TOWARDS A TIMBRAL CLASSIFICATION SYSTEM FOR MUSICAL EXCERPTS

Aurélien Antoine, Duncan Williams and Eduardo Miranda

Interdisciplinary Centre for Computer Music Research (ICCMR)
University of Plymouth, UK
`aurelien.antoine@postgrad.plymouth.ac.uk`

ABSTRACT

Searching for audio samples within a library can be a tedious and time-consuming task. In this paper, we report on the design of a pilot automatic classification system that utilises timbral properties to automatically classify audio samples. At this stage of the study, we have decided to work only with orchestral audio samples. In addition, we conducted a perceptual experiment to evaluate the performance of the system across five timbral attributes: breathiness, brightness, dullness, roughness and warmth. Promising classification results indicate that this approach may be suitable for further work that could also benefit some music production tasks.

1. INTRODUCTION

Searching for audio samples within a large library can be a tedious and time-consuming task for composers and sound designers. To assist them in this process, they can utilise new developments in music information retrieval (MIR)–for a survey of MIR systems the reader is referred to [1]. The system presented in this paper utilises timbre properties to automatically classify audio samples. Timbre represents a complex musical property that is often defined as all the sound properties except pitch, loudness and duration, which allow us to distinguish and recognise the sound of two different sounds [2]. Furthermore, several works have demonstrated the importance of acoustic features in the definition of this multidimensional attribute [3, 4].

Our initial aim for developing this system was to overcome the time-consuming task of listening to a large sound file database when composing. This search and listening task may be part of the compositional process, but it can also be a situation that composers would prefer to circumvent or make more efficient. However, a timbre-based classification system could also have applications for music production. For example, it could aid sound engineers in evaluating their mixings. The perceptual evaluation of timbral qualities utilised in this system could also be implemented in intelligent systems for music production.

In order to make this system user-friendly, we targeted verbal descriptors of timbral qualities by means of their underlying acoustic correlates, which can sometimes be complex and potentially overlapping. Terms like brightness or roughness are words from everyday language used to describe perceived musical timbres. These terms are more intuitive than their acoustic correlates (e.g. spectral centroid, critical bands).

2. THE CLASSIFICATION SYSTEM

The prototype system is implemented in the `Matlab` environment with a simple user interface. It currently integrates five timbral attributes: breathiness, brightness, dullness, roughness and warmth. We calculate the timbral index as follow:

Breathiness. Fundamental amplitude against noise content and the spectral slope [5]. The bigger the ratio between fundamental amplitude and the noise content, the breathier the sound.

Brightness. Spectral centroid [6]. The higher the spectral centroid, the brighter the sound.

Dullness. Spectral centroid [7]. A low spectral centroid value indicates that the sound is dull.

Roughness. Distance between adjacent partials in critical bandwidths and also the energy above the 6th harmonic [3].

Warmth. Spectral centroid and energy in its first three harmonics [2]. A low spectral centroid and a high energy in the first three harmonics indicate that the sound is warm.

To use the system, the user must first define the directory containing the audio files. Next, the system analyses each file and uses the acoustic correlates mentioned previously to calculate its timbral index for each attribute. File names and timbral indexes are then stored in matrices (one matrix per attribute), which are sorted in ascending order at the end, to return the best results at the top.

3. PILOT STUDY

In order to evaluate the accuracy of our classification system, we decided to run an experiment with human participants to determine the correlation between the humans' responses and the system's ratings. This perceptual evaluation was also designed to test the efficiency of the acoustic analysis when working with polyphonic timbre (timbral mixture emerging from several instruments).

3.1. Method

Training files. As training sources for the implementation of the system, we used 90 sound files generated beforehand with

`Orchids`[1], a computer-aided orchestration program developed by IRCAM[2]. These sound files consisted of short audio samples of orchestration ($\approx$ 2 to 4 seconds) and composed of several orchestral instruments, including violins, flutes and trumpets.

Stimuli. For the listening test, we decided to use 3 sound files for each timbral attribute. These stimuli were taken from the 90 orchestral sound files mentioned previously. For each attribute, we have the selected the files rated as the best, medium and worst results according to the classification performed by the system.

Procedure. We conducted the perceptual evaluation with 20 individuals, in the same room and using the same playback equipment (circumaural headphones). Each participant had to listen to 15 sound samples in total, which consisted of 3 sound samples ($\approx$ 2 to 4 seconds each) for each timbral attribute. For each attribute, participants were asked to rate the 3 audio samples using the Verbal Attribute Magnitude Estimation (VAME) rating method [8], which quantifies the applicability of the descriptor (e.g., bright $\leftrightarrow$ not bright), on a five-point Likert scale.

3.2. Results and Discussions

Due to space limitations, graphic representations of this study's results are available on the internet at the following address: `http://goo.gl/YkdOA6`.

For the timbral attribute breathiness, brightness and dullness, participants' responses and the system's rating were similar for each sample. It appears to be a correlation between the participants' responses and the classification system. The mean and standard deviation for each sample for breathiness, brightness and dullness support this interpretation.

For the attribute warmth, we can note that participants responses were similar to the system's rating for the sample identified as best result. However, there is a difference for the two other samples. This will require further investigation on our calculations for the warmth index and also on the sound samples selected in order to identify the reason participants rated these two samples similarly while there was a difference in the system's rating.

Finally, for the attribute roughness, participants rated the sample identified as medium result as the roughest sound while the system suggested a different sample. We can note that participants' responses and the system's rating were similar for the sample identified as the least rough. However, these results do not validate our calculations for the attribute roughness and further investigations are required to improve our system's rating for this attribute.

4. CONCLUSIONS AND FURTHER WORK

In this paper, we have presented a computer system developed to classify musical excerpts according to five verbal descriptor of timbral qualities. The prototype classification system is developed in `Matlab` with five timbral attributes currently implemented: breathiness, brightness, dullness, roughness and warmth. To evaluate the performances of the system, we ran a preliminary perceptual study with 20 individuals, using orchestral audio samples generated by `Orchids`. However, the same method could be applied with any source sound set (e.g. synthesised sounds).

Listener's ratings were in good agreement with the system's classification for the attributes breathiness, brightness and dullness. For the attributes roughness and warmth, the results showed a high amount of inter-participant variation in the listeners responses. This could be addressed by increasing the number of participants but at the pilot stage of this research, the high number of participants which would be required (likely hundreds) is not practical. Therefore we propose that this variation be addressed in further work by improving the classification system in response to participant ratings.

Nevertheless, these early results are promising with respect to the success of four of the five selected attributes, particularly given the absence of fully agreed or realised metrics. We think such a system facilitating the use of a particular audio descriptor to estimate attributes could be useful in musical composition as well as in intelligent music production tasks.

5. REFERENCES

[1] M. Schedl, E. Gómez, and J. Urbano, "Music information retrieval: Recent developments and applications," *Foundations and Trends® in Information Retrieval*, vol. 8, no. 2-3, pp. 127–261, 2014.

[2] R. Pratt and P. Doak, "A subjective rating scale for timbre," *Journal of Sound and Vibration*, vol. 45, no. 3, pp. 317–328, 1976.

[3] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*, vol. 22. Springer Science & Business Media, 2007.

[4] J. F. Schouten, "The perception of timbre," in *Reports of the 6th International Congress on Acoustics*, vol. 76, p. 10, 1968.

[5] M. A. Landera and R. Shrivastav, "Effects of spectral slope on perceived breathiness in vowels," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3339–3340, 2006.

[6] E. Schubert, J. Wolfe, and A. Tarnopolsky, "Spectral centroid and timbre in complex, multiple instrumental textures," in *Proceedings of the international conference on music perception and cognition, North Western University, Illinois*, pp. 112–116, 2004.

[7] D. B. Bolger, *Computational Models of Musical Timbre and the Analysis of its structure in Melody*. PhD thesis, University of Limerick, 2004.

[8] R. A. Kendall and E. C. Carterette, "Verbal attributes of simultaneous wind instrument timbres: I. von bismarck's adjectives," *Music Perception: An Interdisciplinary Journal*, vol. 10, no. 4, pp. 445–467, 1993.

---

[1]http://www.forumnet.ircam.fr/product/orchids-en/

[2]Institut de Recherche et Coordination Acoustique/Musique