EVOLUTION OF TIMBRE DIVERSITY IN A DATASET OF BRAZILIAN POPULAR MUSIC: 1950-2000

Rodrigo Borges
University of São Paulo
rcborges@ime.usp.br

Marcelo Queiroz
University of São Paulo
mqz@ime.usp.br

ABSTRACT

In this paper we discuss a method for assessing the temporal evolution of timbre diversity in an annotated dataset, and apply it to a collection of Brazilian music from the 1950's to the 2000's. Previous work have explored audio analysis for measuring the variety of acoustic features or the stylistic evolution in American Popular Music in the period 1950-2010. We aim in this study to verify up to what point a similar methodology could be applied to a considerably different dataset (Brazilian popular music) in a comparably long period (1950-2000). The measure of timbre diversity, based on Shannon's entropy function, displays its lowest value for 1950-1955, abrupt decaying from 1975 to 1990 and an increasing trend from this point until 2000.

1. INTRODUCTION

Empirical musicologists relate the recent availability of large collections of digital music with a more scientific approach to music history [1, 2]. Digital audio allows acoustic descriptors to be extracted automatically and to be applied in many tasks in the field of Music Information Retrieval (MIR) [3]. We propose using acoustic descriptors to analyze the evolution of timbre diversity in a dataset of Brazilian popular music.

Among all music dimensions, timbre is the one which resists most formalization attempts, being frequently defined by opposition, as the sound quality which is not pitch, not intensity and not duration; it is informally referred to as the sound color, or more objectively as associated to the spectral composition and its dynamic variations [4]. For practical purposes and Music Information Retrieval tasks, the acoustic descriptor named Mel Frequency Cepstral Coefficients (MFCCs) is the most frequent timbre-related characteristic extracted from music sound signals, and will be addressed in the following sections.

We attempt to express timbre diversity as a measure of how these acoustic descriptors are distributed through all possible regions in an MFCC representation space, for each group of songs belonging to the same period. Periods with higher timbre diversity should display MFCC distributions that span many regions, whereas periods of low timbre diversity should display highly concentrated MFCC distributions.

We use Serrà et al [1] and Mauch et al. [2] as references for our work, and aimed to verify to what extent their methodology can be reproduced in a different dataset distributed in a period of comparable length, namely the dataset "100 greatest Brazilian music records" [5] comprising Brazilian popular music from the 1950's to the 2000's, compiled by the specialized music magazine Rolling Stone in 2007.

The text is structured as follows. In the next section we present the two studies taken here as a basis, focusing on how they deal with the timbre dimension. The methodology for measuring timbre diversity evolution is then detailed, comprising the dataset description as number of records, artists, years and songs; feature extraction; codeword representation and diversity measurement. Results are presented, discussed and compared to the ones in the literature. Conclusions and future work are presented in Section 5.

2. RELATED WORK

Serrà et al. [1] analyze harmony, timbre and loudness descriptors extracted from 464,411 distinct music recordings from a public collection known as the "million song dataset", using recordings from 1955 to 2010. They calculated *codeword* representations for harmony and timbre, and use a power law model for expressing the diversity of the distribution of these features over the years. The idea behind this method is to assume codewords as representatives of particular harmonic or timbre structures, and to associate higher degrees of diversity to samples with a more balanced distribution: e.g. if songs from a specific year use diversified harmonic and timbre combinations, this distribution should be more balanced, but if these songs use only relatively few of them, then the distribution will be more concentrated towards fewer codewords, i.e. it would be less balanced.

In order to take feature successions into account, the authors also proposed modeling each song using transition networks, where each node represents a codeword and each link represents a temporal transition. The measures of average shortest path length, clustering coefficient and assortativity with respect to a random network, were interpreted in terms of higher or lower diversity of harmonic and timbre elements. For the specific case of timbre, the diversity rate reached its peak in the year 1965, and started to decrease from there. Despite interesting evolutionary observations, such as the ”loudness race”, corresponding to a constant increase in the loudness level over the years, or the timbre diversity peak value happening in the year of 1965, the authors point out to a general lack of significant statistical trends in the evolution of harmonic or timbre elements in contemporary western popular music in the period considered.

Mauch et al. [2] investigated the “US Billboard Hot 100” between 1960 and 2010, aiming to measure musical diversity and evolution of disparities, as well as demonstrating quantitative trends of harmonic and timbre properties. As motivation cues, the authors asked three questions, to be answered during the analysis: (1) did North American popular music variety increase or decrease over time?; (2) were evolutionary changes continuous or discontinuous?; and (3) if they were discontinuous, when did discontinuities occur?

They chose to represent the acoustic properties in a fashion similar to the previous authors, but using the term *topics* instead of codewords. 16 topics were calculated, 8 based on MFCC (for timbre-related aspects) and 8 based on Chroma (for harmony-related aspects). Topics were calculated with a hierarchical generative model named Latent Dirichlet Allocation (LDA).

Having calculated timbre and harmonic topics for each song, it was possible to study the evolution of topics over the years. Authors also had access to expert-based annotations, that made possible the association of semantic information to each topic, for example ”drums, agressive, percussive” in the case of a particular timbre topic, and ”natural minor” for a harmonic topic. Temporal evolution in the frequency of some topics revealed clear trends, as for example the topic named ”energetic, speech and bright”, that starts increasing in occurrence from 1980 on.

Four measures of diversity are presented by [2]: the first measure is simply the number of songs in each time period, used to verify that other diversity measures are not affected by the size of a subsample. The second measure accounts for the year-wise diversity of acoustic style clusters in the data. The third is the effective number of music topics for each year, averaged across the harmonic and timbre topics. The fourth corresponds to disparity, or the variety of measurements in the matrix of principal components derived from the topics.

By using the Kmeans clustering algorithm, authors estimated that 13 clusters would better separate data in terms of the distribution of topics. These clusters are associated to musical styles and their evolution over the years is discussed. A Self Similarity Matrix was also calculated to assess topic distribution over the years, by computing the similarity between topic distribution of different time periods. This matrix was used to detect discontinuities, that according to these authors took place specially in 1983 and 1991.

3. METHODOLOGY

In order to verify to what extent the methodology applied in [2] can be reproduced in a different dataset we adapted it to our music collection as explained in the following subsections.

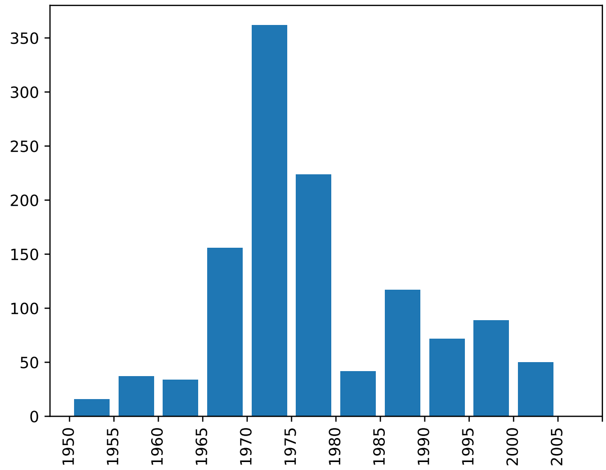

Figure 1. Distribution of songs over the 5 year periods .

3.1 Dataset

The dataset consists of 100 records released from 1950 to 2000, by 60 artists, summing up 1199 songs elected by the specialized music magazine Rolling Stone as the ”100 greatest Brazilian music records list” [5]. This collection was published as representative of the opinions of 60 music researchers, producers and journalists, based on how influential they thought these records were to others artists.

Since the number of songs in each 5-year period is very unbalanced (see Figure 1), we selected random subsamples based on the period with the least number of songs, in order to allow for a more stable comparison. A table with all artist names, number of records, number of songs and number of years spanned by each artist in the dataset is presented in Table 1.

3.2 Timbre Feature Extraction

Mel frequency cepstral coefficients (MFCCs) were originally developed for automatic speech recognition and were later found to be useful for music information retrieval [3]. Even though timbre as a concept is very hard to define, since it encompasses many acoustic dimensions, MFCCs captures relevant timbre-related acoustic characteristics of the signal spectrum, and were also used in our reference work [2].

MFCC data were extracted with the Librosa library [1] using 13 coefficients, windows of 2048 samples and 75% overlap between windows. That sums up to 10.844.508 frames extracted for the whole dataset.

3.3 Codeword Representation

Codeword Representation is a technique for representing high cardinality data, allowing data to be clustered in fewer groups of similar elements, and representing each sample as a histogram. The first idea behind this technique is to

[1] http://github.com/librosa/librosa

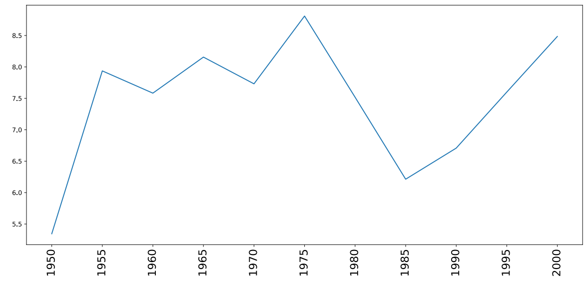

Figure 2. The entropy calculated in bits for each 5-year period.

apply unsupervised clustering techniques (e.g. Kmeans) to estimate how many clusters would ensure that the data can be well separated. In our case we used 10 clusters to represent all MFCC arrays extracted from all songs. At this point each MFCC is identified as belonging to one of the 10 clusters, and each song can be seen as a temporal succession of transitions betweens clusters. These transitions are counted for each song, ending up with a histogram indicating its distribution of MFCCs over the 10 clusters. The histograms are normalized with respect to time and become the Codeword Representation of each song.

3.4 Diversity metric

According to Mauch et al. [2], maximum diversity is achieved when frequencies are uniformly distributed in the histogram, and minimum diversity corresponds to all MFCCs belonging to a single cluster. As suggested by these authors, we take Shannon's entropy function as a measure of diversity. The average proportion of frames over each cluster $\bar{q}$ for a given 5-year period is given by

$$\bar{\mathbf{q}} = (\bar{q_1}, \bar{q_2}, \ldots, \bar{q_{10}}). \tag{1}$$

We calculate the diversity defined as

$$D = exp\left(-\sum_{i=1}^{10} \bar{q}_i \ln \bar{q}_i \right) \tag{2}$$

The maximum entropy value is attained when all $\bar{q}_i$ are equal and $D = 10$. The minimum value occurs when only one cluster is represented, and $D = 1$.

4. RESULTS

In Figure 3 it is possible to notice how the probabilities are distributed over the decades (horizontal axis) and over the clusters (vertical axis). There are periods when these probabilities are more uniformly distributed through all clusters (1970, 1975,1980), in contrast to periods when they are much more concentrated in fewer clusters (1950,1985,1990). It also possible to see clusters that present almost constant proportion over the decades, as the case of the cluster number 8, in opposition to cluster 7 that presents a peak in 1950 and then decays with time.

Another visualization in the right side of Figure 3, with the same values from the previous matrix but with columns and rows sorted by similarity. Two dendrograms are presented as representing the similarity of periods (columns) and clusters (rows), the tree-like grouping is built from euclidean distances between elements of the array. The closest periods, as indicated in the upper dendrogram, are 1970 and 2000, followed by the next most similar pair, 1965 and 1990. 1950 is considerably different when compared to all other periods.

The entropy is then calculated for each group of songs of the set and the results are shown in Figure 2. 1975 presented highest degree of diversity when compared to the other periods, 1950 presents the lowest one.

Mauch et al. [2] applies a similar measurement of diversity in a similar period of time, but mixing timbre and harmonic features, and finds the lowest value in 1985. After then it starts to increase and reach its highest value around 2000 for the case of American Popular Music, similarly to what was presented here.

5. CONCLUSIONS

We applied a part of the methodology from Mauch et al. [2] while trying to transpose the same analysis to a different music dataset. The results of the entropy-based diversity metric bring interesting trends for discussion, which might lead to interesting musicological interpretations and insights.

6. REFERENCES

[1] J. Serrà, A. Corral, M. Boguñá, M. Haro, and J. L. Arcos, "Measuring the evolution of contemporary western popular music," *Scientific Reports*, vol. 2, 07/2012 2012. [Online]. Available: http://www.nature.com/srep/2012/120726/srep00521/full/srep00521.html

[2] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, "The evolution of popular music: Usa 1960–2010," *Royal Society Open Science*, vol. 2, no. 5, 2015. [Online]. Available: http://rsos.royalsocietypublishing.org/content/2/5/150081

[3] E. Pampalk, "Computational models of music similarity and their application in music information retrieval," Ph.D. dissertation, Vienna University of Technology, Vienna, Austria, March 2006. [Online]. Available: http://www.ofai.at/~elias.pampalk/publications/pampalk06thesis.pdf

[4] J. M. Grey, *An exploration of musical timbre*. Dept. of Music, Stanford University, 1975, no. 2.

[5] R. S. Brazil, ""Os 100 maiores discos da Música Brasileira" (100 greatest Brazilian music records list)," http://rollingstone.uol.com.br/listas/os-100-maiores-discos-da-musica-brasileira/, 2018, [Online; accessed 27-April-2018].

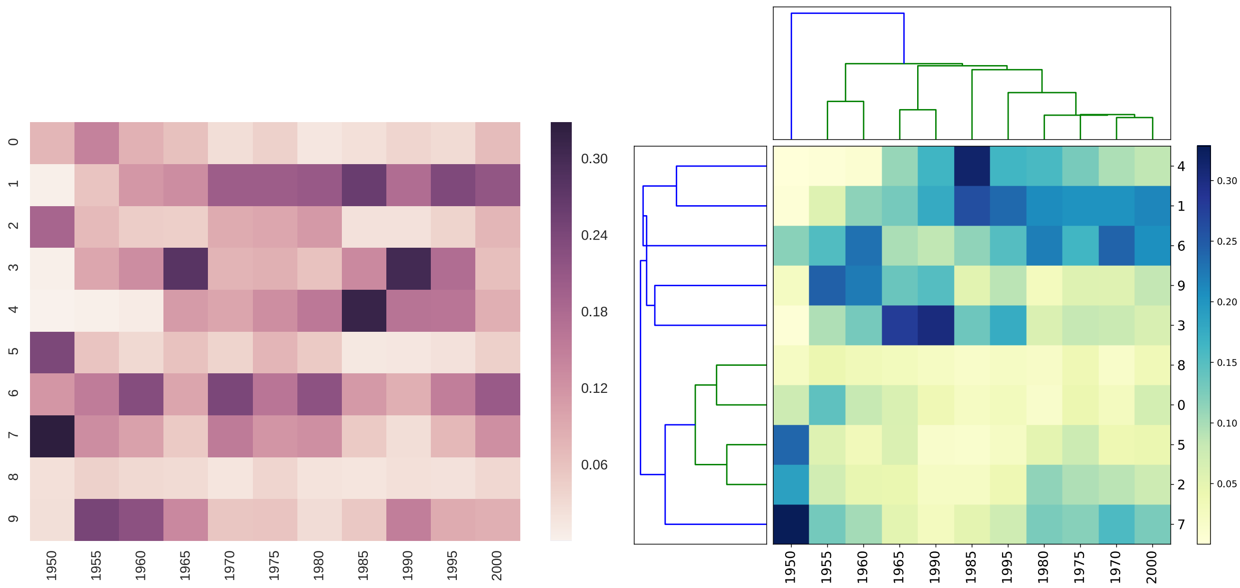

Figure 3. Left: Probabilities of each cluster summed for each 5-year period. Right: Matrix with the probabilities distributed over the clusters and over the 5-year periods, sorted by similarity between columns and rows. The two dendrograms (top and left) indicate the euclidean distances between periods and clusters, respectively.

ID	Artist	# records	# songs	ID	Artist	# records	# songs
0	Caetano Veloso	6	55	31	Júpiter Maçã	1	14
1	Gilberto Gil	5	48	32	Nélson Cavaquinho	1	13
2	Os Mutantes	5	44	33	Secos e Molhados	1	13
3	Roberto Carlos	4	48	34	Itamar Assumpção	1	13
4	Jorge Ben	4	45	35	Blitz	1	13
5	Gal Costa	4	38	36	Elizeth Cardoso	1	13
6	Tim Maia	3	42	37	Tom Zé	1	12
7	Racionais Mc's	3	32	38	Ângela Rorô	1	12
8	João Gilberto	3	30	39	O Rappa	1	11
9	Tom Jobim	3	30	40	RPM	1	11
10	Chico Science/Nação Zumbi	2	37	41	Erasmo Carlos	1	11
11	Sepultura	2	32	42	Os Paralamas Do Sucesso	1	11
12	Milton Nascimento	2	32	43	Ultraje a Rigor	1	11
13	Los Hermanos	2	29	44	Maria Bethânia	1	11
14	Raul Seixas	2	27	45	Luiz Melodia	1	10
15	Mundo Livre S/A	2	27	46	Moacir Santos	1	10
16	Titãs	2	26	47	Arnaldo Baptista	1	10
17	Marisa Monte	2	25	48	Banda Black Rio	1	10
18	Cartola	2	24	49	Novos Baianos	1	9
19	Legião Urbana	2	23	50	Rita Lee & Tutti Frutti	1	9
20	João Donato	2	22	51	Gilberto Gil; Jorge Ben	1	9
21	Paulinho da Viola	2	22	52	Arrigo Barnabé	1	8
22	Elis Regina	2	22	53	Aracy de Almeida	1	8
23	Ira!	2	20	54	João Gilberto; Stan Getz	1	8
24	Chico Buarque	2	20	55	B. Powell; V. de Moraes	1	8
25	Dorival Caymmi	2	14	56	Egberto Gismonti	1	8
26	N. Leão; Z. Kéti; J. do Vale	1	23	57	Caetano; Gal; Gil; Mutantes	1	2
27	Doces Bárbaros	1	17	58	Caetano Veloso;Gilberto Gil	1	1
28	Raimundos	1	16	59	Caetano Veloso; Gal Costa	1	1
29	Walter Franco	1	14	60	Nara Leão	1	1
30	Elis Regina; Tom Jobim	1	14				

Table 1. ID, artists, number of records and songs present in the database. We have gathered different names used by the same artists: e.g. Tom Jobim = Antônio Carlos Jobim; Os Mutante = Mutantes; Jorge Ben Jor = Jorge Ben