Musical Genre Similarity Summary

Stephen Sinclair MUMT 611

17th March 2006

1 Introduction

Automatic recognition musical genre based on audio data is a fundamental problem in the classification and management of large and diverse musical databases, such as what we currently see in handheld personal music devices and web-based music storefronts. Relying on humans to classify databases of upwards of two million songs is time-consuming and mostly intractable. A good genre recognition system would be desireable for organization of such large databases, playlist generation, and recommendation engines for online purchasing.

Of course, the definition of the word genre may be debated. Aucouturier and Pachet (2003) acknowledges that is it merely an assumption that genre is simply music that "sounds similar". However, for the purposes of automatic classification based solely on audio data, and barring the availability of futher metadata, it must be assumed that it is a reasonable generalization.

An early paper regarding retrieval of audio recordings based on acoustic similarity was Foote (1997). Though Foote did not mention genre explicitly, he categorized audio features using a "tree-structured quantization" which is highly reminiscent of a genre hiearchy.

The earliest paper found on recognition of "musical types" based on audio data was Soltau (1998), exploring the use of neural networks to classify rhythmic structures. It was one of the few, if not the only paper to also compare machine results to human subjects' ability to classify similar genres. The findings were that the machine learning algorithm made similar mistakes in separating "rock" and "pop" styles from each other, based on beat structure.

Previous to this, there was at least one study on classification of musical style based on symbolic (MIDI) data (Dannenberg et al. 1997). However the definition of "style" was related to playing style rather than musical genre per se. Results were that various machine learning techniques could recognize playing style varied from 77% to 90% accuracy. A possible conclusion is that symbolic style recognition is a very different problem from genre recognition on audio recordings. Though more work has been done on symbolic music recognition (McKay and Fujinaga 2004; Ponce de León et al. 2004; Kranenburg and Backer 2004), it will be considered a different topic, since in general musical "style" is not the same as genre, for example analysing the difference between various composers within the Classical genre. The remainder of this summary will concentrate on analysis of audio recordings.

2 A Short History

In recent years, perhaps because of the increasing relevancy of this problem due to the popularity of large-capacity portable music players, there has been a considerable amount of research dedicated to genre recognition. Another incentive has been the Music Information Retrieval Evaluation Exchange (MIREX) competitions which have fostered competitive and open research into MIR topics.

The first classification paper to specifically use the word "genre" was Tzanetakis et al. (2001), in which a variety of timbral features are defined, as well as the introduction of the "beat histogram" for characterization of rhythmic patterns. Genre categories used in this paper were Classical, Country, Disco, HipHop, Jazz and Rock, a typical example of whatAucouturier and Pachet (2003) would later refer to as an overly simplistic genre hiearchy. In the same year, Logan and Salomon (2001) studied similarity between individual recordings. However this paper did not refer to genre classification, instead focusing on song-to-song "audio similarity".

Also in 2001, Elias Pampalk wrote a master's thesis on a system for visualization of genre classification entitled "Islands of Music" (Pampalk 2001). It was based on the principal of Self-Organizing Maps introduced by T. Kohonen in 1982. The feature set used for classification was related more to loudness perception than timbral perception, which would eventually lead to his 6th-place MIREX'05 entry (Pampalk 2005), discussed below.

In 2002, Jiang et al. (2002) proposed the use of "spectral contrast features", which are meant to reflect the relative distribution of harmonic and inharmonic components in the spectrum. A negatively titled paper discussed the issues in using timbral properties to evaluate musical similarity, and introduced a method of exploiting contradictions between different similarity measures(Aucouturier and Pachet 2002)

The following year, the same authors published an exhaustive summary of the field, breaking down similarity techniques into the categories of timbral, rhythmic, and pitch-based similarity (Aucouturier and Pachet 2003). A discussion of collaborative filtering and data-mining techniques (co-occurence analysis) for extracting textual metadata was also included. Also published in 2003 was an initial look at methods for comparison of audio similarity algorithms (Pampalk et al. 2003), a subject which would soon become an important issue for the MIREX competitions.

Pampalk (2004) published a toolbox of Matlab functions to be used for computing music similarity. Zhu et al. (2004) focused on determining genre by explicitly trying to match instrumental features. Specifically, they used spectral recognition algorithms against a set of known spectral signatures to recognize musical instruments used in a recording, and also record the mean and standard deviation of pitches, extracted from the audio by polyphonic pitch detection routines, assuming that the classification routines are adequate to smooth error rates. They found this method to effectively increase accuracy over MFCC-only methods in genre classification between Jazz, Rock, Pop and Classical by a factor of 19%.

3 State of the Art

In reviewing the literature, it was found that a large portion of publications related to genre classification in 2005 were published in the proceedings of the ISMIR conference and the MIREX competition. Most of this research seemed to follow similar techniques as the previous publications, attempting to fine-tune feature selection and classification techniques in order to increase performance measures.

The winner of MIREX'05 used standard MFCC techniques as well as some other features, but also took advantage of the AdaBoost algorithm to increase classifier accuracy (Bergstra et al. 2005). Boosting classifiers in a similar way was also proposed in the same year by Bağci and Erzin (2005). The next-best entry also used MFCC features, but implemented support vector machines (SVM) as the classifier (Mandel and Ellis 2005). An entry by West (2005) used spectral irregularity as a feature in addition to MFCC.

The entry by Pampalk (2005) introduced the use of fluctuation patterns in perceptual loudness as an additional characteristic of the audio stream, along with two values extracted from FP data which he calls Focus and Gravity. This can be seen as an attempt to go beyond simple spectral analysis of the recording. However it should be noted that his technique did not beat out Bergstra's MFCC-based entry, coming in at 6^{th} place. He went into further detail on the technique in his ISMIR submission (Pampalk et al. 2005). There were several other MIREX submissions, which may be found at the MIREX web site.

References

Aucouturier, J.-J. and F. Pachet. 2002. Music similarity measures: What's the use? *Proceedings of the 3rd International Symposium on Music Information Retrieval*.

Aucouturier, J.-J. and F. Pachet. 2003. Representing musical genre: A state of the art. *Journal of New Music Research. 32*(1), 83–93.

Bağci, U. and E. Erzin. 2005. Boosting classifiers for music genre classification. *International Symposium on Computer and Information Sciences*.

Bergstra, J., N. Casagrande, and D. Eck. 2005. Genre classification: Timbre- and rhythm-based multiresolution audio classification. MIREX genre classification contest.

Dannenberg, R. B., B. Thom, and D. Watson. 1997. A machine learning approach to musical style recognition. *Proceedings of the International Computer Music Conference*, pp. 344–7.

Foote, J. T. 1997. Content-based retrieval of music and audio. C.-C. J. Kuo et al. (Eds.), *Multimedia Storage and Archiving Systems II, Proceedings of SPIE*, Volume 3229, pp. 138–47.

Jiang, D.-N., L. Lu, H.-J. Zhang, J.-H. Tao, and L.-H. Cai. 2002. Music type classification by spectral contrast feature. *Proceedings of the IEEE International Conference on Multimedia and Expo*.

Kranenburg, P. V. and E. Backer. 2004. Musical style recognition - a quantitative approach. *Proceedings of the Conference on Interdisciplinary Musicology*.

Logan, B. and A. Salomon. 2001. A music similarity function based on signal analysis. *Proceedings of the IEEE International Conference on Multimedia and Expo*.

Mandel, M. I. and D. P. Ellis. 2005. Song-level features and support vector machines for music classification. MIREX genre classification contest.

McKay, C. and I. Fujinaga. 2004. Automatic genre classification using large high-level musical feature sets. *Proceedings of the International Conference on Music Information Retrieval*.

Pampalk, E. 2001. Islands of music: analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology.

Pampalk, E. 2004. A matlab toolbox to compute music similarity from audio. *Proceedings of the International Conference on Music Information Retrieval*.

Pampalk, E. 2005. Speeding up music similarity. MIREX genre classification contest.

Pampalk, E., S. Dixon, and G. Widmer. 2003. On the evaluation of perceptual similarity measures for music. *Proceedings of the International Conference on Digital Audio Effects*.

Pampalk, E., A. Flexer, and G. Widmer. 2005. Improvements of audio-based music similarity and genre classification. *Proceedings of the International Conference on Music Information Retrieval*.

Ponce de León, P. J., C. Pérez-Sancho, and J. Iñesta. 2004. A shallow description framework for music style recognition. A. Fred et al. (Eds.), *Lecture Notes in Computer Science*, Volume 3138, pp. 876–84. Springer-Verlag, Germany.

Soltau, H. 1998. Recognition of musical types. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*.

Tzanetakis, G., G. Essl, and P. Cook. 2001. Automatic musical genre classification of audio signals. *Proceedings of the International Conference on Music Information Retrieval*.

West, K. 2005. Mirex audio genre classification. MIREX genre classification contest.

Zhu, J., X. Xue, and H. Lu. 2004. Musical genre classification by instrumental features. *Proceedings of the International Computer Music Conference*.