

A cluster analysis of benchmark acoustic features on Brazilian music

Leonardo Antunes Ferreira, Estela Ribeiro, Carlos Eduardo Thomaz

¹Centro Universitário FEI

Av. Humberto de Alencar, 3972-B – 09850-901, São Bernardo do Campo, SP

leonardoantunes123@gmail.com, estela.eng@hotmail.com, cet@fei.edu.br

Abstract. *In this work, we extend a standard and successful acoustic feature extraction approach based on trigger selection to examples of Brazilian Bossa-Nova and Heitor Villa Lobos music pieces. Additionally, we propose and implement a computational framework to disclose whether all the acoustic features extracted are statistically relevant, that is, non-redundant. Our experimental results show that not all these well-known features might be necessary for trigger selection, given the multivariate statistical redundancy found, which associated all these acoustic features into 3 clusters with different factor loadings and, consequently, representatives.*

1. Introduction

In the last decade, we have witnessed a significant increase in the number of scientific studies that have identified cognitive differences between musicians and non-musicians [1, 2, 3, 4, 5]. Most of these studies relied on artificial stimuli [4, 5], but recent works on this issue are using naturalistic music pieces, with their corresponding original stimuli, to successfully indicate the cognitive patterns related to human music processing [1, 2, 3].

All these recent works are based on acoustic features extraction from the audio signals to identify the so-called triggers [1], which are instants in the music time series with a rapid increase in the value of a acoustic feature that can be related with cognitive responses. To the best of our knowledge, none of these works has explored such trigger selection approach on Brazilian music.

Therefore, we evaluate whether examples of Brazilian Bossa-Nova and classical music pieces can achieve similar results reported in the aforementioned literature. Moreover, we propose and implement a computational framework based on multivariate statistical analysis to disclose whether all these benchmark acoustic features commonly used in music information retrieval (MIR) tasks are statistically relevant (that is, non-redundant) intending to reduce the number of features required to set triggers for music perception.

2. Methodology

Our methodology can be divided into 3 steps: (I) Music selection; (II) Feature extraction; (III) Feature selection. In the first step we have selected 4 music pieces for the analysis: Garota de Ipanema with two versions, one performed by Tom Jobim (GI1) and the other an instrumental version played by Zimbo Trio (GI2), Bachianas Brasileiras N°2 - O Trenzinho do Caipira composed by Heitor Villa Lobos (BB), and Hungarian Dance N°5 from Johannes Brahms

(HD) as a literature and comparative reference. All songs are available on Spotify or IMSLP music libraries and have been gathered with a sampling rate of 44.1 kHz and saved in WAV format.

In the second step, the following 12 benchmark acoustic features utilized to describe audio signals are calculated using the well-known MIRtoolbox (Version 1.71) [1, 6, 7] and Matlab 2015a, decomposing the audio into a 50 milliseconds window overlapped by 50%: (1) Root Mean Square Energy (RMS), (2) Zero Crossing Rate (ZCR), (3) Spectral Rolloff, (4) Spectral Roughness, (5) Brightness, (6) Spectral Entropy, (7) Spectral Flatness, (8) Spectral Skewness, (9) Spectral Kurtosis, (10) Spectral Centroid, (11) Spectral Spread and (12) Spectral Flux. A detailed description of these features can be found in the user manual of the MIRtoolbox [7].

In the third step, we estimate whether all these 12 features, concatenated directly from the previous analysis windows, are statistically relevant through a cluster analysis using Factor Analysis (FA) with *varimax* rotation [8]. The number of factors retained in FA is equal to the number of principal components extracted from data that have eigenvalue greater than 1. We have selected the representative acoustic feature within each cluster as the one with the highest loading on each factor. The K-means algorithm has been applied to identify clusters that describe statistically similar acoustic features.

3. Results

Our results show that for all songs using only three principal components is possible to explain more than 80% of the data variance. Thus, the number of factors used in FA is equal to three. The factor loadings obtained by FA disclose the correlation among the acoustic features and the clusters presented in Figure 1 were obtained with the K-means algorithm.

Clearly, from Figure 1, we can see the same inter-cluster grouping of the acoustic features for all music pieces. Thus, regardless of the musical genre or whether it is an instrumental or vocal music piece, there are the same acoustic feature memberships between the same clusters. Some of these memberships might be expected. For instance, the S. Skewness (acoustic feature 8) and S. Kurtosis (9) are in the same cluster 1 for all four songs, since both provide information about the type and magnitude of audio signals departures from normality. However, it is interesting to notice that the importance of each feature within each cluster is not equal. In other words, for each song (GI1, GI2, BB and HD), there are 3 non-redundant clus-

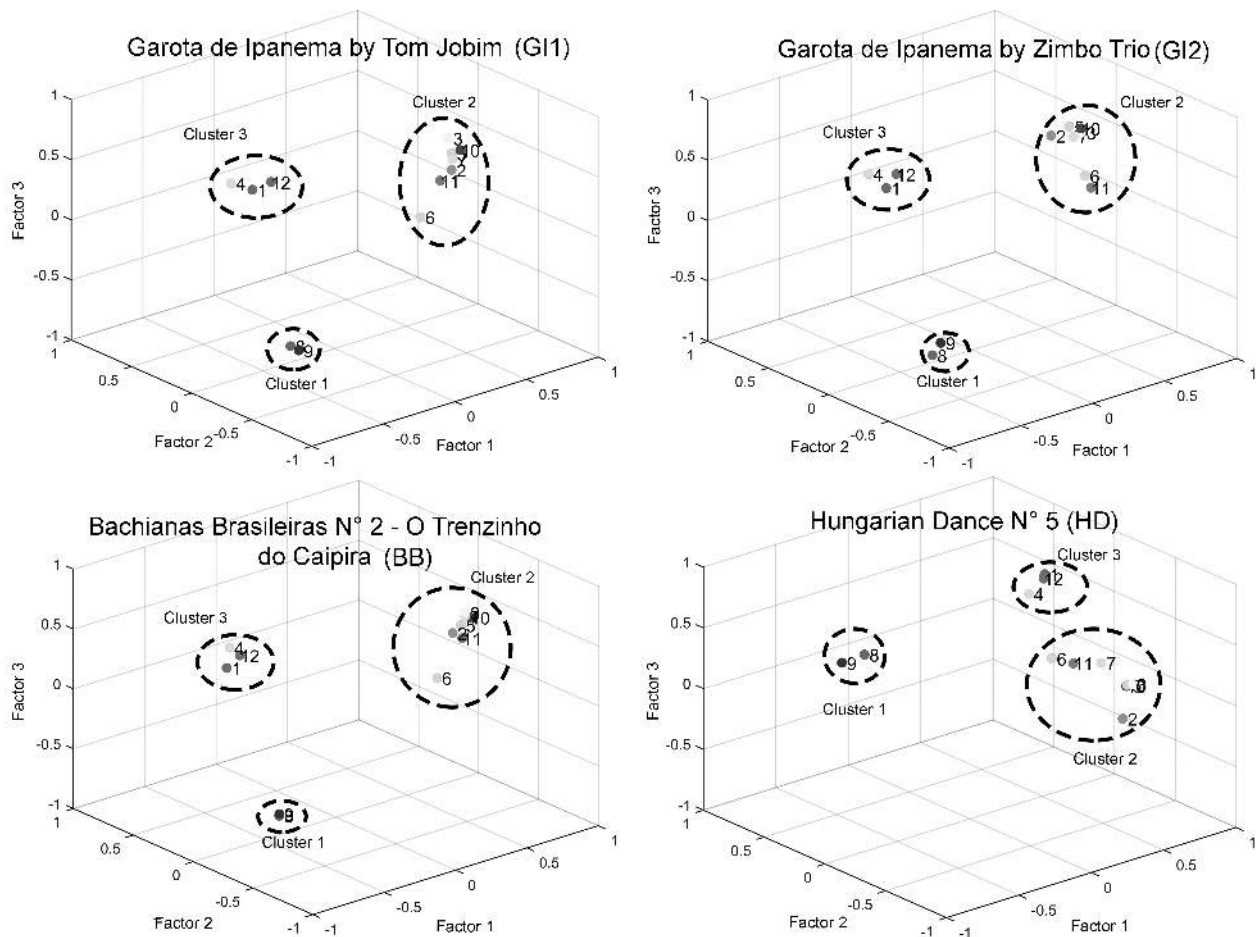


Figure 1: Factor loadings of the acoustic features extracted from the audio signals, being possible to observe the formation of clusters between the following features: Cluster 1 - 8 (S. Skewness) and 9 (S. Kurtosis); Cluster 2 - 2 (ZCR), 3 (S. Rolloff), 5 (Brightness), 6 (S. Entropy), 7 (S. Flatness), 10 (S. Centroid) and 11 (S. Spread); Cluster 3 - 1 (RMS), 4 (S. Roughness) and 12 (S. Flux).

ters, but with distinct acoustic feature representatives, as follows [cluster1, cluster2, cluster3]: GI1 [9, 10, 1]; GI2 [8, 5, 1]; BB [8, 10, 1]; HD [9, 3, 1].

4. Conclusion

In this work, we showed that it is possible to reduce the number of acoustic features required to investigate the cognitive patterns evoked during music listening, given the statistical redundancy found by FA, which grouped all the 12 benchmark acoustic features into only 3 clusters.

This is an exploratory study that indicates a similar clustering behavior between a couple of Brazilian Bossa-Nova and classical music pieces when performing benchmark acoustic features extraction. The same inter-cluster behavior might be achieved in other musical genres, requiring further investigation. However, the distinct intra-cluster acoustic feature loadings highlight the importance of selecting the most relevant features within each cluster to properly representing statistically each music piece for perception analysis.

Acknowledgments

This study was financially supported in part by FEI (PBIC 074/18), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, and the INCT MACC, process 465586/2014-7. We would like to thank all the volunteers who took part in the experiment.

References

- [1] Hanna Poikonen, Vinoo Alluri, Elvira Brattico, Olivier Lartillot, Mari Tervaniemi, and Minna Huotilainen. Event-related brain responses while listening to entire pieces of music. *Neuroscience*, 312:58–73, 2016.
- [2] Andjela Markovic, Jürg Kühnis, and Lutz Jäncke. Task context influences brain activation during music listening. *Frontiers in human neuroscience*, 11:342, 2017.
- [3] Pasi Saari, Iballa Burunat, Elvira Brattico, and Petri Toivainen. Decoding musical training from dynamic processing of musical features in the brain. *Scientific Report*, 708(8):1–12, 2018.
- [4] Peter Vuust, Elvira Brattico, Miia Sppanen, Risto Naatanen, and Mari Tervaniemi. The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch neg-

ativity paradigm. *Neuropsychologia*, 50(7):1432–1443, June 2012.

- [5] C.A. Mikutta, G. Maissen, A. Altorfer, W. Strik, and T. Koenig. Professional musicians listen differently to music. *Neuroscience*, 268:102 – 111, 2014.
- [6] Vinoo Alluri, Petri Toiviainen, Iiro P Jääskeläinen, Enrico Glerean, Mikko Sams, and Elvira Brattico. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage*, 59(4):3677–3689, 2012.
- [7] Oliver Lartillot. *MIRtoolbox 1.6.1 Users Manual*, 2014.
- [8] Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson, 6th edition, 2007.