soluções de transcrição automática de música. Basicamente este processo consiste em obter o espectro de frequências do sinal de áudio, calcular o nível de energia de cada nota correspondente a frequência de cada uma das 12 notas musicais que compõem a escala temperada. Tradicionalmente usa-se a transformada de Fourier para tal tarefa desde meados de 1999 com Fujishima [9] e, desde então, vários esforços são empreendidos para otimizar o Pitch Class Profile (PCP) construído via transformada discreta de Fourier (DFT).

Vários estudos abordaram o problema de construção de chroma feature em músicas. Tais estudos construíram soluções embasadas em DFT ou STFT (Short Fast Fourier Transform) para identificação de melodias: [19], [1], [2], [10], [21], [8] e [14], esse último usando filtros partícula para otimização. Há também estudos de soluções de chroma feature com foco em transcrição automática de acordes via DFT: [12], [15], [11], [20], [5], [17], [7], [3], [4] e [13].

Tais estudos focaram soluções adaptadas em DFT e pouco se tem outras alternativas para caracterização de notas. Há também várias limitações citadas da DFT em trabalhos como o de Harte [11]. O estudo de Mauch [18] por exemplo foca a aplicação de NNLS para aumentar a

identificação de notas e o trabalho [23] foca a modelagem matemática para a representação da distribuição das notas em suas frequências. Porém há ainda a carencia de métodos alternativos em relação a DFT para a construção de chroma feature.

O presente trabalho foca a apresentação de um método alternativo e mais eficiente que a DFT, no que tange a identificação das notas, para a construção do chroma feature. A alternativa Chroma Convolution Method (CCM), que é uma operação essencialmente de domínio temporal, possui características particulares que otimizam a distinção de notas e timbres de instrumentos musicais. O CCM pode ser usado no lugar da DFT ou STFT para extrair sons polifônicos com mais acurácia, identificando notas de diferentes instrumentos musicais.

Este paper está organizado da seguinte forma: Section 2 descreve o uso, aspectos gerais e limitações da DFT; Section 3 apresenta a conceituação do método proposto, características e processo de construção do chroma feature usando o CCM; Section 4 apresenta dois experimentos de comparação dos métodos DFT e CCM; Section 5 delimita discussões sobre os resultados da eficácia do CCM; Section 6 apresenta conclusões e trabalhos futuros.

2. DISCRETE FOURIER TRANSFORM (DFT)

Chroma feature ou pitch class profile (PCP) tem sido usado quase exclusivamente como um front-end para o reconhecimento de acordes ou extrato de melodias de áudio gravado. Fujishima [9] desenvolveu um sistema de transcrição automática de acordes em tempo real, onde ele deriva um chroma feature de 12 dimensões a partir do DFT do sinal de áudio. Desde então a transformada discreta de Fourier (DFT ou FFT) tem sido muito utilizada para a construção das características cromáticas (chroma feature) do áudio. A função dessa transformada é traduzir informações que estão em domínio temporal para domínio frequencial de tal forma a projetar, em bases ortonormais, o valor de cada componente senoidal presente no sinal tratado. Essa projeção se dá pelo somatório do produto interno das senóides (exponenciais complexas) pelo sinal [22].

Visto que essa transformada somente oferece informações em termos de frequências, surgiu-se a necessidade de adaptá-la para a visualização das variações das frequências ao longo do tempo. Essa adaptação se denomina short time Fourier transform (STFT) [6]. Esse processo de janelamento se resume a dividir o sinal em partes, para serem analisadas individualmente no domínio da frequência, ou seja, a cada instante de tempo referente a cada trecho é



© Jose Pedro de Santana Neto, Henrique Gomes de Moura, Fernando William Cruz.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Jose Pedro de Santana Neto, Henrique Gomes de Moura, Fernando William Cruz. "Paper Template For ISMIR 2015", 16th International Society for Music Information Retrieval Conference, 2015.

possível obter uma análise frequencial.

Porem essa tecnica possui as seguintes limitacoes:

- surgimento de frequencias fantasmas (alising) que podem dificultar a identificacao das notas musicais que foram verdadeiramente tocadas;
- problemas de vazamento gerados pelo truncamento do sinal, podendo gerar frequencias que dificultam a visualizacao do espectro (leakage);
- dificuldades na determinacao do tamanho da janela pois trata-se de um parametro de amostragem que pode limitar a banda de frequencia a ser analisada.

Observa-se que as limitacoes apresentadas sao, em grande parte, oriundas do processamento digital de sinais necessario para a constituicao do espectro de frequencias. Deste modo, surge a necessidade de se pesquisar alternativas para o problema de construcao do chroma feature no dominio do tempo, preservando assim as caracteristicas originais do sinal de audio.

3. CHROMA CONVOLUTION METHOD (CCM)

O chroma convolution method (CCM) objetiva a projecao dos trechos do sinal sobre cada um dos sinais das notas musicais procuradas. Observa-se que quando o trecho do sinal eh projetado sobre o sinal de uma nota musical, atraves da convolucao, esta nota eh amplificada e as demais sao suprimidas.

Seja uma nota musical formada a partir de uma senoide monocromatica. Quando esse sinal eh convoluido com um trecho de sinal de audio eh possível extrair como resultado o nivel de energia relacionado a esta frequencia. Tal energia eh mensurada a partir da seguinte equacao Eqn (1):

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

Neste contexto as caracteristicas cromaticas (chroma feature) foram construidas a partir dos seguintes procedimentos:

1. dividiu-se o audio a ser analisado em sinais com duracao pre-determinada;
2. funcoes senoidais monocromaticas foram utilizadas para construir notas musicais no dominio do tempo, correspondendo a escala cromatica musical;
3. convoluiu-se cada parte do sinal de audio com as notas musicais;
4. a energia da convolucao foi extraida a partir da equacao Eqn (1);
5. por fim cada energia foi somada com suas respectivas oitavas, originando o chroma feature, esquematicamente apresentado em Figure 1.

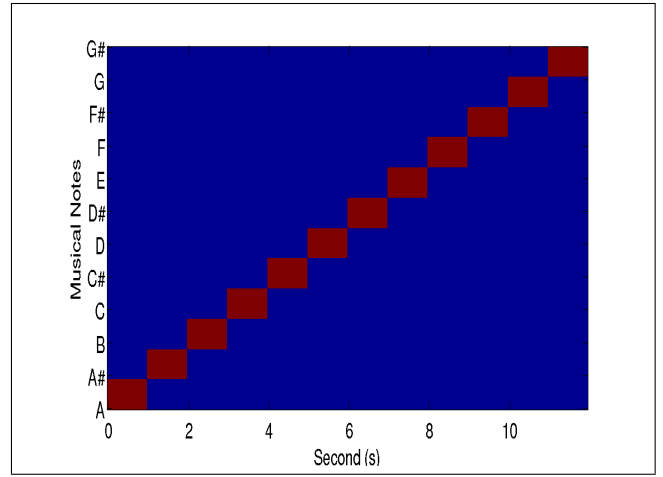


Figure 1. Schematic representation of the chroma feature. On this chart, a full chromatic scale starting from A up to G#.

A Figure 2 representa uma visao esquematica do procedimento descrito. Ao final do processo as caracteristicas cromaticas (chroma feature) podem ser visualizadas a partir de um espectrograma de 12 dimensoes.

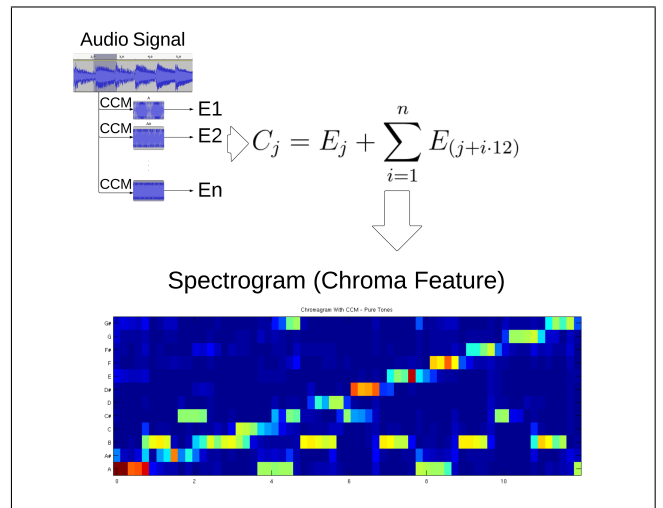


Figure 2. Schematic of process to build chroma feature using CCM.

4. EXPERIMENTS AND RESULTS

Com o intuito de demonstrar a eficacia do CCM em relacao ao tradicional metodo STFT [16], no que diz respeito a identificacao das notas tocadas a partir de suas caracteristicas cromaticas (chroma feature), foram realizados 2 experimentos usando a linguagem de programacao MATLAB. Como entrada de dados usou-se sinais de audio gravados por instrumentos musicais reais¹.

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

4.1 Experiment 1: Identification of Musical Notes

O primeiro experimento se trata de identificacao de notas musicais numa melodia de piano. Tais notas foram tocadas segundo a Figure 3.



Figure 3. Recorded notes played on piano.

Utilizou-se convolucao com 96 tons puros de diferentes frequencias para gerar o chroma feature. As janelas de ambas as propostas foram configuradas no tamanho de 0.120 segundos e a taxa de amostragem do audio foi de 44.1 kHz.

4.1.1 STFT Results

Segue chroma feature encontrado utilizando o metodo tradicional da STFT em Figure 4. Observa-se que a escala cromatica foi identificada, porem com erros em relacao ao momento exato em que as notas foram tocadas. Nota-se, por exemplo, que uma nota E foi identificada no comeco quando na verdade deveria ter sido identificada nos instantes proximos de 7 segundos.

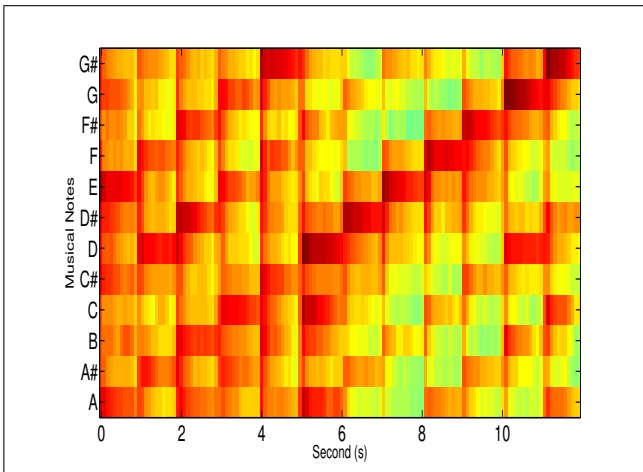


Figure 4. Obtained chroma feature by using the STFT method.

Foi tambem feito uma analise de eficiencia das notas gravadas no audio com a sequencia de notas identificadas no chroma feature. As notas identificadas foram selecionadas atraves da maior intensidade em cada instante de tempo. A Table 1 representa as notas tocadas e as que foram identificadas no chroma feature. Houveram 9 de 14 notas tocadas em instantes corretos, ou seja, 64.3% de fidelidade com a melodia original.

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.

4.1.2 CCM Results

Segue chroma feature encontrado utilizando o CCM em Figure 5.

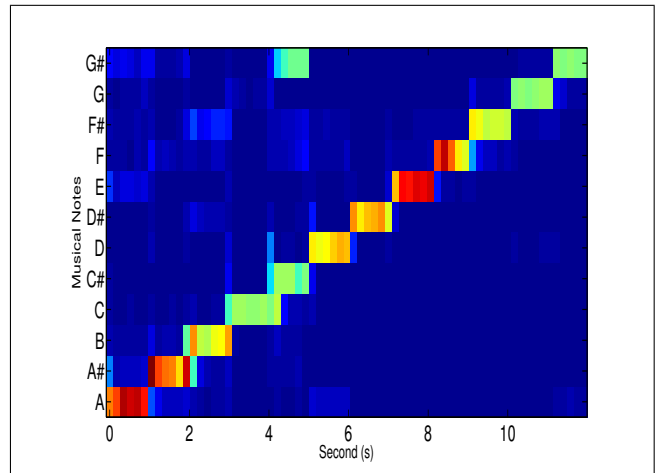


Figure 5. Obtained chroma feature by using the Chroma Convolution method.

Observa-se na figura Figure 5 que a melodia apresentou uma maior clareza. Uma tabela comparativa das notas tocadas e identificadas foi feita tambem atraves da estrategia dos maximos. Em vista dos resultados de Table 2, houve 13 de 14 notas tocadas em instantes corretos, ou seja, 92.9% de fidelidade com a melodia original.

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

O segundo experimento diz respeito a identificao de notas musicais de diferentes instrumentos. Para tal foram tocadas as seguintes melodias no piano e violao juntas representadas em 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.

Utilizou-se convolucao com 12 tons de violao e 12 tons de piano de diferentes frequencias gerando dois chroma feature respectivamente, um para cada instrumento. As janelas de ambas as propostas foram configuradas no tamanho de 0.120 segundos e a taxa de amostragem do audio foi de 44.1 kHz.

4.2.1 STFT Results

Segue chroma feature encontrado utilizando o metodo tradicional da STFT em Figure 7. Observa-se nessa figura que a escala cromatica do piano foi parcialmente identificada. As notas do violao sao praticamente imperceptiveis.

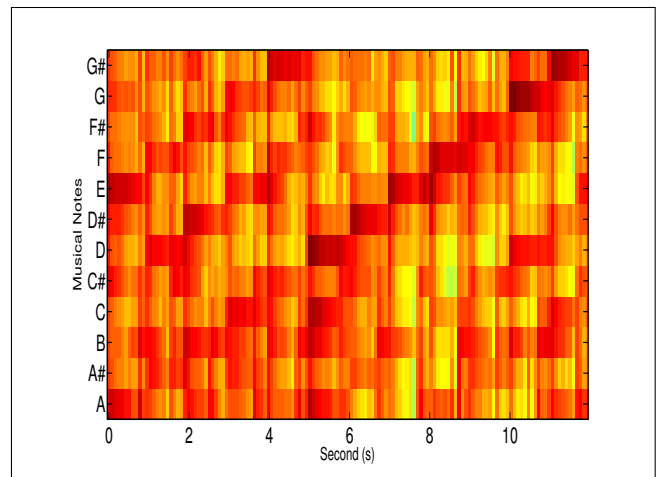


Figure 7. Chroma feature of audio using STFT.

Foi tambem feito uma analise de eficiencia das notas gravadas dos dois instrumentos com a sequencia de notas identificadas no chroma feature, usando a maior intensidade em cada instante de tempo. Segue tabela comparativa em 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.

Primeiramente eh importante ressaltar que nao eh possivel obter um chroma feature para cada instrumento usando o metodo STFT. Deste modo, utilizando a estrategia da maxima intensidade a cada instante de tempo, so foi possivel obter um valor de eficiencia global em torno de 38.7%, ou seja, a cada instante de tempo uma nota das duas notas tocadas pelos instrumentos, piano e violao, deveria ter sido identificada. Os resultados dessa analise se encontram em Table 3.

4.2.2 CCM Results

Segue chroma feature encontrado utilizando o CCM de tons de piano:

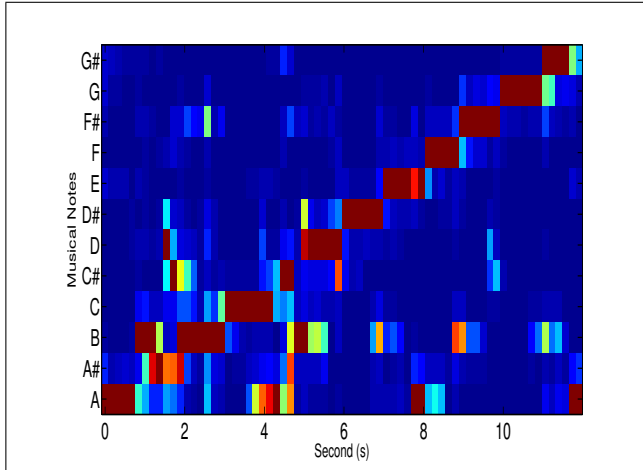


Figure 8. Chroma feature of audio using CCM of piano tones.

Observa-se na figura Figure 8 que a melodia executada no piano foi identificada com maior clareza. Sao visiveis tambem algumas notas executadas no violao. Uma tabela comparativa das notas tocadas identificadas foi feita tambem atraves da estrategia dos maximos. Em vista dos resultados de Table 4, houveram 13 de 18 notas de piano em instantes corretos, ou seja, 72.2% de fidelidade com a melodia original. Quanto ao violao, houveram 5 de 18 notas em instantes corretos, ocasionando em 27.7%. Tal resultado mostra que neste experimento o CCM amplificou as notas executadas pelo piano e suprimiu as notas executadas pelo violao.

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.

Segue chroma feature encontrado utilizando o CCM de tons de violao em Figure 9. Observa-se na figura que a melodia executada pelo violao tambem foi identificada com maior clareza. Sao visiveis tambem algumas notas da escala cromatica executada no piano, que ficaram, em contraste com o resultado apresentado pela Figure 8, suprimidas. Uma tabela comparativa das notas tocadas e identificadas foi feita tambem atraves da estrategia dos maximos. Em vista dos resultados de Table 5, houveram 9 de 19 notas executadas no violao identificadas corretamente, ou seja, 47.4% de fidelidade com a melodia original. Quanto ao piano, houveram 2 de 19 notas identificadas corretamente, ocasionando em 10.5%.

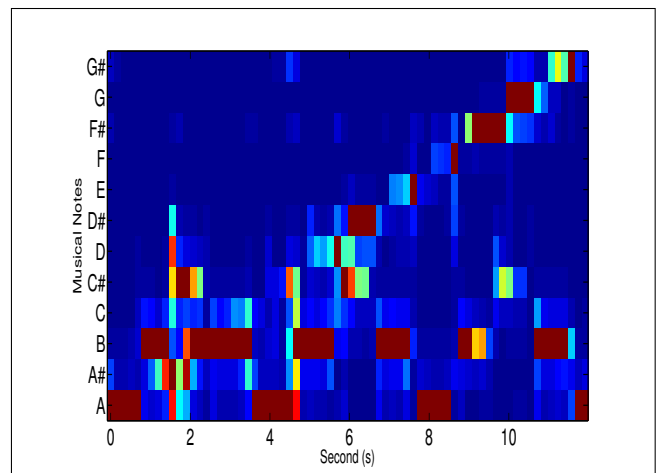


Figure 9. Chroma feature of audio using CCM of acoustic guitar tones.

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.

5. DISCUSSION

Os resultados alcançados pelo CCM se apresentam mais satisfatórios do que o método tradicional STFT, porém com um maior custo computacional agregado aos cálculos. O tempo de processamento do método da STFT foi de aproximadamente 0.1 segundos contra 13.2 segundos do CCM, no primeiro experimento. Ressalta-se que o elevado custo computacional do CCM não representa uma limitação em termos práticos pois o método pode ser executado, sem grandes dificuldades, por um computador pessoal.

No primeiro experimento, o CCM apresentou uma melhora na eficiência de aproximadamente 28.6% em relação ao método tradicional STFT. De certa forma o CCM valoriza a energia das notas que tenham correlação com o sinal de referência, o que favorece a identificação. Outro ponto importante é o fato de o CCM suprimir quais informações do sinal que não estejam correlacionadas com os sinais de referência utilizados na convolução.

Outro aspecto importante é a possibilidade, oferecida pelo CCM, de focar a atenção num instrumento específico a partir da utilização de seu timbre, sem custos adicionais para o algoritmo. Notas de referência gravadas no piano foram melhor identificadas no experimento 2 pois as notas do violão foram suprimidas. O resultado inverso foi observado quando o mesmo sinal de áudio foi submetido aos sinais de referência produzidos pelo violão. A ordem de grandeza das diferenças encontradas para a identificação das notas privilegiadas pelos sinais de referência em relação as notas suprimidas foi da ordem de 40.7%, em média. O método tradicional STFT não oferece essa possibilidade a menos que otimizações sejam praticadas no sentido de considerar o timbre dos instrumentos de uma maneira sim-

ilar.

6. CONCLUSIONS AND FUTURE WORKS

Este trabalho apresentou uma nova forma de construção do chroma feature ou PCP, a partir da utilização de um novo método, aqui denominado por chroma convolution method (CCM). Os resultados obtidos foram confrontados com o algoritmo tradicional STFT revelando uma melhor eficiência do CCM na identificação das notas musicais.

Outra vantagem encontrada para o CCM, em relação ao método tradicional STFT, diz respeito à detecção de melodias polifônicas, produzidas por mais de um instrumento. O CCM opera com sinais de referência, para as notas musicais a serem identificadas, que podem ser obtidas de instrumentos reais, fato que, naturalmente, inclui informações de timbre na identificação.

Este trabalho apresentou uma maneira padrão e básica de utilização do CCM, sem quaisquer procedimentos elaborados para redução de ruído no processo de identificação das notas. Da mesma forma que o método tradicional STFT vem sendo desenvolvido para aplicações mais sofisticadas, como o NNLS [18] e filtro de partículas [14], o método CCM pode ser aprimorado para que melhores resultados sejam alcançados.

Vale ressaltar que o CCM e o método STFT foram aqui comparados em suas aplicações básicas, ou seja, muito se tem para fazer em relação à aplicação do CCM sobre a identificação de acordes e melodias polifônicas, considerando o grande volume e a qualidade dos trabalhos já produzidos pelas soluções tradicionais no domínio da frequência.

7. REFERENCES

- [1] Majid A Al-Taei, 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. Cite-seer, 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 Retrieval 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.