

PAPER TEMPLATE FOR ISMIR 2015

First author

Affiliation1

author1@ismir.edu

Second author

Retain these fake authors in

submission to preserve the formatting

Third author

Affiliation3

author3@ismir.edu

ABSTRACT

The abstract should be placed at the top left column and should contain about 150-200 words.

1. INTRODUCTION

O processo de extracao de caracteristicas cromaticas eh um problema constante em solucoes de transcricao automatica de musica. Esse processo basicamente se da em processar um sinal de audio no dominio da frequencia, calcular o nivel de energia de cada nota correspondente a frequencia e somar as oitavas de cada uma das 12 notas. Tradicionalmente usa-se a transformada de fourier para tal tarefa desde meados de 1999 com Fujishima [9] e, desde entao, varios esforcos sao empreendidos para otimizar o Pitch Class Profile (PCP) construido via transformada discreta de fourier (DFT).

Varios estudos abordaram o problema de construcao de chroma feature em musicas. Tais estudos construiu solucoes embasadas em DFT ou STFT (Short Fast Fourier Transform) para identificacao de melodias: [19], [1], [2], [10], [21], [8] e [14], esse ultimo usando filtros particula para otimizacao. Ha tambem estudos de solucoes de chroma feature com foco em transcricao automatica de acordes via DFT: [12], [15], [11], [20], [5], [17], [7], [3], [4] e [13].

Tais estudos focaram solucoes adaptadas em DFT e pouco se tem solucoes alternativas para caracterizacao de notas. Ha tambem varias limitacoes citadas da DFT em trabalhos como [11]. O estudo de Mauch [18] por exemplo foca a aplicacao de NNLS para aumentar a identificacao de notas e o trabalho [23] foca a modelagem matematica para a representacao da distribuicao das notas em suas frequencias. Porem ha ainda a carencia de metodos alternativos em relacao a DFT para a construcao de chroma feature.

O presente trabalho foca a apresentacao de um metodo alternativo e mais eficiente que a DFT, no que tange a identificacao das notas, para a construcao do chroma feature. A alternativa Chroma Convolution Method (CCM), que eh uma operacao essencialmente de dominio temporal, possui caracteristicas particulares que otimizam a distincao de notas e timbres de instrumentos musicais. O CCM pode ser

usado no lugar da DFT ou STFT para extrair sons polifonicos com mais acuracia, identificando notas de diferentes instrumentos musicais.

Este paper esta organizado da seguinte forma: Section 2 descreve o uso, aspectos gerais e limitacoes da DFT; Section 3 apresenta a conceituacao do metodo proposto, caracteristicas e processo de construcao do chroma feature usando o CCM; Section 4 apresenta dois experimentos de comparacao dos metodos DFT e CCM; Section 5 delimita discussoes sobre os resultados da eficacia do CCM; Section 6 apresenta conclusoes e futuros trabalhos.

2. DISCRETE FOURIER TRANSFORM (DFT)

Um chroma feature ou um pitch class profile (PCP) tm sido usados quase exclusivamente como um front-end para o reconhecimento de acordes ou sistemas de extraco de melodias de audio gravado. Fujishima [9] desenvolveu um sistema de transcricao automatica de acordes em tempo real, onde ele deriva um chroma feature de 12 dimensoes a partir do DFT do sinal de audio. Desde entao a transformada discreta de fourier (DFT ou FFT) tem sido muito utilizada para a construcao das caracteristicas cromaticas (chroma feature) do audio. A funcao dessa transformada eh traduzir informacoes que estao em dominio temporal para dominio frequencial de tal forma a projetar, em bases ortonormais, o valor de cada componente senoidal presente no sinal tratado. Essa projecao se da pelo somatorio do produto interno das senoides (exponenciais complexas) pelo sinal [22].

Visto que essa transformada somente oferece informacoes em termos de frequencias, surgiu-se a necessidade de adapta-la para a visualizacao das variacoes das frequencias ao longo do tempo. Essa adaptacao 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 dominio da frequencia, ou seja, a cada instante de tempo referente a cada trecho eh possivel obter uma analise frequencial.

Porem essa tecnica possui as seguintes limitacoes:

- surgimento de frequencias fantasmas (aliasing) 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);



© First author, Second author, Third author.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** First author, Second author, Third author. "Paper Template For ISMIR 2015", 16th International Society for Music Information Retrieval Conference, 2015.

- 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.

3. CHROMA CONVOLUTION METHOD (CCM)

Em vista das limitaes apresentadas da STFT, foi desenvolvido um metodo no dominio do tempo que preserva as caracteristicas originais do sinal. 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 so 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 possivel 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 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.

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

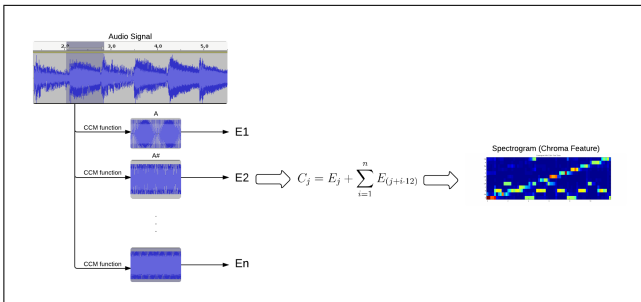


Figure 1. Schematic of process to build chroma feature using CCM.

4. EXPERIMENTS AND RESULTS

Com o intuito de demonstrar a eficacia do CCM em relacao ao tradicional metodo STFT [16] no que diz respeito 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 reais ¹.

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 2:



Figure 2. Notes that were played and recorded 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 5. 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.

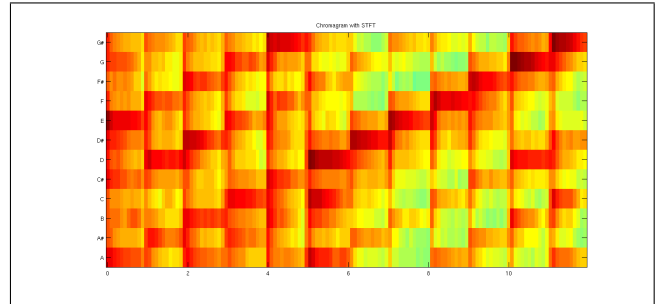


Figure 3. Chroma feature of audio using STFT.

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

Foi tambem feito uma analise de eficiencia das notas gravadas no audio com a sequencia de notas identificas no chroma feature. As notas identificadas foram selecionadas atraves da maior intensidade em cada instante de tempo. Segue tabela comparativa:

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 notes and identified notes.

Dado os dados da tabela Table 1, houve 9 de 14 notas tocadas em instantes corretos, ou seja, 64.2857% de fidelidade com a melodia original.

4.1.2 CCM Results

Segue chroma feature encontrado utilizando o CCM:

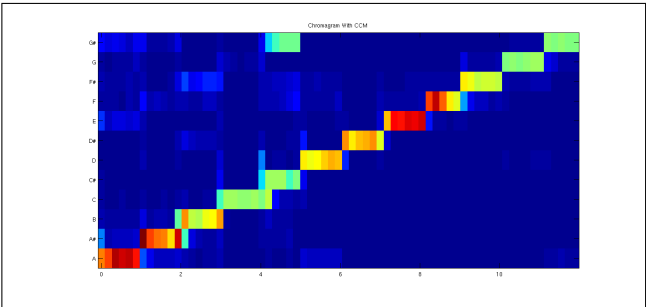


Figure 4. Chroma feature of audio using CCM.

Observa-se na figura Figure 4 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, 92.8571% de fidelidade com a melodia original.

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:

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 notes and identified notes from chroma feature.



Figure 5. 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:

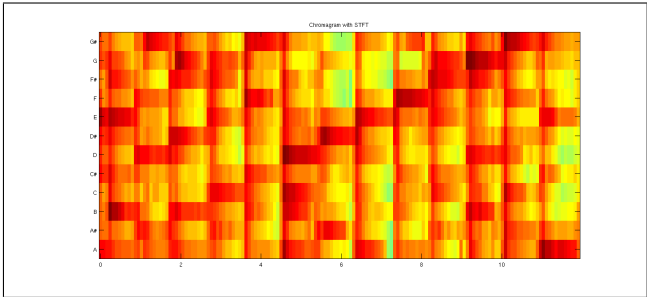


Figure 6. Chroma feature of audio using STFT.

Observa-se na figura Figure 6 que a escala cromatica do piano foi parcialmente identificada. As notas do violao sao praticamente imperceptiveis.

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:

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 notes of chroma feature.

Segundo os dados da tabela Table 3, houve 12 de 31 notas de piano tocadas em instantes corretos, ou seja, 38.7096% de fidelidade com a melodia original do piano. Quanto ao violao, somente 1 nota foi tocada em instante certo, ou seja, 3.2258% de fidelidade com a melodia original do violao.

4.2.2 CCM Results

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

Observa-se na figura Figure 7 que a melodia apresentou uma maior clareza. Eh visivel tambem a escala cromatica tocada pelo piano e algumas poucas notas do violao. Uma tabela comparativa das notas tocadas identificadas foi feita tambem atraves da estrategia dos maximos:

Em vista dos resultados de Table 4, houve 13 de 18 notas de piano em instantes corretos, ou seja, 72.2222% de fidelidade com a melodia original. Quanto ao violao, houve 5 de 18 notas em instantes corretos, ocasionando em

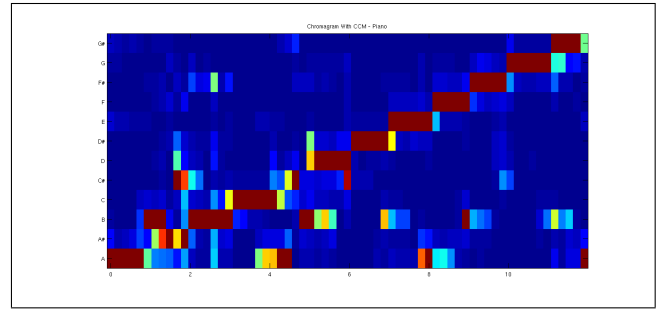


Figure 7. Chroma feature of audio using CCM of piano tones.

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 of chroma feature with piano tones.

27.7777%. Tal resultado mostra que o CCM usando tons de piano realmente amplificou as suas notas e suprimiu as do violao.

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

Observa-se na figura Figure 8 que a melodia apresentou uma maior clareza. Eh visivel tambem a escala cromatica tocada pelo piano suprimida e as notas do violao amplificadas. Uma tabela comparativa das notas tocadas e identificadas foi feita tambem atraves da estrategia dos maximos:

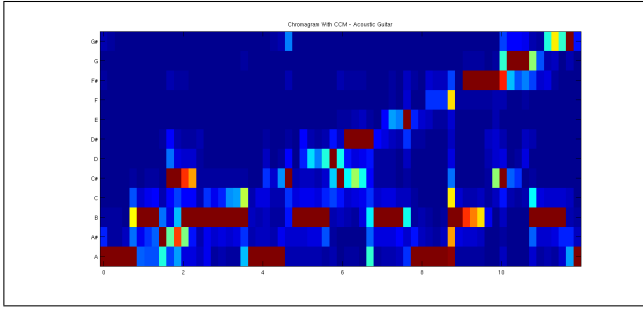


Figure 8. Chroma feature of audio using CCM of acoustic guitar tones.

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 and identified notes.

Em vista dos resultados de Table 5, houve 9 de 19 notas de violao em instantes corretos, ou seja, 47.3684% de fidelidade com a melodia original. Quanto ao piano, houve 2 de 19 notas em instantes corretos, ocasionando em 10.5263%. Tal resultado mostra que o CCM usando tons de violao realmente amplificou as suas notas e suprimiu as do piano.

5. DISCUSSION

Em vista dos resultados alcançados, ha limitacao e beneficios no uso da convolucao em relacao a transformada de fourier janelada. Uma limitacao encontrada foi o tempo de processamento que no metodo da STFT foi de 0.100138 segundos enquanto no CCM foi de 13.266711 segundos para o primeiro experimento.

Do ponto de vista dos beneficios, o primeiro experimento mostra que o CCM eh mais eficiente do que o STFT para caracterizacao de notas. O STFT obteve um rendimento de 64.2857% enquanto que o CCM obteve 92.8571%, ocasionando em 28.5714% a mais de notas identificadas com o CCM. Tal fato se deve pela reducao de ruidos e

intensificacao das frequencias sonoras realmente gravadas pelo instrumento musical.

Do segundo experimento observa-se um beneficio muito importante no uso do CCM, a possibilidade de distinguir instrumentos musicais tocados juntos. Alem da STFT nao possuir a possibilidade de geracao de um chroma feature para cada instrumento, ela gerou resultados baixos em comparacao com CCM na identificacao das melodias do piano (38.7096%) e violao (3.2258%).

Usando o CCM gerou-se, para cada instrumento, um chroma feature. O chroma feature do piano possuiu identificacao de 72.2222% de notas de melodia de piano e 27.7777% de melodia de violao. Esse fato demonstra que o chroma feature do piano realmente destaca as suas notas de timbre caracteristico. O chroma feature do violao tambem possuiu tal capacidade com 47.3684% de fidelidade para melodia do violao e 10.5263% para melodia de piano. Eh notavel que o CCM possui um potencial maior de identificacao que o STFT para ambos instrumentos alem de destacar mais diferentes instrumentos.

6. CONCLUSIONS AND FUTURE WORKS

Em vista do que foi proposto, foi apresentado uma nova forma de construçao da chroma feature usando o chroma convolution method (CCM). Comparou-se o metodo proposto com o tradicional STFT e o CCM mostrou-se mais satisfatorio.

A eficiencia do CCM frente ao STFT nao foi somente pela deteccao de melodias monofonicas, mas tambem pela identificacao de melodias polifonicas com mais de um instrumento. O CCM possui a capacidade diferencial de trabalhar com timbres de instrumentos musicais para a distincao dos mesmos no audio.

Pode-se concluir que melhorias que foram feitas na forma simples da STFT em trabalhos correlatos podem ser aplicadas no CCM, aumentando assim seu potencial. Tambem eh de significativa importancia demonstrar a otimizacao dos resultados de sistemas de transcriçao automatica de acordes usando o CCM.

Assim como recorrente com o uso da STFT, plausivel de implementacao de filtros para reduo de ruido e equalizacao no CCM, pois tais resultados expostos nao possuem nenhum tipo de pre-processamento de audio. Eh passivel de ressaltar que outros estudos comparativos devem ser feitos para conhecer e estabelecer as limitacoes e a eficacia do CCM.

7. REFERENCES

- [1] Majid A Al-Tae, 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 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.