

Software para Caracterização de Timbre através de Espectrogramas e Mel-espectrogramas

João Pedro S. Pereira,¹ Victor Fernandes,² Tadeu N. Ferreira³
 Escola de Engenharia - UFF, Niterói, RJ

O timbre pode ser definido como uma característica intrínseca que difere perceptivelmente um som de outro. A música está inserida de forma significativa em processos de digitalização. Técnicas ligadas ao MIR (*Music Information Retrieval*) são utilizadas nesse processo, e o timbre é uma característica que pode ser explorada. A proposta deste trabalho é a análise de sinais musicais por meio de um software livre. Uma interface gráfica de usuário (GUI) desenvolvida fornece espectrogramas e mel-espectrogramas, tendo em vista que os mesmos são muito importantes para caracterizar o timbre. O desenvolvimento da aplicação foi feita em MATLAB. Neste trabalho, a caracterização do timbre em sua qualidade é um objetivo, o que possibilita a diferenciação entre dois instrumentos de nomes distintos. Porém, o timbre também pode ser categorizado segundo a sua identidade, o que caracteriza instrumentos diferentes de um mesmo grupo [1].

Métodos de processamento digital de sinais foram utilizados para a análise dos áudios provenientes de instrumentos musicais. A interface gráfica desenvolvida é uma ferramenta que provê meios para análise de Fourier, através da *Fast Fourier Transform* (FFT), e também a extração dos recursos principais: espectrogramas e Mel-espectrogramas. Seu funcionamento está completamente descrito em [5]. Além disso, também estão disponíveis o banco de dados de áudio, apresentação visual do software, arquivo de download da aplicação, e código-fonte.

$$S(t, \omega) = |STFT(t, \omega)|^2. \quad (1)$$

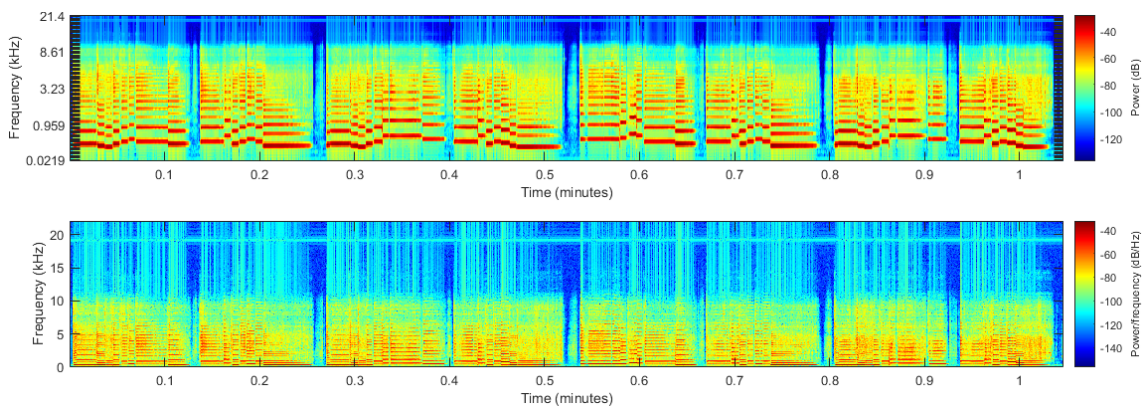


Figura 1: Extração de recursos: (a) MEL-espectrograma, (b) Espectrograma. Fonte: Autoral

¹jp_pereira@id.uff.br
²fernandesvictor@id.uff.br
³tadeu_ferreira@id.uff.br

O espectrograma $S(t, \omega)$ carrega a densidade espectral de potência de um sinal janelado [2]. É adquirido a partir da *Short Time Fourier Transform* (STFT), representada algebricamente pela equação (1), onde t denota tempo e ω é a frequência. Os parâmetros que mais se adequaram na extração dos espectrogramas dos sinais foram: *overlap size* de 1024, 2048 pontos na FFT, largura da janela de 2048, taxa de amostragem de 44.100 Hz e a janela *Hanning*. O objetivo na escolha dos parâmetros é possibilitar que a representação gráfica dos sinais tenha boa resolução temporal e frequencial, como representado na figura 1, pois o espectrograma é a representação visual da forma como a intensidade das diferentes frequências variam ao longo do tempo. Além disso, a definição de bons parâmetros para a STFT evitam um vazamento espectral (*leakage*) nos resultados.

A utilização de outra escala pode ajudar a visualização dos padrões formados no espectrograma, essa é uma das motivações para usarmos a escala melódica (MEL) na representação do espectrograma. Outra característica notável que a escala MEL possui, é a natureza perceptual da mesma, que é consoante a percepção auditiva humana. Desse modo, é possível imputar à escala MEL a capacidade de humanizar o espectrograma ou qualquer outra técnica de análise de áudio, a escala se adéqua à maneira que o ouvido humano percebe as nuances de frequência, pois a audição humana não percebe as frequências de maneira linear. Logo, a escala MEL se mostra adequada para representar frequências em aplicações de áudio, onde a percepção humana é fundamental.

O espectrograma, assim como sua representação em escala de frequência melódica, é eficaz em mostrar alguns comportamentos característicos de emissão dos instrumentos, porém, as características espectrais retiradas dos sinais podem descrever melhor a forma da magnitude do espectro. Tal conceito entra em concordância com o aspecto de envoltória do espectro, usado para caracterizar o próprio timbre [3]. A utilização conjunta de diversas características espectrais é interessante para caracterizar timbre do instrumento [4]. A figura 2 demonstra a característica *roll-off*, extraída do mesmo áudio dos espectrogramas apresentados na figura 1, tocado por saxofone um tenor.

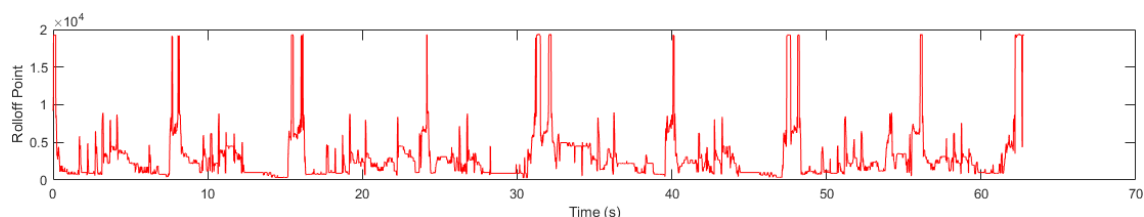


Figura 2: Característica espectral Roll-off. Fonte: Autoral

Referências

- [1] A. Lerch. **An introduction to audio content analysis: Applications in signal processing and music informatics**. Hoboken, New Jersey: Wiley-IEEE Press, 2012.
- [2] M. Müller. **Fundamentals of Music Processing**. Germany, Erlangen: Springer Cham, 2015.
- [3] M. Müller, D. P. W. Ellis, A. Klapuri e G. Richard. “Signal Processing for Music Analysis”. Em: **IEEE journal of selected topics in signal processing** 5.6 (2011), pp. 1088–1110.
- [4] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis e S. McAdams. “The timbre toolbox: Extracting audio descriptors from musical signals”. Em: **The Journal of the Acoustical Society of America** 130.5 (2011), pp. 2902–2916.
- [5] J. P. Pereira. **Repositório Github - Timbral analysis**. Online. Acessado em 02/02/2024, <https://github.com/jps-pereira/Timbral-Analysis.git>.