Análise de padrões acústicos em áudios musicais para detecção de *triggers*

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

Departamento de Engenharia Elétrica
FEI – São Bernardo do Campo, SP – Brazil

Marcoantoniosf2207@gmail.com, estela.eng@hotmail.com, cet@fei.edu.br

Abstract. *This work analyzes in detail an international public database of musical audios about the acoustic features present in the existing samples. A total of one thousand audios classified in ten musical genres were analyzed in relation to their acoustic characteristics to look for patterns in the triggers present in the samples of each musical genre and the amount of these for application in subsequent studies of brain activations generated by listening to these songs. The results show that among the ten musical genres present in the audio database, two of these (Disco and Metal) do not have enough triggers for such an application.*

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 está presente ao longo da vida humana e a escuta musical é uma das atividades de lazer mais praticadas enquanto estamos acordados [Mehl & Pennebaker 2003]. 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 [Rentfrow & Gosling 2003]. Quando ouvimos uma música pela primeira vez, em alguns segundos decidimos entre trocá-la ou ouvi-la até o final [Istók et al. 2013]. Estudos indicam que as preferências musicais dos indivíduos revelam traços de sua personalidade e habilidades cognitivas [Greenberg et al. 2015]. Dessa forma, ao escolher o tipo de música que gostamos, realizamos análises implícitas que correspondem com as nossas preferências musicais [Istók et al. 2013; Greenberg et al. 2015].

Os gêneros musicais são utilizados para classificar músicas, e apesar de existirem padrões entre as músicas de um determinado gênero, esta classificação é vaga. Existem mais de mil gêneros musicais e muitos destes são considerados subgêneros, portanto há redundância entre esses [Soleymani et. al. 2015]. Isto motivou pesquisas com o objetivo de aprimorar a classificação de gêneros de forma automática que obtiveram acurácia similar à classificação feita por seres humanos [Tzanetakis e Cook 2002]. Indo além, uma

nova metodologia de recomendação musical foi proposta, não utilizando os gêneros já estabelecidos, mas buscando as características intrínsecas nos áudios. Estas características foram avaliadas por voluntários e por algoritmos de análise do sinal de áudio, então os resultados foram agrupados e uma nova ferramenta de recomendação foi desenvolvida [Soleymani et. al. 2015].

Um áudio musical possui diversas características que são estímulos para o cérebro, tais como harmonia, melodia e ritmo [Tzanetakis e Cook 2002], gerando ativações cerebrais. A percepção musical é o resultado de uma série de reações químicas, mecânicas e neurais que acontecem a partir do momento que os tímpanos captam um som e este é convertido em um sinal interpretado pelo cérebro [Peretz e Zatorre 2005].

Utilizando-se dessas características acústicas, estudos recentes [Poikonen et al. 2016; Ribeiro e Thomaz 2019; Ferreira, Ribeiro e Thomaz 2019; Ribeiro 2020] analisaram as atividades neurais geradas durante a escuta musical, por meio de equipamento de eletroencefalografia (EEG), baseando-se em instantes capazes de gerar respostas neurais significativas em instantes denominados de *triggers*. Estes estudos demonstram que é possível identificar potenciais evocados nos sinais de EEG por meio das métricas utilizadas para analisar as características acústicas dos áudios [Poikonen et al., 2016], assim como diferenciar músicos de não-músicos [Ribeiro e Thomaz 2019].

Este trabalho estende as metodologias de identificação de triggers [Poikonen et al. 2016] e de análise das características acústicas [Ribeiro e Thomaz 2019; Ferreira, Ribeiro e Thomaz 2019; Ribeiro 2020] em uma base de áudios composta por trechos de músicas de gêneros variados para averiguar a presença desses *triggers* nestas amostras e verificar a possibilidade de utilização desses áudios em estudos de ativações cerebrais pertinentes.

2. Métodos

A metodologia pode ser dividia em três etapas: (1) Extração de características acústicas; (2) Extração de *triggers*; (3) Análise dos *triggers*.

2.1. Extração de características acústicas

Para primeira etapa, utiliza-se a base de áudios pública e internacionalmente conhecida denominada GTZAN [Tzanetakis e Cook 2002] que contém mil trechos musicais de trinta segundos de duração, contendo dez gêneros musicais com cem músicas em cada um deles, sendo que os gêneros musicais presentes na base GTZAN são: (1) *Blues*, (2) Música Clássica, (3) *Country*, (4) *Disco*, (5) *Hip-Hop*, (6) *Jazz*, (7) *Metal*, (8) *Pop*, (9) *Reggae* e (10) *Rock*.

Segundo Lerch (2012) e Kness e Schedl (2016), as características acústicas podem ser separadas em níveis, sendo essas de baixo nível ou alto nível. As características de baixo nível são definidas de curta duração, gerando um valor para cada trecho de áudio analisado. Estas características não são diretamente relacionadas a música e seus aspectos, mas descrevem o sinal de áudio no domínio do tempo ou da frequência. A partir dessas características são construídas as características de alto nível, que podem descrever as propriedades musicais como melodia, harmonia, entre outras. De todos os áudios da base GTZAN foram extraídas características acústicas de baixo nível [Lerch 2012, Kness e Schedl 2016] capazes de gerar respostas neurais significativas [Poikonen et al. 2016, Ribeiro e Thomaz 2019]. Essas características estão especificadas na Tabela 1.

A ferramenta utilizada para extração das características foi a MIRtoolbox (versão 1.7.2) [Lartillot 2019], no software Matlab R19a. Os sinais das características exploradas foram decompostos em janelas de 50 ms de duração com 50% de sobreposição [Ribeiro e Thomaz 2019; Ferreira, Ribeiro e Thomaz 2019; Ribeiro 2020].

Tabela 1. Características acústicas utilizadas

Número	Característica acústica
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	*Sprectral Flux*
11	*Spectral Roughness*
12	*Spectral Entropy*

2.2. Extração de *triggers*

Na segunda etapa, são extraídos os *triggers*, que são instantes no áudio em que há estímulo capaz de gerar respostas neurais significativas e ocorrem por mudanças acentuadas nas características acústicas [Poikonen et al. 2016]. Segundo a metodologia de Poikonen et al. (2016) 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 *Preceeding Low-Feature Phase* (PLFP) e em seguida ocorre um crescimento rápido, ultrapassando o limiar superior V_{sup} chamado de Magnitude de Rápido Crescimento (MoRI). Um *trigger* ocorre após essa sequência, quando o sinal ultrapassa o limiar superior, como mostra a Figura 1. Os valores de PLFP e MoRI são de 500 ms e 150 ms, respectivamente [Ribeiro e Thomaz 2019, Ribeiro 2020].

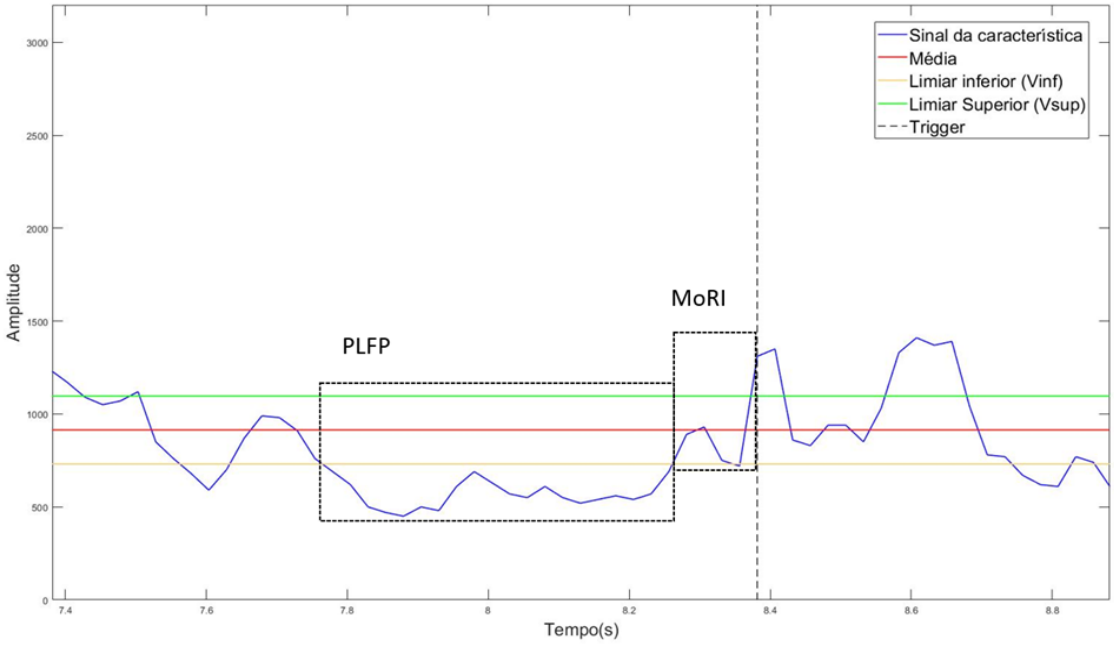

Figura 1. Exemplo de *trigger.*

2.3. Análise de *triggers*

A última etapa compreende a detecção de padrões entre a presença de *triggers* nos diferentes gêneros, obtendo a quantidade de *triggers* por amostra de áudio, permitindo a avaliação de cada áudio individualmente e dos gêneros musicais como um todo.

Estudos apontam que há redundância entre as informações extraídas das características acústicas [Ribeiro 2020]. Esta redundância foi determinada por meio de Análise Fatorial que permite reduzir a quantidade de variáveis utilizadas a partir da correlação encontrada entre as mesmas.

3. Resultados

A extração de *triggers* realizada nos gêneros musicais mostrou que as amostras de áudio investigadas não possuem muitos triggers. Isso se dá em grande parte por conta do comprimento dos áudios, que são de apenas 30 segundos, resultando em muitas amostras sem incidência de *triggers.* As características acústicas são calculadas no domínio do tempo ou da frequência, e o sinal de curta duração não gera muitas amostras que satisfaçam as regras descritas na seção 2.2 para detecção de um instante no tempo significativo. A Figura 2 apresenta os gráficos de distribuição da quantidade de *triggers* para cada característica acústica, em cada um dos dez gêneros musicais.

A Análise Fatorial apresentou três agrupamentos de acordo com as informações trazidas por cada característica. Logo não é necessário utilizar todas, mas as de maior carga fatorial de cada grupo [Ribeiro, Thomaz 2019, Ribeiro 2020]. Estas características são RMS no grupo 1 (formado por RMS, *Roughness* e *Spectral Flux*), *Skewness* no grupo 2 (composto por *Skewness* e *Kurtosis*) e *Spectral Rolloff* no grupo 3 (formado por ZCR, *Spectral Rolloff*, *Brightness, Spectral Entropy*, *Spectral Flatness*, *Spectral Centroid* e *Spectral Spread*). A Figura 3 mostra os agrupamentos encontrados por meio da Análise Fatorial.

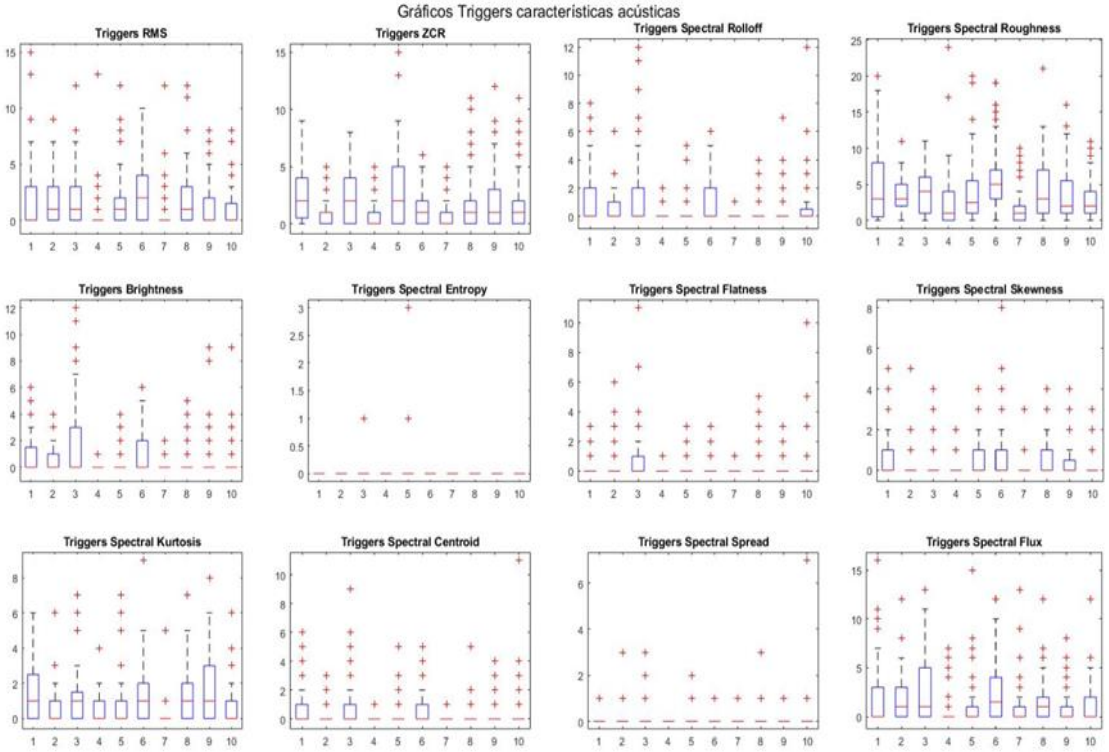

Figura 2. Distribuição de *triggers* de cada característica acústica.

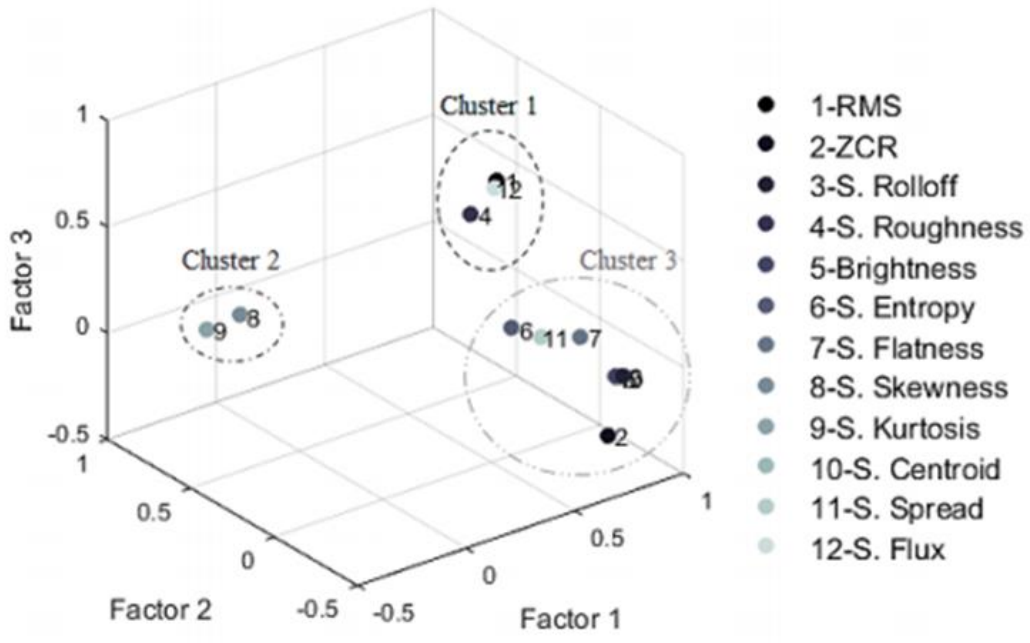

Figura 3. Agrupamento de características acústicas por carga fatorial.

A Figura 2 mostra que muitas características têm sua quantidade média de *triggers* em zero. Porém as Figuras 4, 5 e 6 apresentam a distribuição dos áudios para cada uma das características acústicas representantes dos três grupos citados anteriormente,

evidenciando que dentre as cem amostras de áudio, a maior parte dos gêneros musicais possui uma pequena quantidade de amostras que contém ao menos um *trigger*.

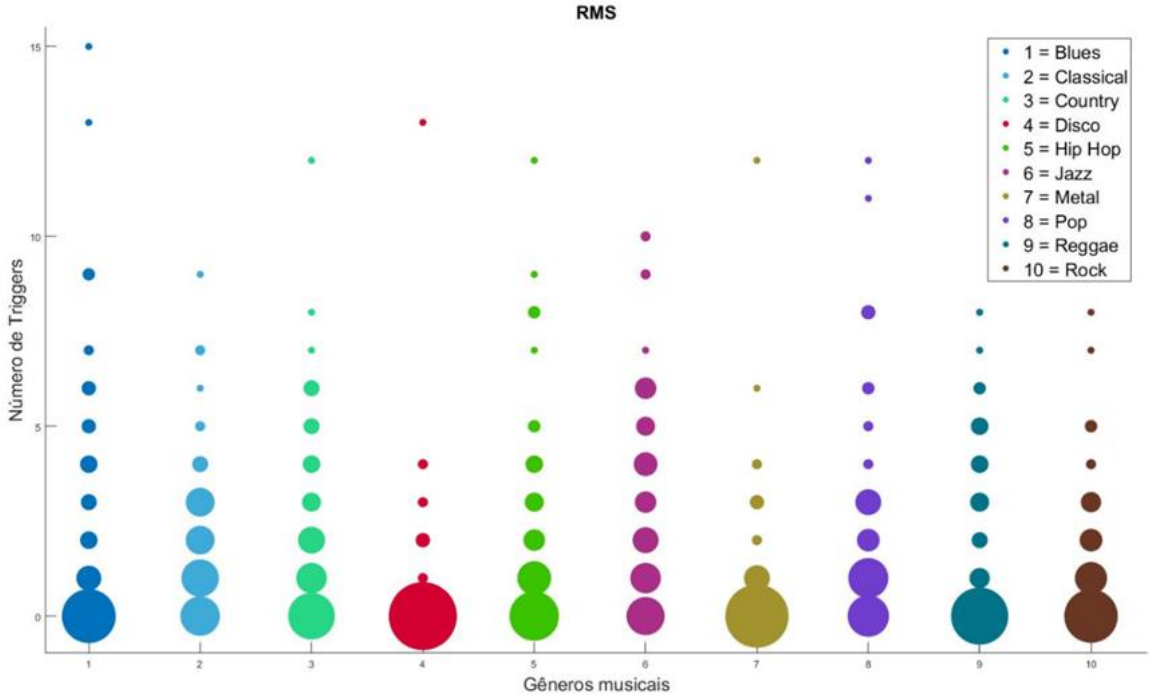

Figura 4. Distribuição de áudios por número de *triggers* – RMS.

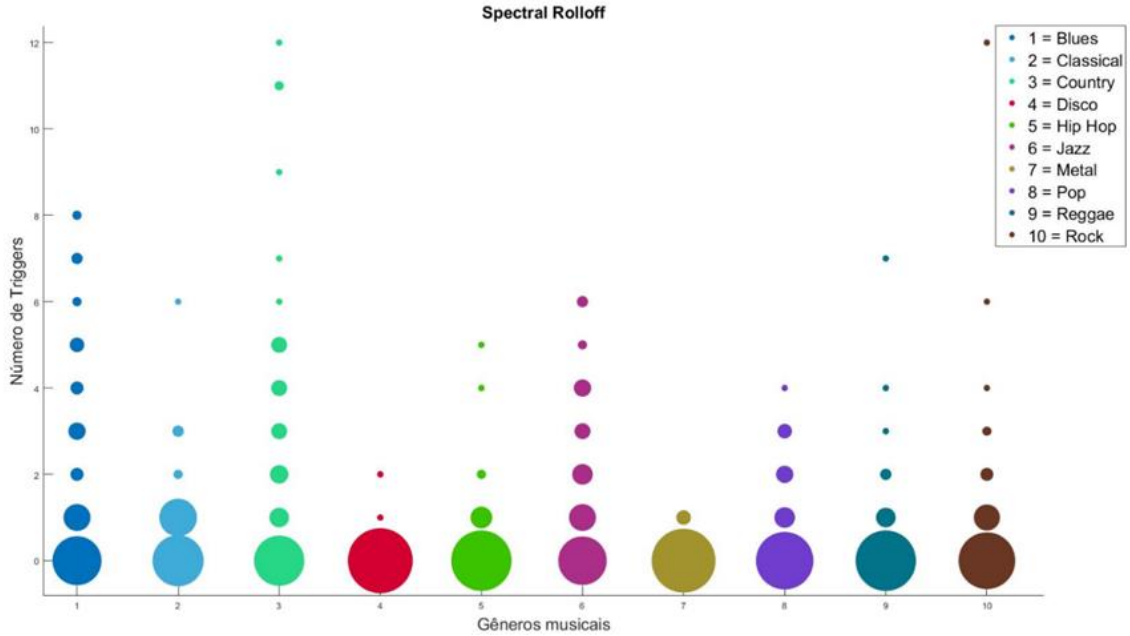

Figura 5. Distribuição de áudios por número de *triggers* – *Spectral Rolloff.*

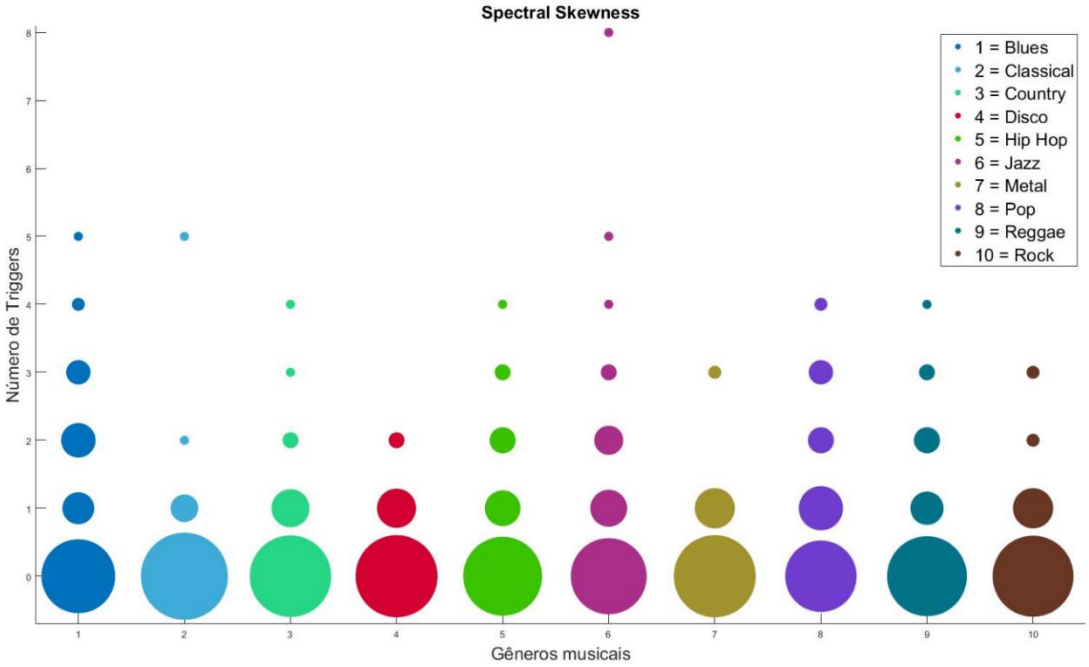

Figura 6. Distribuição de áudios por número de *triggers* – *Spectral Skewness*.

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 trechos musicais não possuem *triggers* suficientes. Portanto, a utilização desses áudios em estudos envolvendo análise de ativações cerebrais não seria possível. A classificação dos gêneros musicais a partir da quantidade de *triggers* também não foi possível, já que houve muitas amostras de diferentes gêneros que não apresentaram estes instantes ao longo do áudio.

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 ao conteúdo de *triggers* para a característica *Spectral Rolloff,* que é a característica acústica mais significativa do terceiro agrupamento, portanto 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 detectar previamente e avaliar apenas os áudios que possuem as melhores quantidades de *triggers*.

Agradecimentos

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

Referências

Ferreira, L. A., Ribeiro, E. and Thomaz, C. E. (2019) "A cluster analysis of benchmark acoustic features on Brazilian music", In: SBCM, pp.1–3.

Greenberg, D. M., et al. (2015) "Musical Preferences are linked to cognitive styles", In: PLOS ONE, pp.1–22.

Istók, E., et al. (2013) "'I love Rock n' Roll' – Music genre preference modulates brain response to music", In: Biological Psychology, v. 92, pp.142–151.

Lartillot, O. (2019) "MIRtoolbox 1.7.2 user's manual", Aalborg: Department of architecture, design e media technology.

Lerch, A. (2012) An introduction to audio content analysis, New Jersey: IEEE, 1st edition.

Kness, P. and Schedl, M. (2016), Music similarity and retrieval: an introduction to audio and web-based strategies, Heidelberg: Springer, 1st edition.

Mehl, M. R., Pennebaker, J. W. (2003) "The Sounds of Social Life: A Psychometric Analysis of Students' Daily Social Environments and Natural Conversations", In: Journal of Personality and Social Psychology, v. 84, n. 4, p.857–870.

Peretz, I., Zatorre, R. J., (2005) "Brain Organization for Music Processing", In: Annual Reviewe of Psychology, v. 56, n. 1, p.89–114.

Poikonen, H., et al. (2016) "Event-related brain responses while listening to entire pieces of music", In: Neuroscience, v. 312, p.58–73.

Rentfrow, P. J., Gosling, S. D. (2003) "The do re mi's of everyday life: The structure and personality correlates to music preference", In: Journal of Personality and Social Psychology, v. 84, n. 6, p.1236–1256.

Ribeiro, E. (2017) "UM ESTUDO SOBRE PREDIÇÃO DE MUSICALIDADE POR MEIO DA ANÁLISE DE SINAIS DE EEG", In: Dissertação de Mestrado em Engenharia Elétria, FEI, São Bernardo do Campo, Brasil.

Ribeiro, E., Thomaz, C. E., (2019) "Whole brain EEG analysis of musicianship", In:

Music Perception, v. 37, pp. 42–56.

Ribeiro, E. (2020) “ANÁLISE E RECONHECIMENTO DE PADRÕES COGNITIVOS EM ESCUTAS MUSICAIS E SONOROS EM ÁUDIO”, In: Tese de Doutorado em Engenharia Elétrica, FEI, São Bernardo do Campo, Brasil.

Soleymani, M., et al. (2015) “Content-Based music recommendation using underlying music preference structure”, In: IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6.

Tzanetakis, G., Cook, P. (2002) “Musical genre classification on audio signals”, In: IEEE Transactions on Speech and Audio Processing, v. 10, n. 5, p.293–302.