19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

PERCEPTUAL CONSTRAINTS FOR MUSIC INFORMATION RETRIEVAL SYSTEMS

PACS: 43.75.Xz

McKinney, Martin F.[1]; Skowronek, Janto[1]; Novello, Alberto[1]
[1] Philips Research Laboratories; High Tech Campus 36, 5656 AE Eindhoven, The Netherlands; {martin.mckinney,janto.skowronek,alberto.novello}@philips.com

ABSTRACT

In developing systems for music information retrieval (MIR), perceptual constraints often provide useful guidelines for design and evaluation. We show here three different types of MIR systems and illustrate how perceptual constraints shape their development.

Perceptual similarity of music is an important attribute for music browsing, searching and playlisting. We are conducting a large-scale experiment in which listeners rank the similarity of musical-excerpt pairs. From the data we will create a low-dimensional space of perceptual music similarity and then attempt to assign the dimensions to acoustic qualities.

Previous work has shown that the perceptual tempo of music can be ambiguous in that tempi of different metrical levels can be equally perceived. Our system for automatic music tempo extraction takes this tempo ambiguity into account and provides estimates of the relative perceptual salience of different tempi for individual pieces of music.

Finally, we have developed a method for music "mood" classification that is based on extensive perceptual research aimed at identifying the most objective mood types across listeners. Here, we use mood labels that were judged consistent across listeners and thus provide a more useful set of communication criteria to relate the mood of music.

INTRODUCTION

Music consumers have access to vast (and growing) amounts of music audio due to recent advances in audio compression technology as well as increases in storage capacities and broadcast bandwidths. To deal with this wealth of music, users need tools to navigate, browse and search for music. Traditionally, such tools have been based on *annotated* metadata, such as genre and artist, but recent developments in automatic music information retrieval (MIR) are making it possible to extract some types of metadata directly from the music audio signal itself [1]. These additional metadata, combined with traditional annotated metadata, provide a much richer set upon which sophisticated systems for music navigation can be built.

While many aspects of algorithms for MIR can be constrained by, or defined in terms of, musicological, sociological, and/or computational factors, we have found that perceptual criteria are often valuable guides. This is true not only for the evaluation of, but also for the development and construction of MIR systems. While the advantage of using perceptually-based constraints depends somewhat on the application context, we present here three cases in which perceptual knowledge clearly benefits or is needed for the development of algorithms. More specifically, we have examined automatic extraction of music similarity ratings, tempo extraction, and music mood classification.

MUSIC SIMILARITY

Music similarity is an important characteristic for music browsing and playlisting. Music consumers and listeners often describe their desire for a particular type of music by saying, "I'd like something that sounds like this song or that artist." There has been much recent work on methods to automatically evaluate similarity between two musical pieces based on signal analysis (see [6] for an overview). While some studies have used listening tests as a method for measuring the performance of such systems, very little work has been done on actually characterizing perceptual similarity. Data on, or a model of, music similarity perception would be a valuable tool in guiding systems for music similarity evaluation.

We recently developed a methodology for characterizing the perceptual similarity between pieces of music using triadic comparisons [5]. Listeners were asked to rank the similarity between pairs of musical excerpts, which were presented three at a time. They indicated which pair of the three excerpts were the most similar and which were the most dissimilar. We then generated a (dis)similarity matrix from the rankings (assigning a distance of 1 for the most similar pairs, 3 for the least similar pair and 2 for the remaining "middle" pair. Finally, we performed multidimensional scaling on the data to yield a multi-dimensional "perceptual similarity" space, in which all eighteen musical excerpts were placed. Figure 1 shows the results of the MDS for two dimensions. The figure shows clustering of certain genres and suggests that music genre is a strong factor that correlates highly with similarity ratings. While not necessarily a surprising result, our finding that genre, which represents a constellation of factors, correlates with similarity provides support for our method. A logical next step is to now broaden our database with a larger-scale rating experiment and then use the data on perceptual similarity to calculate a mapping between the music audio feature space and the perceptual similarity space. A larger-scale perceptual experiment is currently underway.

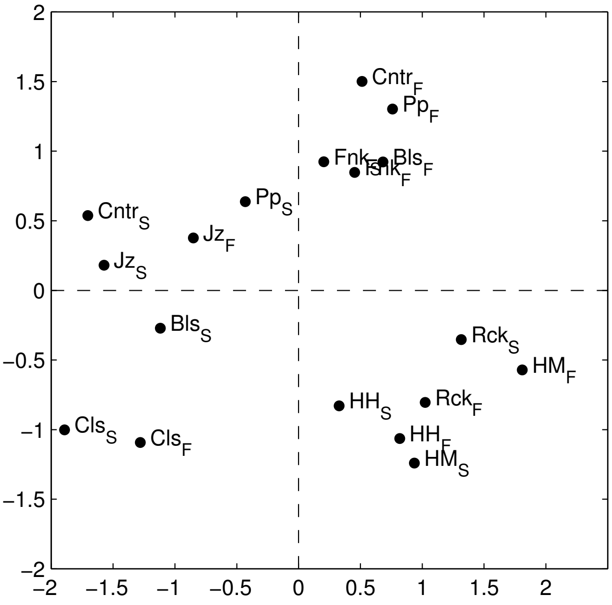

Figure 1: Position of music excerpts in a 2-dimensional space obtained through MDS. The stress of the MDS fit was 0.244.

TEMPO EXTRACTION

Music tempo is a useful statistic for music selection, comparison, mixing and playlisting. While it is common to characterize music tempo by the *notated* tempo on sheet music in beats per minute (BPM), this value often does not correspond to the *perceptual* tempo [3].

For the applications listed above, it is in fact the perceptual tempo that is of interest and a single value is clearly not adequate. We have therefore established a method and protocol in which we report two tempo values as well as their relative perceptual salience. This method was used for the evaluation criteria in the MIREX 2005 tempo extraction contest [4]. We performed a series of experiments in which we asked listeners to tap to the beat of musical excerpts and computed the two most salient tempi and their relative salience. Figure 2 shows an example of a histogram of tapped tempi for a single excerpt of music. In this case the responses were almost evenly split between two tempi, each at a different metrical level of the musical excerpt.

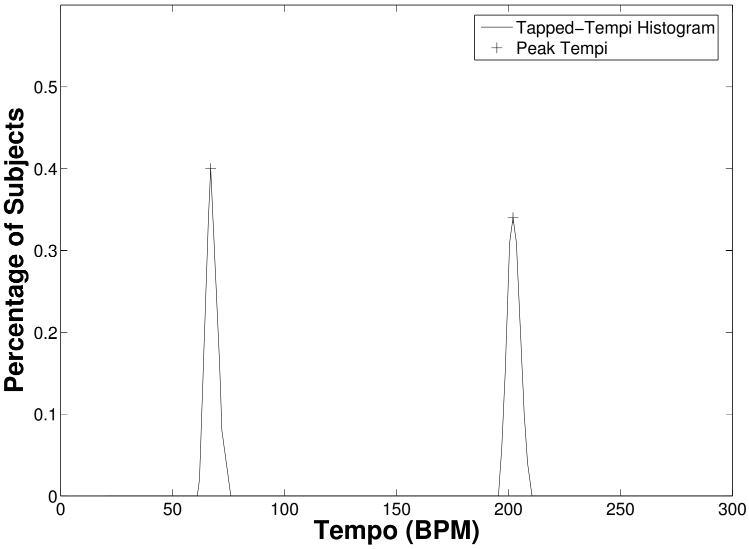

Figure 2: Histogram of subjects' tapped tempi for a single excerpt of music. Two distinct peaks in the histogram indicate a split of perceived tempi across listeners and support the idea of annotating tempo with more than a single value.

In addition to using perceptual criteria for the *evaluation* of tempo extractors, we have also seen benefits in applying perceptually-based signal processing models to algorithms for automatic tempo extraction. Our tempo extractor conforms to a conventional two-stage structure, in which a *driving* signal is derived from the audio signal in the first stage, and is then used to drive a series of periodicity detectors in the second stage. In our case, the driving signal is a pulse train derived from an onset detection process, where each pulse represents the onset of an auditory object. The onset detector features two processing components, modeled after perceptual phenomena, that help boost performance, not only in the detection of onsets but also in tempo extraction [7].

MUSIC MOOD CLASSIFICATION

Interest in automatic music mood classification is increasing because it could enable people to browse and manage their music collections by means of the music's emotional expression complementary to the widely used music genres. In the development of an automatic music mood classifier, we must treat the high degree of subjectivity associated with mood. Our approach to reduce subjectivity is first to define mood classes on which users show a relative agreement when applying them to music and second to use only that music in our ground-truth database that conveys emotions well enough such that listeners can assign a mood label to the music track.

With a series of four experiments we intended to setup of a proper ground-truth database for music mood classification and develop and evaluate a corresponding classification algorithm. The first subjective experiment [9] aimed at identifying a set of mood classes that have been judged consistent across listeners. Experiment two had the purpose of collecting music material for the desired ground-truth database. We gathered subjective ratings on the chosen mood classes for a large number of music excerpts and chose only excerpts for the database that got a clear judgment from the subjects. In experiment three we trained individual detector models for each mood class, because the used classes were not mutually exclusive. The detectors used a set of different types of features [2, 3, 10, 8] and quadratic discriminant analysis (QDA). With a randomized

Mood Class	Performance
arousing-awakening	85.1 ± 1.7 %
angry-furious-agressive	90.1 ± 1.6 %
calming-soothing	90.9 ± 2.4 %
carefree-lighthearted-light-playful	77.1 ± 1.7 %
cheerful-festive	79.8 ± 1.9 %
emotional-passionate-touching-moving	82.2 ± 1.1 %
loving-romantic	80.2 ± 1.7 %
peaceful	88.5 ± 1.7 %
powerful-strong	80.7 ± 1.8 %
sad	85.0 ± 1.8 %
restless-jittery-nervous	90.6 ± 1.5 %
tender-soft	89.5 ± 0.6 %

Table I: Classification performance (mean ± standard error across bootstrap repetitions) of the individual mood detectors.

80/20 split of the training and test data and using repetitions with bootstrapping, we estimated the classification performance of the detectors. Results are shown in Table I. In a fourth experiment, which is currently under preparation, subjects will evaluate the quality of the mood estimations for abitrary music that is not part of our thoroughly designed ground-truth mood database.

The obtained classification results show that an automatic estimation of music mood, as we defined it by means of the subjective experimental method, is in general possible.

CONCLUSIONS

We have seen performance benefits from using perceptual criteria and models in systems for automatic music information retrieval. Our most comprehensive example comes from our music tempo extractor where it was not only the evaluation criteria that were perceptually-based, but acoustic processing components as well. Our work in music similarity and music mood classification suggest that careful collection and treatment of perceptual data can provide robust design and evaluation guidelines in these areas as well.

References

[1] J. S. Downie. Music information retrieval. Annual Review of Information Science and Technology **37 (2003), no. 1** 295–340

[2] M. F. McKinney, J. Breebaart. Features for audio and music classification. In *Proceedings of the 4th International Conference on Music Information Retrieval*. Johns Hopkins University, Baltimore, MD (2003)

[3] M. F. McKinney, D. Moelants. Extracting the perceptual tempo from music. In *Proceedings of the 5th International Conference on Music Information Retrieval*. Pompeu Fabra University, Barcelona (2004)

[4] M. F. McKinney, D. Moelants. Mirex 2005: Tempo contest. In *Proceedings of the 6th International Conference on Music Information Retrieval*. Queen Mary, University of London, London (2005)

[5] A. Novello, M. F. McKinney, A. Kohlrausch. Perceptual evaluation of music similarity. In *Proceedings of the 7th International Conference on Music Information Retrieval*. University of Victoria, Victoria, Canada (2006)

[6] E. Pampalk. *Computational Models of Music Similarity and their Application to Music Information Retrieval*. Ph.D. thesis, Vienna University of Technology, Vienna (2006)

[7] J. E. Schrader. *Detecting and interpreting musical note onsets in polyphonic music*. Master's thesis, Eindhoven University of Technology, Eindhoven (2003)

[8] J. Skowronek, M. McKinney. *Intelligent Algorithms in Ambient and Biomedical Computing*, volume 7 of *Philips Research Book Series*. Springer, Dordrecht, NL (2006) 103–118, 103–118

[9] J. Skowronek, M. McKinney, S. van de Par. Ground truth for automatic music mood classification. In *Proceedings of the 7th International Conference on Music Information Retrieval*. Victoria, Canada (2006)

[10] S. van de Par, M. McKinney, A. Redert. Musical key extraction from audio using profile training. In *Proceedings of the 7th International Conference on Music Information Retrieval*. Victoria, Canada (2006)