Danilo Makoto Ikuta

Extração de Componentes de Arquivos de Música

São Paulo – Brasil

2016

Danilo Makoto Ikuta

Extração de Componentes de Arquivos de Música

Centro Universitário Senac

Orientador: Gabriel de Faria Andery

São Paulo – Brasil
2016

Resumo

Segundo a ABNT (2003, 3.1-3.2), o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (...) As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Palavras-chaves: latex. abntex. editoração de texto.

Abstract

This is the english abstract.

Key-words: latex. abntex. text editoration.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

Fig.	Area of the i^{th} component
456	Isto é um número
123	Isto é outro número
lauro cesar	este é o meu nome

Sumário

1 Introdução

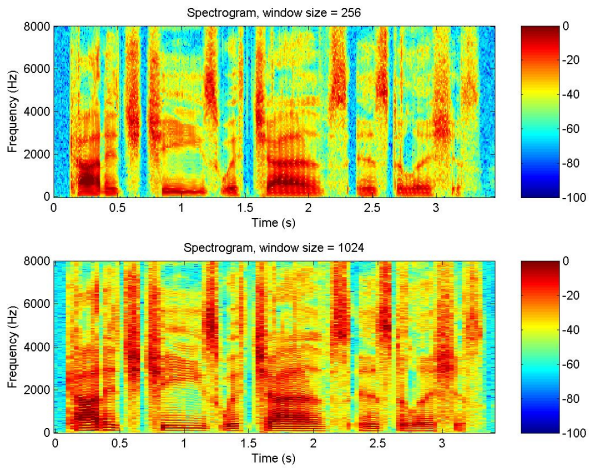
Cifras, partituras, covers, vídeo aulas e outras fontes podem auxiliar um músico a aprender a tocar uma determinada música em um instrumento musical. Para músicas mais conhecidas, tal material de apoio pode ser encontrado em livros, revistas ou na internet. Entretanto, esse material pode não existir ou estar em outra língua, dependendo da música e do instrumento que se deseja aprender, podendo exigir certa habilidade para discernir instrumentos e padrões dentro da música.

Para músicos menos experientes, aprender uma música utilizando-se apenas da habilidade auditiva("de ouvido") pode ser uma tarefa árdua. As dificuldades podem incluir a identificação de acordes, identificação de um instrumento ou melodia em músicas onde há múltiplos instrumentos, simultaneamente, e identificar qual ou quais notas estão sendo tocadas em dado momento da música. Possíveis formas de auxiliar esse processo de aprendizado incluem a extração de uma certa característica e a exclusão de características indesejadas.

O campo de recuperação de informação musical(MIR, do inglês Music Information Retrieval) trata, dentre outros assuntos, da identificação e extração de instrumentos(fontes sonoras), melodia, percussão, acompanhamento entre outras características de uma música.

Para a análise de áudio, costuma-se usar o espectrograma(figura 1), que representa o som através de um gráfico frequência X tempo, onde cada coordenada possui uma certa intensidade (amplitude).

Figura 1 – Exemplo de espectrograma. Fonte: http://www.ee.columbia.edu/~csmit/ELEN_E6820/images/assignment_1/specgrams.jpg



Com a finalidade de auxiliar músicos menos experientes, o desenvolvimento de

uma aplicação capaz de gerar arquivos de áudio derivados da análise e extração de componentes de uma música, como um instrumento específico, vozes, percussão, melodia e acompanhamento é descrito.

Serão comparadas técnicas tanto voltadas à instrumentos e características específicos quanto métodos com foco na abrangência de instrumentos e características.

Dentre os desafios a serem ponderados, estão a relação entre componentes harmônicos e de ruído dos instrumentos e a interação entre múltiplos instrumentos em um curto intervalo de tempo, no qual a frequência e/ou timbre pode afetar a qualidade da análise.

2 Revisão Bibliográfica

As técnicas descritas neste capítulo podem ser divididas nas seguintes categorias:

2.1 Obtenção do Espectrograma

2.1.1 STFT(Short-time Fourier Transform)

Segundo Roads et al. (1995), um espectrograma é uma forma de se visualizar informações sonoras no decorrer do tempo , em um gráfico de tempo e frequência, onde uma escala de cor representa a amplitude(intensidade) em cada intervalo de frequência e tempo. Atualmente, a forma mais comum de se calcular o espectrograma é utilizando a transformada de Fourier de tempo curto(STFT, Short-time Fourier Transform).

Para Fourier(Ivor (1995)), qualquer sinal periódico, independente da complexidade, pode ser representado como uma somatória de diversas ondas simples(em especial ondas senoidais), variando em frequência, amplitude e fase. Para a análise de sinais digitais de áudio, a STFT(shourtt-time Fourier transform) é utilizada.

A STFT consiste na separação do sinal de entrada em janelas(sogmentos) de tempo, geralmente sobrepostos e análise individual de cada janela. Para um sinal de entrada $x[m]$ com duração M, a saída $X[l, k]$ é a transformada de Fourier em cada quadro

Categoria	Técnicas
Obtenção do espectrograma	STFT MR-FFT
Pré-processamento	Somatória Sub-harmônica MFCC NMF
Obtenção de Informações do Espectrograma e Outras Fontes	Estimativa de Tremolo/Vibrato HCS Estimativa de F0 Cálculo de ESIs Matrizes de Similaridade Modelo Universal de Voz
Refinamento das Informações Obtidas	Rastramento Harmônico Extração de Contorno Mais Provável HMM Geração de Máscara Tempo-frequência
Pós-processamento	Suavização de Contorno

Tabela 1 – Relação de categorias e técnicas

l e cada intervalo de frequência k. Segue-se a equação:

$$X[l,k] = \sum_{m=0}^{M-1} h[m]x[m+lH]e^{-j(2\pi/N)km}, \tag{2.1}$$

onde $h[m]$ é a janela de seleção da entrada $x[m]$, l é o índice do quadro e H o avanço de tempo em amostras.As frequências de cada índice k são dadas pela relação

$$f_k = (k/N) \times f_s, \tag{2.2}$$

onde f_s corresponde à frequência de amostragem e N ao comprimento da janela em amostras.

2.1.2 MR-FFT(Multi-resolution Fourier Transform)

Proposto por Dressler (2006) e utilizado em Hsu e Jang (2010), tem como foco o uso de múltiplas resoluções dependendo da faixa de resolução analisada.

2.2 Pré-processamento

2.2.1 Somatória Sub-harmônica

O método de somatória sub-harmônica, proposto por Hermes (1986) e implementado em Cao et al. (2007) e Hsu et al. (2009) representa cada quadro com as estruturas harmônicas realçadas, em especial as provenientes de vocais. O algoritmo é dado por:

$$H_t(f) = \sum_{n=1}^{N} h_n P_t(nf), \tag{2.3}$$

onde $H_t(f)$ é o valor de somatória sub-harmônica da frequência f no quadro t, $P()$ é o espectro calculado via STFT, N a quantidade dos n componentes harmônicos considerados e h_n é o peso do n-ésimo componente harmônico.

2.2.2 MFCC(Mel-Frequency Cepstral Coefficients)

Hsu et al. (2009)

2.2.3 NMF(Fatorização de Matriz Não-negativa)

Rafii et al. (2013)

2.3 Obtenção de Informações do Espectrograma e Outras Fontes

2.3.1 Estimativa de Tremolo/Vibrato

Hsu e Jang (2010) e Regnier e Peeters (2009) partem do princípio de que poucos instrumentos, além da voz humana, apresentam simultâneamente ambos vibrato e tremolo (VERFAILLE; GUASTAVINO; DEPALLE, 2005). Vibrato se refere à modulação de frequência(FM) ou a micro-variação de tom, ao passo que o tremolo refere-se à modulação de amplitude(AM) ou micro-variações na intensidade. Considerando a taxa de vibrato/tremolo em torno de 6 Hz para a voz humana, calcula-se as amplitudes(extensão) de tremolo/vibrato para cada parcial p_k, com duração de t_i até t_j. Para o cálculo do vibrato, a transformada de Fourier das frequências f_{p_k} é:

$$F_{p_k}(f) = \sum_{t=t_i}^{t_j} (f_{p_k}(t) - \mu_{f_{p_k}}) e^{-2i\pi f \frac{t}{L}}, \tag{2.4}$$

onde $\mu_{f_{p_k}}$ é a frequência média da parcial $p_k(t)$ e $L = t_j - t_i$. Em seguida, é calculada a extensão relativa:

$$\Delta f_{relp_k}(f) = \frac{F_{p_k}(f)}{L\mu_{f_{p_k}}}, \tag{2.5}$$

enfim, calcula-se a extensão relativa entre 6Hz:

$$\Delta f_{p_k} = \max_{f \in [4,8]} \Delta f_{relp_k}(f). \tag{2.6}$$

Para o cálculo de tremolo, utiliza-se as mesmas equações para o cálculo de vibrato, mas substituindo f_{p_k} por a_{p_k}

2.3.2 HCS(Harmonic Coded Structure)

O HCS é utilizado por Joo et al. (2011) no processo de encontrar os tons dominantes da melodia em um espectrograma. Uma vez que as amplitudes harmônicas variam com o instrumento e tom citar fonte?, fez-se necessário construir um codebook a partir de amostras de notas em diversos instrumentos e voz humana, agrupando-as de acordo com o algoritmo proposto em (PELLEG; MOORE, 2000).

Calculam-se os HCSs para cada possível frequência fundamental(F0) seguindo as equação:

$$h_\eta[n] = w[n] \sum_{m=1}^{H} b_m \cos(m \cdot 2\pi\eta \cdot n + \phi_m), H = \lfloor \frac{f_s}{2\eta} \rfloor, \tag{2.7}$$

onde f_s é a frequência de amostragem, η a frequência fundamental(F0) do HCS, $w[n]$ a janela de análise, b_m a amplitude da m-ésima harmônica, e ϕ_m a fase de b_m.

2.3.3 Estimativa de Frequências Fundamentais(F0)

2.3.4 Cálculo de ESIs

Hsu et al. (2009), Hsu e Jang (2010)

2.3.5 Matrizes de Similaridade

Rafii e Pardo (2012)

2.3.6 Modelo Universal de Voz

Rafii et al. (2013)

2.4 Refinamento das Informações Obtidas

2.4.1 Rastreamento Harmônico

Cao et al. (2007) (inclui estrutura harmônica estável)

2.4.2 Extração do contorno/estrutura mais provável

Hsu e Jang (2010), Cao et al. (2007), Joo et al. (2011)

2.4.3 HMM(Modelos Ocultos de Markov)

Hsu et al. (2009)

2.4.4 Geração de Máscara Tempo-frequência

Rafii e Pardo (2012), Driedger, Müller e Disch (2014)

2.5 Pós-processamento

2.5.1 Suavização de Contorno

Joo et al. (2011)

3 Metodologia

3.1 Análise e Escolha das Técnicas e Combinações

Citar e jsutificar as técnicas e quais combinações serão implementadas para teste.

3.2 Desenvolvimento

Mostrar o desenvolvimento da aplicação e técnicas utilizadas.

4 Próximos Passos

Como próximos passos estão a pesquisa de mais alguns artigos, com vista na verificação de outras técnicas e relevância das técnicas obtidas. Além disso, deve-se estabelecer as medidas de desempenho para os testes, implementar e analisar diferentes combinações de técnicas e apontar possíveis melhorias.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 6028*: Resumo - apresentação. Rio de Janeiro, 2003. 2 p. Citado na página 2.

CAO, C. et al. Singing melody extraction in polyphonic music by harmonic tracking. In: *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, Vienna, Austria, September 23-27, 2007*. [s.n.], 2007. p. 373–374. Disponível em: <http://ismir2007.ismir.net/proceedings/ISMIR2007_p373_cao.pdf>. Acesso em: 21.05.2016. Citado 2 vezes nas páginas 11 e 13.

DRESSLER, K. Sinusoidal extraction using an efficient implementation of a multi-resolution fft. In: *Proc. of 9th Int. Conf. on Digital Audio Effects (DAFx-06)*. [s.n.], 2006. p. 247–252. Disponível em: <http://www.dafx.ca/proceedings/papers/p_247.pdf>. Acesso em: 12.08.2016. Citado na página 11.

DRIEDGER, J.; MÜLLER, M.; DISCH, S. Extending harmonic-percussive separation of audio signals. In: *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014, Taipei, Taiwan, October 27-31, 2014*. [s.n.], 2014. p. 611–616. Disponível em: <http://www.terasoft.com.tw/conf/ismir2014-/proceedings/T110_127_Paper.pdf>. Acesso em: 13.05.2016. Citado na página 13.

HERMES, D. J. Measurement of pitch by subharmonic summation. *Journal of Acoustic of Society of America*, v. 83, p. 257–264, 1986. Citado na página 11.

HSU, C. et al. Singing pitch extraction from monaural polyphonic songs by contextual audio modeling and singing harmonic enhancement. In: *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009, Kobe International Conference Center, Kobe, Japan, October 26-30, 2009*. [s.n.], 2009. p. 201–206. Disponível em: <http://ismir2009.ismir.net/proceedings/PS2-2.pdf>. Acesso em: 20.05.2016. Citado 2 vezes nas páginas 11 e 13.

HSU, C.; JANG, J. R. Singing pitch extraction by voice vibrato / tremolo estimation and instrument partial deletion. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands, August 9-13, 2010*. [s.n.], 2010. p. 525–530. Disponível em: <http://ismir2010.ismir.net-/proceedings/ismir2010-89.pdf>. Acesso em: 18.05.2016. Citado 3 vezes nas páginas 11, 12 e 13.

IVOR, G. *Joseph Fourier: 1768-1830*. [S.l.]: The MIT Press, 1995. Citado na página 10.

JOO, S. et al. Melody extraction based on harmonic coded structure. In: *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA, October 24-28, 2011*. [s.n.], 2011. p. 227–232. Disponível em: <http://ismir2011.ismir.net/papers/PS2-10.pdf>. Acesso em: 17.05.2016. Citado 2 vezes nas páginas 12 e 13.

PELLEG, D.; MOORE, A. W. X-means: Extending k-means with efficient estimation of the number of clusters. In: *Proceedings of the Seventeenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. (ICML '00), p. 727–734. ISBN 1-55860-707-2. Disponível em: <http://dl.acm-.org/citation.cfm?id=645529.657808>. Acesso em: 09.08.2016. Citado na página 12.

RAFII, Z. et al. Combining modeling of singing voice and background music for automatic separation of musical mixtures. In: *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013, Curitiba, Brazil, November 4-8, 2013*. [s.n.], 2013. p. 41–46. Disponível em: <http://www.ppgia.pucpr-.br/ismir2013/wp-content/uploads/2013/09/63_Paper.pdf>. Acesso em: 15.05.2016. Citado 2 vezes nas páginas 11 e 13.

RAFII, Z.; PARDO, B. Music/voice separation using the similarity matrix. In: *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, Mosteiro S.Bento Da Vitória, Porto, Portugal, October 8-12, 2012*. [s.n.], 2012. p. 583–588. Disponível em: <http://ismir2012.ismir.net/event/papers/583-ismir-2012.pdf>. Acesso em: 15.05.2016. Citado na página 13.

REGNIER, L.; PEETERS, G. Singing voice detection in music tracks using direct voice vibrato detection. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. [s.n.], 2009. p. 1685–1688. Disponível em: <http://articles-.ircam.fr/textes/Regnier09b/index.pdf>. Acesso em: 08.06.2016. Citado na página 12.

ROADS, C. et al. *The Computer Music Tutorial*. [S.l.]: The MIT Press, 1995. Citado na página 10.

VERFAILLE, V.; GUASTAVINO, C.; DEPALLE, P. Perceptual evaluation of vibrato models. In: *Proceedings of Conference on Interdisciplinary Musicology (CIM05)*. [s.n.], 2005. Disponível em: <http://oicrm.org/wp-content/uploads/2012-/03/VERFAILLE_V_CIM05.pdf>. Acesso em: 08.06.2016. Citado na página 12.