Less Acoustic Features Means more Statistical Relevance: Disclosing the Clustering Behavior in Music Stimuli

Estela Ribeiro and Carlos E. Thomaz

***Abstract*—Identification of appropriate content-based features for the description of audio signals can provide a better representation of naturalistic music stimuli which, in recent years, have been used to understand how the human brain processes such information. In this work, an extensive clustering analysis has been carried out on a large and benchmark audio dataset to assess whether features commonly extracted in the literature are in fact statistically relevant. Our results show that not all of these well-known acoustic features might be statistically necessary. We also demonstrate quantitatively that, regardless of the musical genre, the same acoustic feature is selected to represent each cluster. This finding discloses that there is a general redundancy among the set of audio descriptors used, that does not depend on a particular music track or genre, allowing an expressive reduction of the number of features necessary to identify apropriate time instants on the audio for further brain signal processing of music stimuli.**

***Index Terms*—music, cluster analysis, audio analysis.**

I. Introdução

Na área da Recuperação de Informação Musical (MIR - *Music Information Retrieval*) existem muitas características que podem ser extraídas dos áudios, servindo para diferentes propósitos como a classificação do gênero das músicas, similaridade musical, detecção de emoções, entre outros. O desenvolvimento de sistemas de recuperação de áudio requer uma melhor representação desses sinais, considerando a grande quantidade de características acústicas existentes, o que significa que é necessário ter uma identificação mais apropriada das características acústicas utilizadas [1], evitando o uso de descritores redundantes e selecionando aqueles mais relevantes para obtenção de um conjunto compacto e expressivo de descritores.

Uma aplicação emergente para estas características acústicas é utilizá-las como métricas que definem instantes no tempo para analisar os sinais cerebrais de voluntários. Historicamente, estudos que analisam os sinais cerebrais para compreender como o cérebro processa a música têm utilizado estímulos artificiais, como: diferentes sequências sonoras, tons puros e tons complexos, acordes maiores e menores, acordes consonantes e dissonantes [2], [3]. No entanto, estes estímulos representam apenas algumas das múltiplas dimensões que constituem o que chamamos de música e, portanto, o que esses estudos na realidade mostram é como o cérebro processa estes elementos isoladamente. Estudos mais recentes têm buscado contornar esse problema utilizando, como estímulos, gravações de músicas completas [2]–[5]. Neste caso é imperativo identificar instantes específicos para se analisar adequadamente esses sinais cerebrais.

Em mais detalhes, na ultima década, trabalhos apontam que é possível utilizar certas características acústicas para identificar instantes no sinal de áudio que são capazes de gerar respostas neurais significativas durante a escuta musical [2]–[7], contribuindo para o entendimento de como o cérebro humano processa a música. Assim, utilizando estes instantes para descrever a associação entre as mudanças dinâmicas dos áudios com as ativações neurais registradas por equipamentos de fMRI (Imagem por Ressonância Magnética funcional) e EEG (Eletroencefalografia) é possível, inclusive, indicar se uma determinada pessoa recebeu, ou não, treinamento musical em sua vida [5], [6]. Em particular, para sinais de EEG, o método para identificação destes instantes [4] baseia-se no conceito de *triggers*, que são instantes nas séries temporais geradas pelas características acústicas extraídas dos áudios em que ocorrem um alto contraste, determinados de acordo com o modelo proposto.

Neste contexto, esse artigo, estendendo trabalhos anteriores recentes [5], [8], investiga a possibilidade de se encontrar um comportamento de agrupamento (em inglês, *clustering*) entre características de baixo nível extraídas do sinal de áudio. Assim, é feita uma análise estatística multivariada de sinais de áudio de uma base de dados pública e de referência na literatura a fim de avaliar, de forma abrangente, se as características acústicas extraídas dos áudios podem ser consideradas estatisticamente não-redundantes. É levantada a hipótese de que esse comportamento de agrupamento seja similar entre músicas de mesmo gênero musical e considera-se aqui uma metodologia de avaliação exploratória de seleção dessas características acústicas.

II. MATERIAIS E MÉTODOS

A. Base de Áudio

Foi utilizada a base de dados GTZAN [9], [10], por ser pública e bem conhecida na área de MIR. Esta base consiste em 1000 trechos musicais (30 segundos cada) e contém 10 gêneros musicais com 100 músicas cada um, sendo esses: Blues, Classic, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae e Rock.

E. Ribeiro é pesquisadora colaboradora do Laboratório de Processamento de Sinais no Centro Universitário da FEI, São Paulo, CEP 09850-901 Brasil (e-mail: estela.eng@hotmail.com).

C. E. Thomaz é professor do Departamento de Engenharia Elétrica e coordenador do Laboratório de Processamento de Sinais da FEI, São Paulo, CEP 09850-901 Brasil (e-mail: cet@fei.edu.br).

B. Metodologia

Foram selecionadas 12 características acústicas de baixo nível de referência na literatura e utilizadas em outros trabalhos [2], [4], [5], [8], [11], sendo elas: (1) Root Mean Square (RMS) - medida da energia do sinal computada por meio da raiz quadrada média do quadrado da amplitude; (2) Zero Crossing Rate (ZCR) - medida do número de vezes que o sinal cruza o eixo do tempo; (3) Spectral Rolloff - frequência abaixo da qual 85% da energia total está contida no sinal; (4) Spectral Roughness - estimativa da dissonância sensorial; (5) Brightness - medida da quantidade de energia acima de 1500 Hz; (6) Spectral Entropy - medida relativa a Entropia de Shannon do sinal, indicando se o espectro contém picos predominantes ou não; (7) Spectral Flatness - medida da uniformidade do espectro, definido como a razão entre a média geométrica e a média aritmética do sinal, também conhecida como Entropia de Wiener; (8) Spectral Skewness - terceiro momento central da distribuição do espectro do sinal de áudio, relacionado com a assimetria da distribuição; (9) Spectral Kurtosis - quarto momento central da distribuição do espectro do sinal de áudio, indica o achatamento do espectro e mudanças súbitas que podem indicar transientes no áudio; (10) Spectral Centroid - primeiro momento central da distribuição do espectro, é o centro geométrico da distribuição; (11) Spectral Spread - segundo momento central da distribuição do espectro; e (12) Spectral Flux - medida das mudanças temporais no espectro entre janelas sucessivas. Uma explicação detalhada de todas essas 12 características pode ser encontrada no manual da MIRtoolbox [11] e em outros trabalhos [1], [12], [13].

A extração dessas características é realizada utilizando a MIRtoolbox (versão 1.6.1) [11]. Para realizar essa extração, o sinal de áudio é decomposto em janelas de 50 milissegundos (ms) de duração com 50% de sobreposição, gerando w janelas extraídas de cada áudio. Essas características acústicas são capazes de gerar respostas neurais significativas que podem ser registradas pelos sinais de EEG [4], [14], e podem ser úteis para predizer treinamento musical durante tarefas auditivas [5].

Tendo em vista o uso do modelo tradicional de análises de sinais cerebrais que se baseiam em estímulos artificiais, o uso da estatística multivariada, por meio da Análise Fatorial (FA), busca encontrar relações entre as características acústicas utilizadas, a fim de obter uma melhor representação desses dados. Assim, a princípio, questiona-se a necessidade de utilizar todas as características disponíveis para a análise, fazendo sentido manter apenas aquelas que são não-redundantes para um processamento mais eficiente computacionalmente dos dados.

Seja, portanto, a matriz $\mathbf{A}_g$ de cada gênero musical g da base de dados GTZAN representada como:

$$\mathbf{A}_g = \begin{bmatrix} \mathbf{a}_{1,1} & \mathbf{a}_{1,2} & \dots & \mathbf{a}_{1,n} \\ \mathbf{a}_{2,1} & \mathbf{a}_{2,2} & \dots & \mathbf{a}_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_{N,1} & \mathbf{a}_{N,2} & \dots & \mathbf{a}_{N,n} \end{bmatrix}, \tag{1}$$

em que N descreve o total de músicas analisadas e n é o total de características, tal que

$$\mathbf{a}_{i,j}^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_w \end{bmatrix}, \tag{2}$$

em que $i = [1, 2, ..., N]$ e $j = [1, 2, ..., n]$ para o número total w de janelas extraídas por música e por característica (aqui $N = 100$, $n = 12$ e $w = 119900$). Desta forma, $\mathbf{A}_g$ é composta pela concatenação das w janelas de todas as n características acústicas.

A proposta da FA aqui é então descrever a associação entre os valores extraídos de cada janela das características acústicas de maneira não-supervisionada. Dado que a matriz $\mathbf{A}_g$ possui uma matriz de correlação R com respectivamente P e Λ matrizes de autovetores e autovalores [15], tal que

$$R = P^T \Lambda P. \tag{3}$$

Para estimar os parâmetros deste modelo, é utilizado o método das componentes principais [15], pois a sua solução de decomposição espectral simplifica a questão de quantos fatores devem ser retidos na FA [5], convencionalmente escolhendo os fatores cujos autovalores sejam maiores do que 1 para determinar o número adequado de fatores utilizados na análise [15].

Idealmente, é desejado um padrão de resposta em que cada fator apresente alta carga fatorial para um determinado agrupamento (*cluster*) de características acústicas e nos demais fatores as características desse *cluster* apresentem baixa carga fatorial. Portanto, os $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_s]$ fatores podem substituir as n variáveis iniciais, com s fatores comuns rotacionados. O propósito da rotação é rearranjar os fatores de tal forma que tenham uma relação mais simples e interpretável com os padrões implícitos nos dados [16]. O tipo de rotação aplicada pode permitir que os fatores se correlacionem ou não. Assim, esses se diferenciam entre rotações ortogonais ou oblíquas [15]. Neste caso é aplicado a rotação *Varimax*, que é a mais comum utilizada, sendo uma rotação ortogonal. Ou seja, $\mathbf{F}$ são os autovetores identificados pelo método das componentes principais, rotacionados pelo método *Varimax*. Seu uso neste trabalho é explicado por forçar os fatores a não se correlacionarem, simplificando a análise. Desta forma, as associações entre as características acústicas serão mais significativas em termos da variância dos dados [15], [16].

Dessa forma, a FA é aplicada na matriz $\mathbf{A}_g$, resultando na matriz $\mathbf{F}_g$ que contém as cargas fatoriais das n características acústicas extraídas dos áudios, para os s fatores comuns rotacionados. Isso é feito primeiramente para cada um dos g gêneros musicais, e em seguida para todos os áudios sem levar em consideração os gêneros musicais. Ou seja, serão analisados os dados contidos no espaço das variáveis para se avaliar os comportamentos de agrupamentos existentes entre as características acústicas.

Supõe-se que haja diferenças entre gêneros nos *clusters*, que são resultantes da dispersão das cargas fatoriais de cada descritor sonoro. Espera-se que para os diferentes gêneros musicais hajam diferentes *clusters* entre as características acústicas analisadas. Para verificar isso, é analisado se os *clusters* gerados pela FA de cada gênero musical são similares,

ou seja, determina-se o grau de associação entre esses *clusters*. Isso é feito aplicando a correlação de Pearson entre os s fatores comuns rotacionados para cada gênero musical da matriz $\mathbf{F}_g$.

Por fim, é necessário avaliar os *clusters* encontrados. É necessário identificar os *clusters* existentes e avaliar a consistência destes *clusters* para, então, determinar o número adequado de fatores que devem ser extraídos na FA. O algoritmo k-means é empregado para particionar os dados em k conjuntos. Portanto, sejam $\mathbf{F}_g = [\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_s]$ os dados no espaço s-dimensional das variáveis e seja $C_i = C_1, C_2, \ldots, C_k$ os *clusters* em k *clusters*. Duas métricas serão calculadas aqui para determinar o valor ótimo de k e avaliar o número adequado de s fatores extraídos na FA, sendo essas Silhouette e Davies-Boulding índices [17], [18]. Ambos os critérios retornam o número ótimo de k clusters identificados.

1) Silhouette (Si): O critério Silhouette (Si) [17] calcula a medida da similaridade entre as observações dentro de seu próprio *cluster* comparada com os outros *clusters* existentes. Assumimos que os dados já foram agrupados em k *clusters* pela técnica k-means. Assim, o critério Silhouette é definido como

$$S_i = \frac{(b_i - o_i)}{max(b_i, o_i)}, \tag{4}$$

em que o_i é a distância média do ponto i aos outros pontos dentro de C_o, e b_i é a distância média mínima entre o ponto i e os pontos nos diferentes *clusters* C_k para $k \neq o$, minimizado entre *clusters*.

Esta métrica quantifica quão bem o objeto i foi classificado. A solução ótima de *cluster* apresenta o maior valor do índice Silhouette.

2) Davies-Bouldin (DB): O critério Davies-Bouldin (DB) [18] é uma medida para avaliar algoritmos de *clusters* baseada na razão das distâncias intra-cluster e inter-cluster. Novamente, assumimos que os dados já foram agrupados em k *clusters* pela técnica k-means. Assim, o critério DB é definido como

$$DB = \frac{1}{k} \sum_{i=1}^{k} D_i, \tag{5}$$

em que D_i é equivalente ao máximo de $D_{i,j}$ para $i \neq j$, sendo

$$D_{i,j} = \frac{(\bar{d}_i + \bar{d}_j)}{d_{i,j}}, \tag{6}$$

tal que $\bar{d}_i$ é a distância intra-cluster, i.e. a distância média entre cada ponto no *cluster* i ao seu centroide, $\bar{d}_j$ a distância média entre cada ponto no *cluster* j ao seu centroide, e $d_{i,j}$ a distância Euclidiana entre os centroides dos *clusters* i e j.

A solução ótima de *cluster* apresenta o menor valor do índice Davies-Bouldin.

III. RESULTADOS

Os áudios foram analisados para cada gênero musical g. Para a maioria dos gêneros musicais, apenas 2 fatores apresentavam autovalor acima de 1, exceto os gêneros Metal e Pop, que apresentaram 3 fatores com autovalores acima de 1. Assim, inicialmente, foi definida a utilização de apenas 2 fatores para se analisar como as características acústicas extraídas dos áudios se comportavam entre os diferentes gêneros musicais. A Figura 1 apresenta os gráficos de dispersão dos 10 gêneros musicais existentes na base de áudio GTZAN com as cargas fatoriais das características acústicas extraídas nos 2 fatores gerados pela FA. É possível observar que mesmo para os diferentes gêneros musicais, as características acústicas se agrupam similarmente. Sendo assim, identificamos, para cada gênero musical, os seguintes *clusters* entre as características acústicas: Cluster 1 - 1 (RMS), 4 (Spectral Roughness) e 12 (Spectral Flux); Cluster 2 - 8 (Spectral Skewness) e 9 (Spectral Kurtosis); Cluster 3 - 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread).

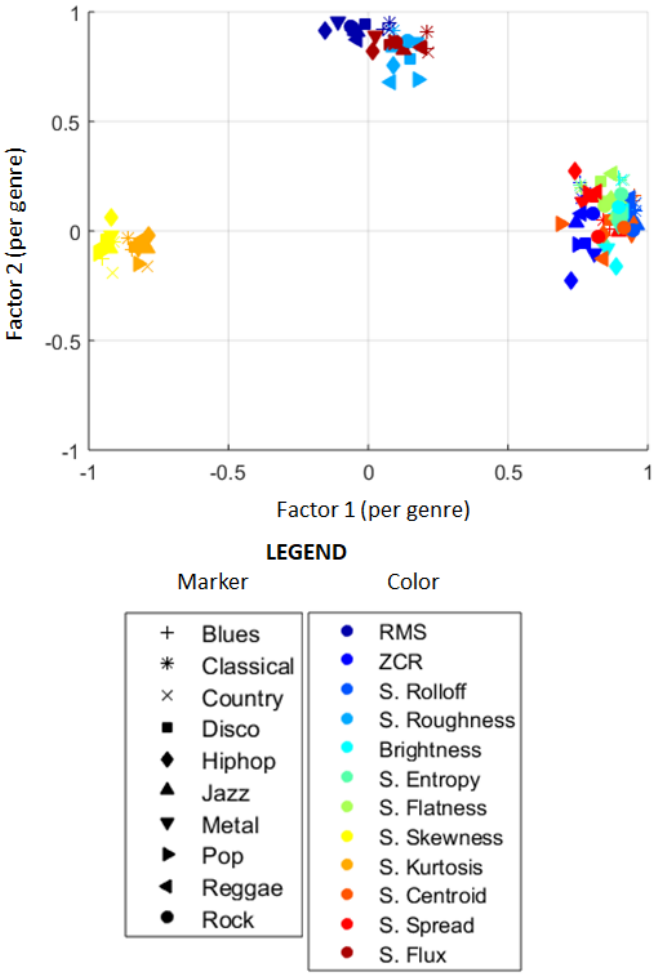

Fig. 1. Carga fatorial das características acústicas extraídas dos 10 diferentes gêneros musicais contidos na base de áudio GTZAN, sendo esses: Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock.

Para avaliar o número adequado de fatores a serem empregados na FA e a fim de determinar o número ótimo de *clusters*, os índices Silhouette (Figura 2) e Davies-Bouldin (Figura 4) foram utilizados, variando o número de k, considerando $k \in \{2, 3, \ldots, 6\}$. Dessa forma, nas Figuras 2 e 4 são apresentados os valores dos índices resultantes dos valores ótimos de k para cada um dos 10 gêneros musicais. Estes resultados mostram que para ambas as métricas, os melhores valores foram apresentados quando utilizado 2 fatores ($s = 2$).

Nessa condição, para todos os gêneros, $k = 3$ apresenta o número ótimo de *clusters*. Para outras condições ($s \neq 2$), o número de k variava significativamente entre gêneros. É possível ver a variação do número de k *clusters* nas Figuras 3 e 5.

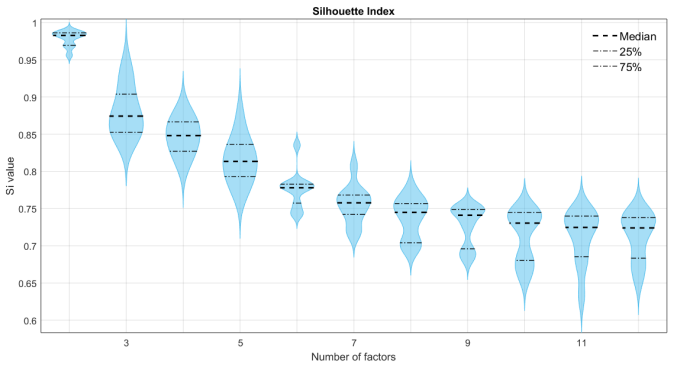

Fig. 2. Distribuição do índice Silhouette variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.

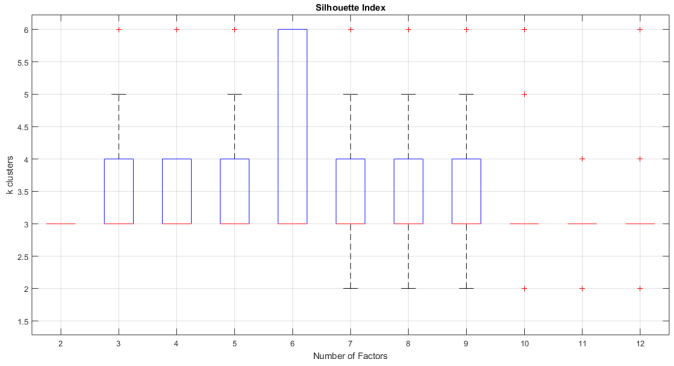

Fig. 3. Número de $k \in \{2, 3, \ldots, 6\}$ identificados utilizando o índice Silhouette, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.

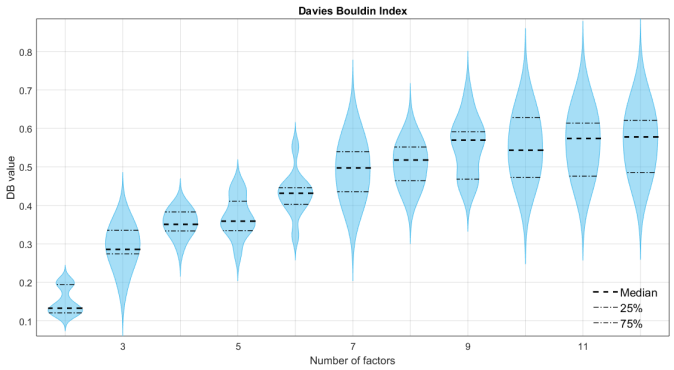

Fig. 4. Distribuição do índice Davis-Bouldin variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.

Para determinar o grau de associação entre as características acústicas dos gêneros musicais contidos na base de áudio GTZAN, é apresentada na Figura 6 a matriz de correlação do Fator 1 da FA, e na Figura 7 a matriz de correlação para o

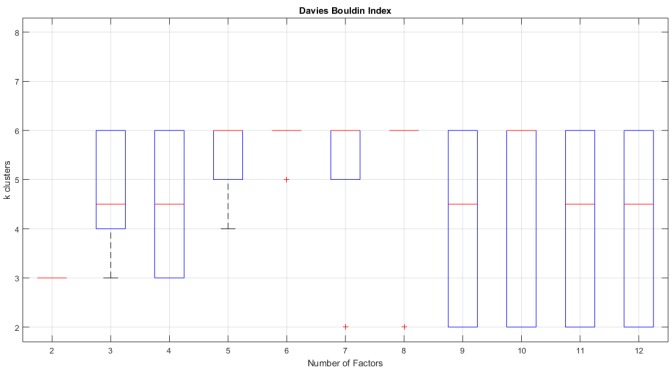

Fig. 5. Número de $k \in \{2, 3, \ldots, 6\}$ identificados utilizando o índice Davis-Bouldin, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.

Fator 2 da FA. É possível ver que, utilizando $s = 2$ como o número de fatores determinado pelos índices Silhouette e Davis-Bouldin, as características acústicas são agrupadas de maneira similar, apresentando alta correlação entre os diferentes gêneros musicais para os dois fatores.

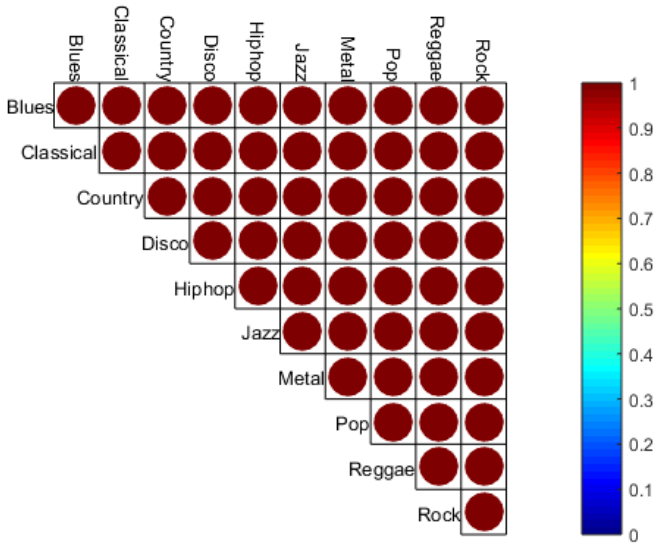

Fig. 6. Matriz de correlação do Fator 1 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA.

Por fim, é realizada a FA para todas as músicas sem considerar os gêneros musicais. A Figura 8 apresenta os protótipos das cargas fatoriais das características acústicas considerando todos os áudios da base de dados GTZAN sem diferenciação entre os gêneros. Assim, usando 2 fatores, é possível observar os mesmos *clusters* encontrados anteriormente, sendo: Cluster 1 - 1 (RMS), 4 (Spectral Roughness) and 12 (Spectral Flux); Cluster 2 - 8 (Spectral Skewness) and 9 (Spectral Kurtosis); Cluster 3 - 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread).

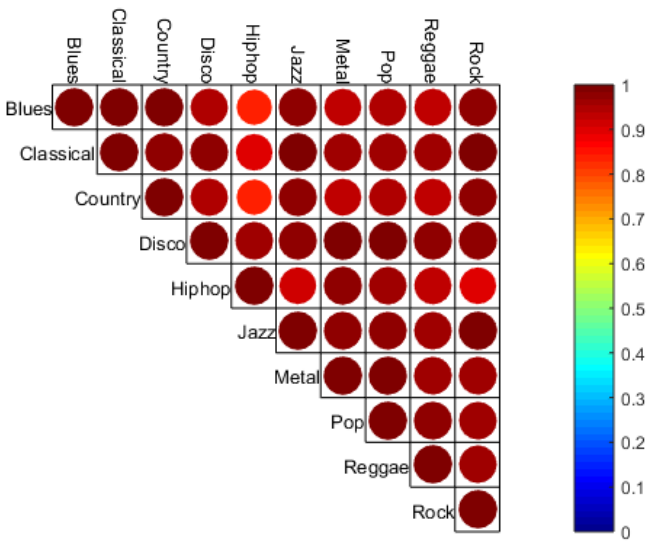

Fig. 7. Matriz de correlação do Fator 2 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA.

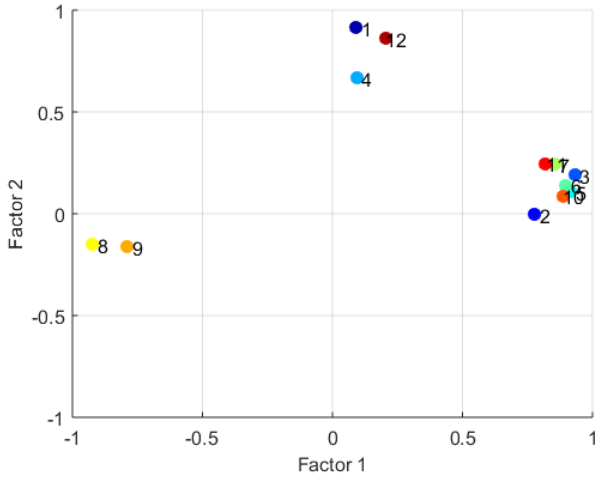

Fig. 8. Protótipo das cargas fatoriais das características acústicas extraídas de todos os áudios contidos na base de áudio GTZAN.

IV. DISCUSSÃO

Neste trabalho, foi aplicada a análise estatística multivariada para entender como se relacionam o conjunto de descritores sonoros comumente explorados na literatura afim. Nossos principais resultados mostram que independente do gênero musical, o mesmo comportamento de agrupamento acontece para as características acústicas investigadas. Isso pode ser observado nas Figuras 6 e 7, em que todas as características acústicas apresentam altos valores de correlação entre diferentes gêneros, significando que os mesmos *clusters* são encontrados.

O objetivo de utilizar a FA é revelar a associação entre um conjunto de características, selecionando as mais relevantes. Isso é importante porque a seleção de características afeta fortemente análises posteriores como a classificação ou a identificação de instantes significativos nos áudios capazes de gerar

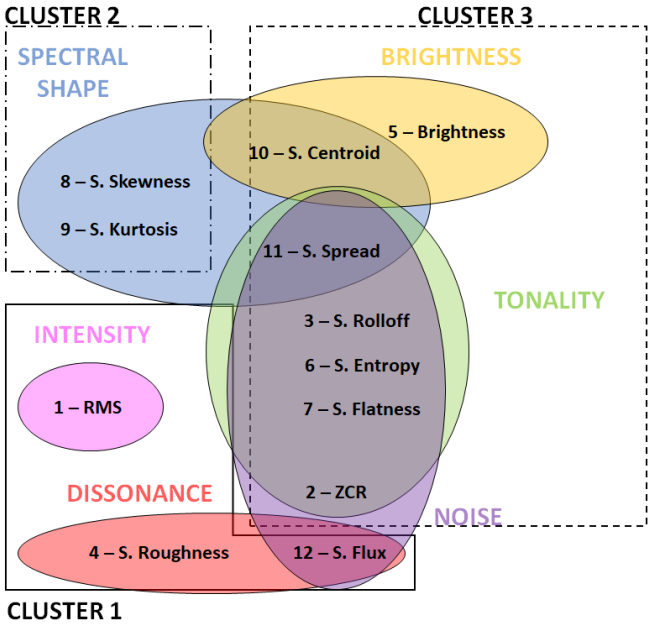

Fig. 9. Descrição esquemática da relação semântica entre os 12 descritores sonoros de baixo nível (representados pelas cores distintas das elipses) e os 3 clusters encontrados matematicamente pela FA.

respostas neurais relevantes durante tarefas auditivas. Isso é um problema de reconhecimento de padrões que estima as funções de densidade em espaços de alta dimensão, dividindo-os em regiões, determinando um comportamento de *cluster* entre os conjuntos de características extraídas dos sinais de áudio [15].

A princípio foi levantada a hipótese de que os gêneros musicais apresentariam diferentes comportamentos de *clusters* para o conjunto de características acústicas extraídas dos áudios. Nossa análise mostra que o mesmo *cluster* acontece, independente do gênero. Dada a natureza de cada característica acústica, é razoável que algumas dessas se agrupem, como no *Cluster* 3, em que todas as características são relacionadas com propriedades da tonalidade e ruído, ou o menor *cluster*, em que Spectral Skewness e Spectral Kurtosis estão relacionados com a forma espectral da distribuição dos dados. As características Spectral Flux e Spectral Roughness estão associadas, como ilustrado esquematicamente na Figura 9, com a sensação de dissonância no áudio e estão agrupadas também. Isso acontece especificamente quando são utilizados 2 fatores na FA. As Figuras 2 e 4 claramente demonstram que $s = 2$ apresenta a melhor relação de *cluster*, confirmando que $k = 3$ é o número ótimo de *clusters* para o arranjo dessas características acústicas. Aparentemente o primeiro e segundo fatores são melhores para explicar a similaridade entre as características acústicas extraídas para todos os gêneros musicais. No entanto, existem diferenças intra-*clusters* entre os gêneros musicais, embora o padrão inter-*cluster* permaneça igual.

Uma abordagem semelhante à apresentada neste trabalho é a de [2]. Em seu experimento, utilizando fMRI, foi realizada a Análise Fatorial em um conjunto de 25 características acústicas, compreendendo características de baixo e alto nível, utilizando a música clássica *Adios Nonino* de Astor Piazzolla. No entanto, para se adequar a frequência de amostragem dos

sinais de fMRI, foi realizada a redução da taxa de amostragem (*down-sampled*) das séries temporais que resultam das características acústicas de tal forma que o estímulo usado (que possuía 7 minutos e 42 segundos de duração) resultou em um vetor de 231 amostras. Esta redução de amostragem impacta diretamente na informação extraída das características, contrastando com a resolução temporal que encontramos no EEG. Mesmo com esta limitação, em [2] foi aplicada a Análise de Componentes Principais (PCA) com rotação *varimax*, retendo 9 componentes principais (PC), encontrando *clusters* entre as características acústicas extraídas. As características de alto nível não formaram *clusters*, mas cada uma correspondeu a um PC específico. Já as características de baixo nível formaram *clusters* semelhantes aos encontrados neste trabalho, sendo esses: (PC2) S. Centroid, S. Rolloff e ZCR (correspondendo ao aspecto do brilho do sinal); (PC3) S. Spread e S. Flatness (correspondendo ao aspecto da complexidade do timbre); (PC9) Roughness e S. Flux (correspondendo ao aspecto "*Activity*"); e RMS separado das outras características (correspondendo ao aspecto da intensidade). Vemos, assim, uma conformidade entre os resultados obtidos por [2] e os encontrados neste trabalho, demonstrando matemática e experimentalmente que para essa base de dados as características acústicas de baixo nível são, de fato, redundantes, havendo um comportamento de agrupamento entre elas para os gêneros distintos de músicas.

Embora não seja o escopo deste trabalho, a forma como foi construída a metodologia é capaz de identificar as falhas existentes na base utilizada, como repetições, pois representam o mesmo ponto no espaço de alta dimensão considerado, e identificações incorretas, destacando *outliers* dentro dos gêneros.

Com a metodologia proposta aqui, diferentemente da abordagem proposta inicialmente para análise do processamento musical [4], em que são exploradas as diferentes características acústicas para encontrar os instantes que devem ser observados nos sinais de EEG durante tarefas auditivas, é possível determinar quais características acústicas melhor representam os áudios (correspondentes às mais significativas dentro dos *clusters* encontrados aqui), e utilizá-las nas análises dos sinais de EEG, garantindo que não existam redundâncias nas informações acústicas observadas nos instantes utilizados como estímulos para as análises neurais.

Finalmente, essa abordagem, proposta para seleção de características, pode ser utilizada em problemas diversos, em que o número de características é maior. Assim, neste artigo foi demonstrado que é possível selecionar características acústicas que melhor representem as músicas apresentadas como estímulo em tarefas auditivas, a fim de investigar como o cérebro processa a música, dada a redundância estatística encontrada na FA, independentemente da própria música e do seu gênero.

V. Conclusão

Neste trabalho, demonstramos que é possível selecionar características acústicas mais adequadas para representar estímulos musicais naturalistas, a fim de investigar como o cérebro processa a música, dada a redundância estatística encontrada pela FA. Esta é uma avaliação válida de FA para seleção de características nesta tarefa específica e para este conjunto particular de descritores. Aqui, demonstramos experimentalmente que há uma redundância entre essas características acústicas de baixo nível, que não estão apenas relacionados a uma música ou gênero em particular. Assim, podemos tirar proveito desse comportamento de *cluster* para análises posteriores.

Agradecimentos

Essa pesquisa foi apoiada pela bolsa do Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 e o INCT MACC, processo 465586/2014-7.

Referências

[1] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder, "Features for content-based audio retrieval," in *Advances in Computers Volume 78 Improving the Web*, pp. 71–150, Elsevier, 2010.

[2] V. Alluri, P. Toiviainen, I. P. Jaaskelainen, E. Glerean, M. Sams, and E. Brattico, "Large-scale brain networks emerge from dynamics processing of musical timbre, key and rhythm," *NeuroImage*, vol. 59, no. 4, pp. 3677–3689, 2012.

[3] V. Alluri, P. Toiviainen, T. E. Lund, M. Wallentin, P. Vuust, K. Nandi, Asoke, T. Ristaniemi, and E. Brattico, "From vivaldi to beatles and back: Predicting lateralized brain responses to music," *NeuroImage*, vol. 83, pp. 627–636, Dec. 2013.

[4] H. Poikonen, V. Alluri, E. Brattico, O. Lartillot, M. Tervaniemi, and M. Huotilainen, "Event-related brain responses while listening to entire pieces of music," *Neuroscience*, vol. 312, pp. 58–73, Jan. 2016.

[5] E. Ribeiro and C. E. Thomaz, "A whole brain eeg analysis of musicianship," *Music Perception: An Interdisciplinary Journal*, vol. 37, no. 1, pp. 42–56, 2019.

[6] P. Saari, I. Burunat, E. Brattico, and P. Toiviainen, "Decoding musical training from dynamic processing of musical features in the brain," *Scientific Report*, vol. 708, no. 8, pp. 1–12, 2018.

[7] N. T. Haumann, M. Lumaca, M. Kliuchko, J. L. Santacruz, P. Vuust, and E. Brattico, "Extracting human cortical responses to sound onsets and acoustic feature changes in real music, and their relation to event rate," *Brain Research*, p. 147248, 2021.

[8] L. A. Ferreira, E. Ribeiro, and C. E. Thomaz, "A multivariate statistical analysis of eeg signals for differentiation of musicians and non-musicians," in *17th Brazilian Symposium on Computer Music (SBCM)* (J. T. Araújo and F. L. Schiavoni, eds.), (Sao Paulo, Brazil), pp. 80–85, Brazilian Computer Society (SBC), 10 2019.

[9] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 293–302, 7 2002.

[10] B. L. Sturm, "The state of the art ten years after a state of the art: Future research in music information retrieval," *Journal of New Music Research*, vol. 43, p. 147–172, 4 2014.

[11] O. Lartillot, *MIRtoolbox 1.6.1 Users Manual*, 2014.

[12] A. Lerch, *An Introduction to audio content analysis. Applications in signal processing and music informatics*. IEEE Press, 1st ed., 2012.

[13] P. Kness and M. Schedl, *Music Similarity and Retrieval: An introduction to audio and web based strategies*. Springer, 1st ed., 2016.

[14] H. Poikonen, P. Toiviainen, and M. Tervaniemi, "Early auditory processing in musicians and dancers during a contemporary dance piece," *Scientific Reports*, vol. 6, Sept. 2016.

[15] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Pearson, 6th ed., 2007.

[16] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, pp. 187–200, 1958.

[17] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53 – 65, 1987.

[18] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

Estela Ribeiro received in 2015 the B. Sc. degree in mechanical engineering from FSA University Center, São Paulo, Brazil. Obtained the M.Sc. degree in electrical engineering from FEI University Center in 2017. Obtained the ph.D. degree in electrical engineering at FEI University Center, São Paulo, Brazil, in 2020. Since 2016, she has received a research fellowship from FEI, LNCT and CAPES to develop research activities on signal processing and pattern recognition. Her research interests include pattern recognition, cognitive perception and machine learning.

Carlos Thomaz received in 1993 the B.Sc. degree in electronic engineering from Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil. After working for six years in industry, he obtained the M.Sc. degree in electrical engineering from PUC-Rio in 1999. In October 2000, he joined the Department of Computing at Imperial College London where he obtained the Ph.D. degree in statistical pattern recognition in 2004. He joined the Department of Electrical Engineering, FEI University Center, São Paulo, Brazil, in 2005, as an Associate Professor, where he has been, since 2006, head of the Image Processing Laboratory. Since 2014 he has been Professor of Statistical Pattern Recognition at FEI. His research interests include pattern recognition, cognitive perception and machine learning. From 2015 to 2018, Professor Thomaz was awarded a Newton Advanced Fellowship from the Royal Society, UK.