

Análise de características acústicas para identificação de *triggers* em gêneros e áudios musicais distintos

Marco Antonio dos Santos Fernandes, Estela Ribeiro, Carlos Eduardo Thomaz

¹ FEI, departamento de Engenharia Elétrica
Av. Humberto de Alencar Castelo Branco, 3972 – B – 09850-901, São Berardo do Campo, SP

marcoantoniosf2207@gmail.com, estela.eng@hotmail.com, cet@fei.edu.br

Resumo

Este trabalho analisa em detalhes uma base pública internacional de áudios musicais acerca das características acústicas presentes nas amostras existentes. Ao todo mil áudios classificados em dez gêneros musicais foram analisados em relação às suas características acústicas, com objetivo de buscar padrões nos *triggers* presentes nas amostras de cada gênero musical e a quantidade destes para aplicação em estudos subsequentes de ativações cerebrais geradas pela escuta destas músicas. Os resultados mostram que dentre os dez gêneros musicais presentes na base de áudios, dois desses (*Disco* e *Metal*) não apresentam número de *triggers* suficiente para tal aplicação.

1. Introdução

A música é complexa e possui vários gêneros musicais, sendo usada em diferentes situações. Em um festival, show ou reunião entre amigos a música se faz presente, podendo até ser o componente essencial para que o evento aconteça [1]. Ao escolher o tipo de música que gostamos, realizamos implicitamente julgamentos que correspondem às nossas preferências musicais [2, 3].

Um áudio musical possui diversas características [4] que são estímulos para o cérebro, tais como harmonia, melodia e ritmo, gerando ativações cerebrais. Estudos recentes utilizando estímulos sonoros [5, 6, 7, 8] surgiram para compreender as respostas neurológicas geradas durante a escuta musical em instantes denominados de *triggers* [5,6].

Este trabalho estende as metodologias de identificação de *triggers* [5] e de análise das características acústicas [6, 7, 8] em uma base de áudios composta por trechos de músicas de gêneros variados para averiguar a presença de *triggers* nestas amostras e verificar a possibilidade de utilização desses áudios em estudos de ativações cerebrais pertinentes.

2. Metodologia

A metodologia pode ser dividida em três etapas: (I) Extração de características acústicas; (II) Extração de

triggers; (III) Análise dos *triggers*.

Para primeira etapa, utiliza-se a base de áudios pública e internacionalmente conhecida GTZAN [4] que contém mil trechos musicais de trinta segundos de duração, contendo dez gêneros musicais com cem músicas cada. Os gêneros musicais estudados são: (1) *Blues*, (2) *Música Clássica*, (3) *Country*, (4) *Disco*, (5) *Hip-Hop*, (6) *Jazz*, (7) *Metal*, (8) *Pop*, (9) *Reggae* e (10) *Rock*.

Destes áudios foram extraídas características acústicas de baixo nível [9, 10] capazes de gerar respostas neurais significativas [5, 6]. As características extraídas são: (1) *Root Mean Square Energy* (RMS), (2) *Zero Crossing Rate* (ZCR), (3) *Brightness*, (4) *Spectral Centroid*, (5) *Spectral Spread*, (6) *Spectral Flatness*, (7) *Spectral Skewness*, (8) *Spectral Kurtosis*, (9) *Spectral Rolloff*, (10) *Spectral Flux*, (11) *Spectral Roughness* e (12) *Spectral Entropy*. A ferramenta utilizada para extração das características foi a MIRtoolbox (versão 1.7.2) [11], no software Matlab R19a. Os sinais das características exploradas foram decompostos em janelas de 50 ms de duração com 50% de sobreposição [6, 7, 8].

Na segunda etapa, são extraídos os *triggers*, que são instantes no áudio em que há estímulo capaz de gerar respostas neurais significativas (ou potencial evocado) e ocorrem por mudanças acentuadas nas características acústicas [5]. Segundo a metodologia de Poikonen et al. (2016) [5] de identificação de *triggers* é necessário determinar os limiares inferiores e superiores (V_{inf} e V_{sup}) que são uma porcentagem do sinal que está acima e abaixo da sua média, definida aqui como 20%. Quando o sinal permanece abaixo de V_{inf} por um período mínimo chamado *Preceding Low-Feature Phase* (PLFP) e em seguida ocorre um crescimento rápido, ultrapassando o limiar superior V_{sup} chamado de Magnitude de Rápido Crescimento (MoRI), há a ocorrência de um *trigger* ao final desta sequência, como mostra a Figura 1. Os valores de PLFP e MoRI são de 500 ms e 150 ms, respectivamente [6, 8].

Na análise dos *triggers*, terceira e última etapa, é realizada a detecção de padrões entre a presença de *triggers* nos diferentes gêneros, estendendo a metodologia de Ferreira et al. (2019) [7] e Ribeiro et al. (2020, 2019) [6,8] das características acústicas, obtendo a quantidade de *triggers* por amostra de áudio, permitindo a sua avaliação quantitativa.

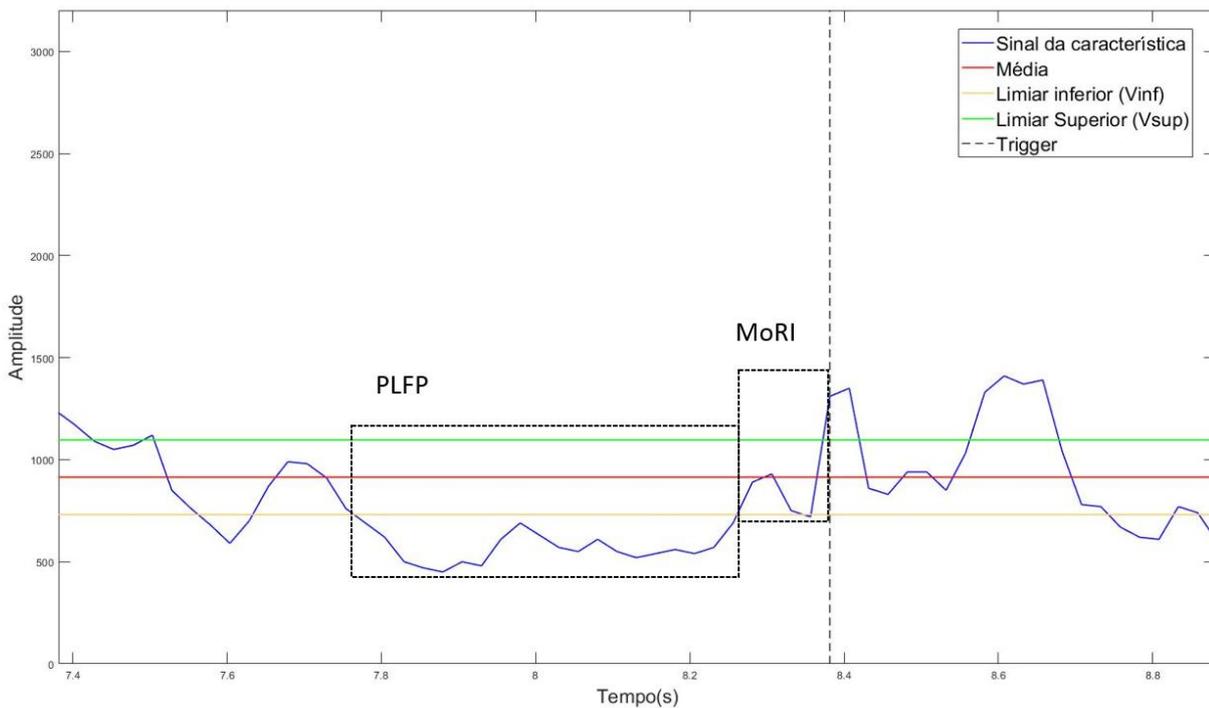


Figura 1: Exemplo de *trigger*.

3. Resultados

Há três agrupamentos entre as características. O primeiro agrupamento é composto por *RMS*, *Roughness* e *Spectral Flux*; o segundo por *Skewness* e *Kurtosis*; O terceiro por *ZCR*, *Spectral Rolloff*, *Brightness*, *Spectral Entropy*, *Spectral Flatness*, *Spectral Centroid* e *Spectral Spread*.

Os agrupamentos foram encontrados através de Análise Fatorial que indicou redundância entre as características acústicas, então foram selecionadas as características que possuíam maior carga fatorial dentro de cada agrupamento, sendo essas: *RMS*, *Spectral Rolloff* e *Spectral Skewness*. Devido ao fato de cada amostra de áudio possuir apenas trinta segundos de duração, estas amostras não apresentam grandes quantidades de *triggers*. A dispersão de *triggers* para estas características selecionadas é importante para que seja possível identificar a quantidade de músicas que podem ser usadas, conforme mostram as Figuras 2, 3 e 4.

A característica *Root Mean Square Energy* (*RMS*) possui poucas amostras válidas para os gêneros *Disco* e *Metal*. *Disco* possui apenas nove músicas com pelo menos dois *triggers* e duas músicas com um *trigger*, logo há poucas amostras que possuem instantes significativos para análise de resposta neural. O gênero *Metal* apresenta números um pouco melhores, sendo dez áudios com pelo menos dois *triggers*. Os demais gêneros musicais apresentam mais de vinte músicas com mais de dois *triggers*, logo oferecem quantidade mínima e suficiente de informação para estudos envolvendo respostas neurais. Alguns áudios apresentam muitos *triggers*, como é o caso de “blues.00026.wav” que possui quinze *triggers*, e áudios

com treze *triggers*, que são “blues.00011.wav” e “disco.00080.wav”, que é um ponto discrepante do gênero *Disco* que tem 89 áudios sem *triggers* RMS.

Spectral Rolloff foi a característica dentre as principais que apresentou menor número de *triggers*. Os gêneros *Disco* e *Metal* possuem dois e cinco áudios, respectivamente, que têm um *trigger*, apenas. Os outros gêneros apresentam pelo menos quinze áudios diferentes com no mínimo um *trigger*, e há casos de um único áudio com mais de dez *triggers* nos gêneros *Rock* e *Country*.

Todos os gêneros possuem ao menos dez áudios com pelo menos um *trigger* detectado pela característica *Spectral Skewness*. O gênero *Classical* foi o gênero com menos amostras, sendo apenas onze que possuem apenas um *trigger*.

Estes resultados mostram que apesar de haver muitas amostras na base de áudios GTZAN, muitos arquivos não possuem *triggers* suficientes. Portanto, a utilização desses áudios em estudos envolvendo análise de ativações cerebrais não seria possível.

4. Conclusão

Este trabalho apresenta os resultados do mapeamento da base de áudios GTZAN para identificar quais das suas amostras possuem *triggers* das principais características acústicas, identificando que apesar de haver muitos trechos de áudio na base, os gêneros *Disco* e *Metal* possuem pouquíssimas amostras significativas em relação a conteúdo de *triggers* para a característica *Spectral Rolloff* que é a característica acústica mais significativa do terceiro agrupamento, portanto é muito importante em relação ao conteúdo acústico do agrupamento. Este achado é

importante para utilização desta base de áudios em estudos de ativações cerebrais motivadas por estas músicas, porque

a partir destes resultados é possível aproveitar apenas os áudios que possuem as melhores quantidades de *triggers*.

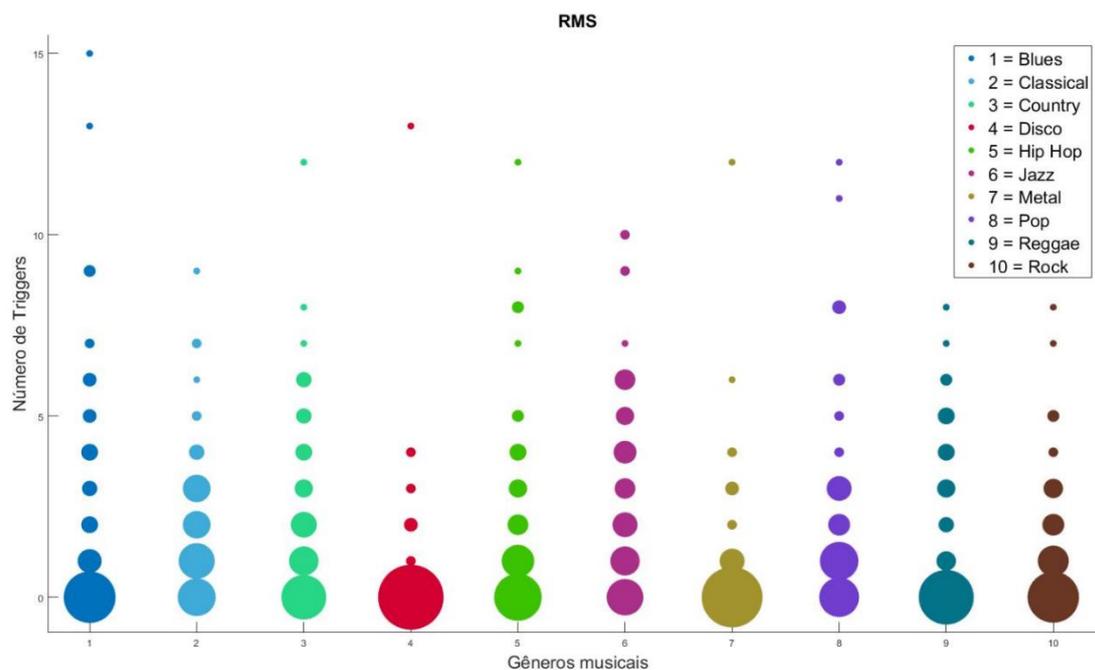


Figura 2: Distribuição de áudios por número de *triggers* – RMS.

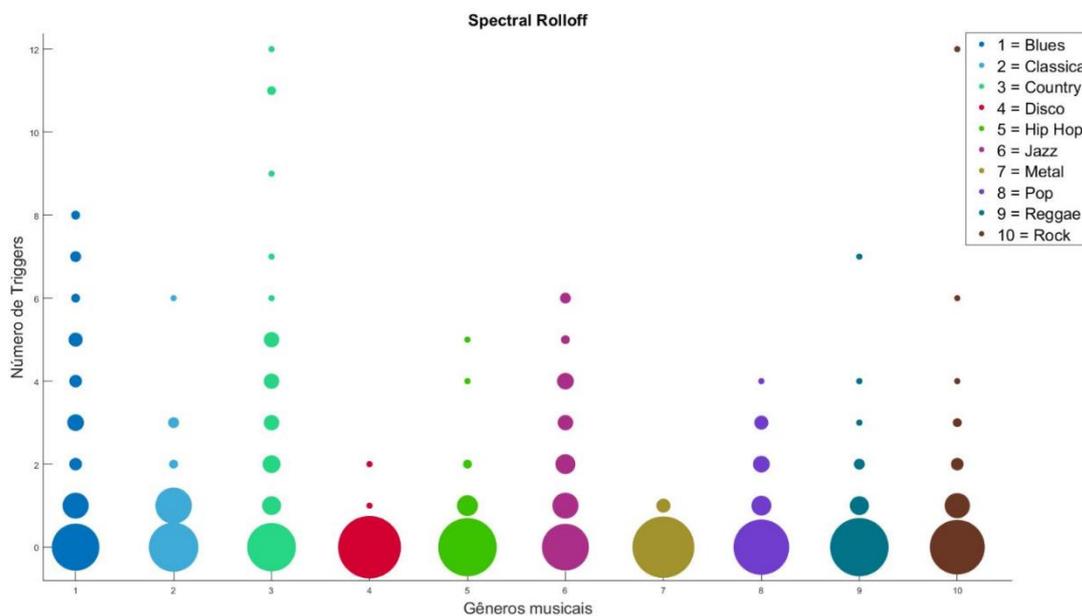


Figura 3: Distribuição de áudios por número de *triggers* – *Spectral Rolloff*.

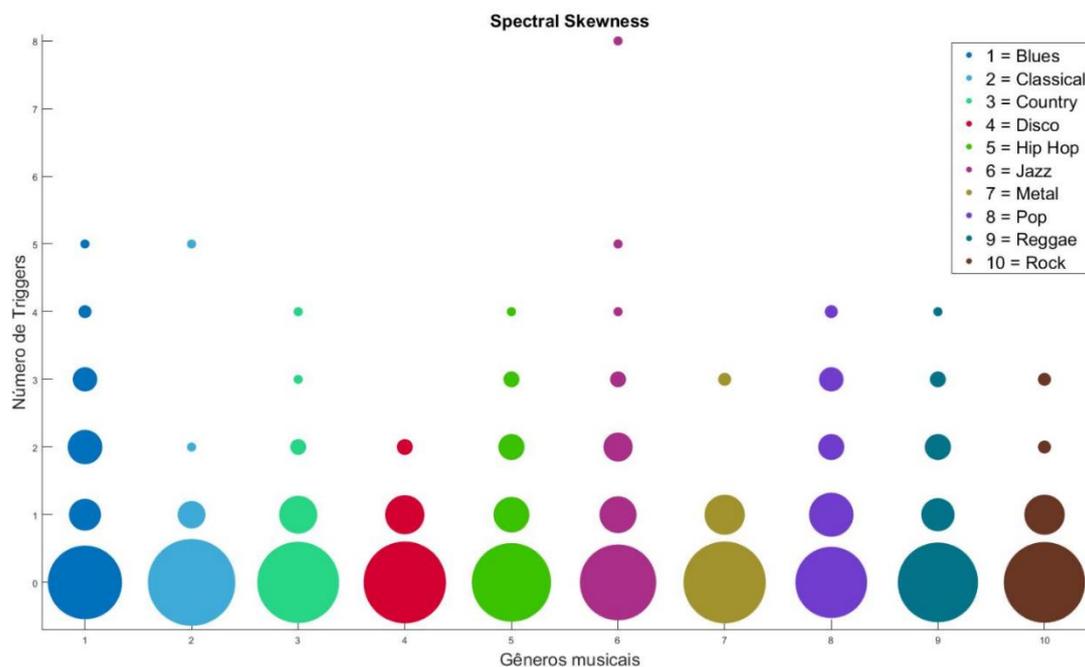


Figura 4: Distribuição de áudios por número de *triggers* – *Spectral Skewness*.

Agradecimentos

Os autores agradecem a bolsa FEI (PBIC006/20) concedida ao primeiro autor para realização desta pesquisa.

Referências

[1] RENTFROW, Peter J.; GOSLING, Samuel D. The do re mi's of everyday life: The structure and personality correlates of music preference. *Journal of Personality and Social Psychology*, v. 84, n. 6, p. 1236-1256, 2003.

[2] GREENBERG, David M.; et al. Musical Preferences are linked to cognitive styles. *PLOS ONE*, pp. 1-22, 2015.

[3] SOLEYMANI Mohammad; et al. Content-Based music recommendation using underlying music preference structure. *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6, 2015.

[4] TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, v. 10, n. 5, p. 293–302, jul. 2002.

[5] POIKONEN, H. et al. Event-related brain responses while listening to entire pieces of music. *Neuroscience*, v. 312, p. 58-73, jan. 2016.

[6] RIBEIRO, Estela; THOMAZ, Carlos E. Whole brain EEG analysis of musicianship. *Music Perception*, v.37, pp. 42-56, 2019.

[7] FERREIRA, Leonardo Antunes; RIBEIRO, Estela; THOMAZ, Carlos Eduardo. A cluster analysis of benchmark acoustic features on Brazilian music. In:

SBCM, pp. 1-3, 2019.

[8] RIBEIRO, Estela. ANÁLISE E RECONHECIMENTO DE PADRÕES COGNITIVOS EM ESCUTAS MUSICAIS E SONOROS EM ÁUDIOS. 2020. Dissertação (Doutorado em Engenharia Elétrica) – Centro Universitário FEI, São Bernardo do Campo.

[9] LERCH, Alexander. An introduction to audio content analysis. 1st ed. New Jersey: IEEE, 2012.

[10] KNESS, Peter; SCHEDL, Markus. Music similarity and retrieval: an introduction to audio and web based strategies. 1st ed. Heidelberg: Springer, 2016.

[11] LARTILLOT, Oliver. MIRtoolbox 1.7.2 users manual. Aalborg: Department of architecture, design e media technology, 2019.