Computer Music Research Group - IME/USP Report for SBCM 2019

Fábio Gorodscy [1] , Guilherme Feulo [1] , Nicolas Figueiredo[1] ,
Paulo Vitor Itaboraí[1] , Roberto Bodo[1] , Rodrigo Borges[1] , Shayenne Moura[1]

[1]FLOSS Competence Center (CCSL), Room 118 – Department of Computer Science /
Institute of Mathematics and Statistics (IME) / University of São Paulo
Rua do Matão, 1010, CCSL – 05508-090 Sao Paulo, SP

compmus.ime@gmail.com

Abstract. *The following report presents some of the ongoing projects that are taking place in the group's laboratory. One of the noteable characteristics of this group is the extensive research spectrum, the plurality of research areas that are being studied by it's members, such as Music Information Retrieval, Signal Processing and New Interfaces for Musical Expression.*

1 Introduction

This report presents the Computer Music Research Group, part of the Department of Computer Science at the Institute of Mathematics and Statistics at the University of Sao Paulo (IME-USP). The group is coordinated by Prof. Dr. Marcelo Queiroz, and is composed of undergraduate, masters and PhD candidates. Its research covers many diverse topics on MIR (singing voice detection, query-by-humming, audio fingerprinting), signal processing (adaptive multi-resolution analysis), physical modeling and augmented instruments, amongst other topics under the computer music area. The group organizes seminars about it's members ongoing research or invited speakers (all recorded and available at `http://compmus.ime.usp.br/en/seminars`) and weekly open meetings in order to update the members about ongoing research and discuss articles and collaborations.

2 Ongoing Projects

2.1 An exploratory work in query-by-humming

Query-by-humming is a common topic in music information retrieval. In the query-by-humming task a hummed record representing imprecisely a target melody, is given to an application which is supposed to retrieve information about the target melody from a dataset. One algorithm addressing the task has to handle deviations in both time and frequency domains.

Fábio Goródscy reviews standard techniques presented in the academic literature and in commercial applications. Algorithms presented in the international conference of music information retrieval are reviewed, as well as the commercial application Soundhound, which are explored and tested. This work compares the performance of several strategies for query-byhumming within a unified query dataset.

Most of the concepts used througout the work can be found in [1]. Related work are [2][3][4][5].

The main goal of this work is to undercover the difficulties in mesuring similarity in this context, by comparing the performance of a commercial tool and also several alternative strategies.

2.2 Development of an efficient adaptive transform for music signals

Music signals can present very heterogeneous spectral characteristics, such as sharp attacks, long stationary tones, vibrato and tremolo all spread across the hearing frequency range. In such cases, fixed resolution spectrograms or even frequency-dependent resolution spectrograms (such as CQT [6]) may not result in a satisfactory representation of the signal. This is the motivation behind adaptive transforms, operators that utilize information about the signal being analyzed to compose a representation that prioritizes frequency or time resolution depending on the signal's characteristics at a given time and frequency band.

These types of algorithms are usually cost-intensive. The project currently in development as Nicolas Figueiredo's Masters thesis is the development of a low-cost adaptive transform algorithm that does not follow the traditional framework of most adaptive transforms [7, 8]. Instead of comparing between different representations (for example, STFTs calculated using 1024, 2048 and 4096-sample analysis windows) and choosing the best one for each time-frequency split, this algorithm uses bandpass filtering and undersampling [9] to isolate "interesting regions" of a given spectrogram and analyze them cheaply in greater detail. The main objectives of this project are to develop an adaptive transform whose computing cost is similar to other representations usually used in MIR tasks, and evaluate it against other multi-resolution and adaptive transforms according to their computing costs and sparsity of the resulting representation.

2.3 Using Active Acoustics techniques in musical instruments and art installations

Active Acoustics is a term used in the New Interfaces for Musical Expression (NIME) research field to describe the usage of sound and vibration inducing devices to drive electronic sounds into physical surfaces [10]. The result of inducing synthesized sounds into a complex sound radiating source can be vastly explored by artists and music technologists.

Using Active Acoustics in order to augment traditional musical instruments is an active topic in the NIME field. Nicolas Figueiredo and Paulo Itaboraí are currently

reviewing the augmented active acoustic instruments's literature in order to develop an augmented banjo. They are currently testing E.Berdahl's results of PID control on an string [11] and trying to empirically expand these results to a banjo's membrane using a BELA Board (originally called BeagleRT [12]), a piezoelectric sensor and a sound transducer.

Another possible application is to explore the acoustical properties (resonances, formants, non-linearities) of physical plates made of different materials to naturally distort and filter the electronic sounds. Paulo Itaboraí is using actuated ceramic plates to expand electroacoustic diffusion systems for acousmatic music performance. Each ceramic piece has an independent audio channel and provides an unique distortion to the sound. This enables electroacoustic composers to embed part of the sound object elaboration and distortion into spatialization gestures. Paulo, in collaboration with the composer Alex Buck, proposed and presented a sound installation called "Acting Voices - Madrigale a Sei Vasi' in the 19th International Conference on New Interfaces for Musical Expression

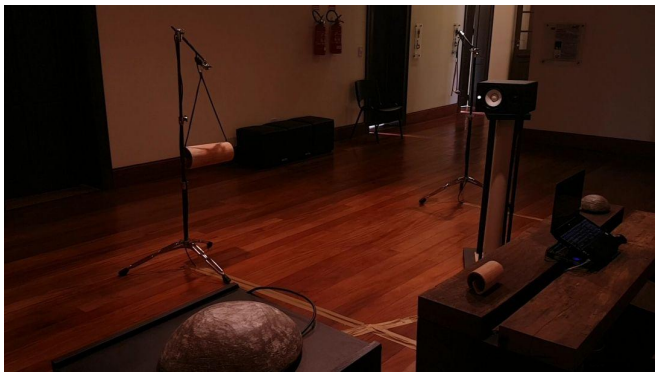

Figure 1: Installation "Active Voices - Madrigale a Sei Vasi" at NIME 2019 conference

2.4 Content Base Music Recommendation Systems

Music recommender systems (MRS) help users interacting with big digital song collections. They operate analyzing information about user's past behavior when listening to music, and suggest to each one of them the next song, album or artist to be heard. The most popular approach for implementing MRS is Collaborative Filtering (CF) [13], which associates each user to a listening profile, and assumes that similar profiles share musical preferences. According to the songs similar users interacted with, it estimates the probability of each unheard song being heard in order to make individual suggestions.

Rodrigo ir working towards one of main weakness in CF based solutions, usually described as the "cold start" problem: when new songs are included in the platform and need to be incorporate in the algorithm even without having any historical data. We propose a solution that associates the acoustic information extracted directly from the songs with user's preference [14]. In the case when there is a strong pattern in the content of the songs an user have heard so far, then it should be possible to recommend to him/her a new song that matches to his preference.

2.5 Experiments on Singing Voice Detection

Singing voice Detection in polyphonic audio signals is the problem that deals with determining which segments of a musical signal (with several sound sources) contain singing voice. This is an active topic in the Music Information Retrieval (MIR) field and has various applications, including automatic singer recognition [15], singing voice separation [16] and melody extraction [17].

Shayenne Moura started working with melody and accompaniment separation [18] and then focused her work on Singing Voice Detection (SVD), also referred as Vocal Detection. Her research is about how the SVD systems were developed in the past and what are the challenges remaining for this task [19]. She is evaluating the impact of using engineered descriptors in comparison with deep embeddings as features on the classification accuracy [20]; also, doing experiments with different mixes from the same pieces to evaluate the vocal detector under these constraints.

2.6 A framework for obtaining Musical Similarity measures

The spread of digital music allowed the appearance of datasets with millions of files. The processing of this huge number of audio files is carried out with techniques of Music Information Retrieval (MIR).

The main goal of this work is to study Musical Similarity which is to determine quantitatively how similar any two given songs are. The concept of musical similarity is subjective and there is no definition of general musical similarity. Therefore, the problem is addressed from similarities of individual musical elements, for instance, melody, harmony, tempo, metric, timbre, etc.

Roberto Bodo is working to reach this goal. He will review several measures of similarity computed between alternative representations of audio files (called audio features). With this, we can determine which songs are closest to each other, within a dataset. From the selected literature, several combinations were identified of audio features and similarity measures. At the current stage of the project, we have implemented a modular framework for obtaining similarity measures based on several features, aggregation strategies and distance models: we handled three types of similarity (timbristic, melodic, and rhythmic), and we calculated similarity matrices for a considerable number of datasets openly available.

In the future, we will explore the replacement of full songs by segments of them, analyze the obtained results and check if they extrapolate to datasets of world music. In addition, we will use deep learning techniques to learn which parts of the songs optimize the quality of the results, and create thumbnails extractors from the trained neural networks.

References

[1] meinard müller. *fundamentals of music processing: audio, analysis, algorithms, applications*. 2015.

[2] Justin Salamon, Joan Serrà, and Emilia Gómez. Tonal representations for music retrieval: from version identification to query-by-humming. *Int. J. of Multimedia Info. Retrieval, special issue on Hybrid Music Info. Retrieval*, 2(1):45–58, Mar. 2013.

[3] Ernesto López, Martín Rocamora, and Gonzalo Sosa. Búsqueda de música por tarareo, 2004.

[4] Bartłomiej Stasiak. Follow that tune – adaptive approach to dtw-based query-by-humming system. *ARCHIVES OF ACOUSTICS*, 39(4):467–476, 2014.

[5] Alexios Kotsifakos, Panagiotis Papapetrou, Jaakko Hollmen, Dimitrios Gunopulos, Vassilis Athitsos, and George Kollios. Hum-a-song: A subsequence matching with gaps-range-tolerances query-by-humming system. 5(10), 2012.

[6] Judith C Brown. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

[7] Alexey Lukin and Jeremy Todd. Adaptive time-frequency resolution for analysis and processing of audio. In *Audio Engineering Society Convention 120*. Audio Engineering Society, 2006.

[8] Florent Jaillet and Bruno Torrésani. Time-frequency jigsaw puzzle: Adaptive multiwindow and multilayered gabor expansions. *International Journal of Wavelets, Multiresolution and Information Processing*, 5(02):293–315, 2007.

[9] Rodney G Vaughan, Neil L Scott, and D Rod White. The theory of bandpass sampling. *IEEE Transactions on signal processing*, 39(9):1973–1984, 1991.

[10] Otso Lähdeoja. Active acoustic instruments for electronic chamber music. 2016.

[11] Edgar Joseph Berdahl. *Applications of feedback control to musical instrument design*. Stanford University, 2010.

[12] Andrew McPherson and Victor Zappi. An environment for submillisecond-latency audio and sensor processing on beaglebone black. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[13] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. *Recommender Systems Handbook*. Springer-Verlag, Berlin, Heidelberg, 1st edition, 2010.

[14] Rodrigo Borges and Marcelo Queiroz. Automatic music recommendation based on acoustic content and implicit listening feedback. *Revista Música Hodie*, 18(1):31 – 43, jun. 2018.

[15] A. L. Berenzweig, D. P. W. Ellis, and S. Lawrence. Using voice segments to improve artist classification of music. In *22nd Int. Conf.: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

[16] Yipeng Li and DeLiang Wang. Separation of singing voice from music accompaniment for monaural recordings. Technical report, Ohio State University Columbus United States, 2005.

[17] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770, Aug. 2012.

[18] Shayenne Moura and Marcelo Queiroz. Melody and accompaniment separation using enhanced binary masks. In *Proceedings of the 16th Brazilian Symposium on Computer Music*, pages 164 – 165, São Paulo, 2017.

[19] Kyungyun Lee, Keunwoo Choi, and Juhan Nam. Revisiting singing voice detection: a quantitative review and the future outlook. In *19th Int. Soc. for Music Info. Retrieval Conf.*, Paris, France, 2018.

[20] Shayenne Moura. Singing voice detection using vggish embeddings. *19th International Society for Music Information Retrieval Conference, Paris, France*, 2018.