MUSIC IMAGERY INFORMATION RETRIEVAL
BRINGING THE SONG ON YOUR MIND BACK TO YOUR EARS

Sebastian Stober
Data & Knowledge Engineering Group
Otto-von-Guericke-Universität Magdeburg, DE
stober@ovgu.de

Jessica Thompson
Bregman Music and Auditory Research Studio
Dartmouth College, USA
jessica.thompson@dartmouth.edu

1. INTRODUCTION

Most existing Music Information Retrieval (MIR) technologies require a user to use a query interface to search for a musical document. The mental image of the desired music is likely much richer than what the user is able to express through any query interface. This expressivity bottleneck could be circumvented if it was possible to directly read the music query from the user's mind. To the authors' knowledge, no such attempt has been made in the field of MIR so far. However, there have been recent advances in cognitive neuroscience that suggest such a system might be possible. Given these new insights, it seems promising to extend the focus of MIR by including music imagery—possibly forming a sub-discipline which could be called *Music Imagery Information Retrieval (MIIR)*. As a first effort, there has been a dedicated session at the *Late-Breaking & Demos* event at the *ISMIR 2012* conference. This paper aims to stimulate research in the field of *MIIR* by laying a roadmap for future work.

2. PRIOR WORK ON NEURAL DECODING

Neural decoding is an exploding area of research in modern neuroscience. Recent work in this area has demonstrated the reconstruction of speech [9] and video [8] stimuli from electrocorticographic (ECoG) and functional magnetic resonance imaging (fMRI) recordings. Little work has been done in this area for music. Musical features such as genre labels [2] and the direction of melodies [5] have been classified from hemodynamic brain activity. A recent study by Schaefer et al. were able to discriminate between electroencephalography (EEG) responses to seven different musical stimuli with a single-trial accuracy of 70% [11]. These results suggest that it may be possible to decode perceived music from the listening brain.

Generally, there is strong evidence that perception and imagination of music share common processes in the brain. In his recent review of the literature on auditory imagery, Hubbard [4] concludes that "auditory imagery preserves many structural and temporal properties of auditory stimuli" and "involves many of the same brain areas as auditory perception." This is also underlined by Schaefer [10, p. 142] whose "most important conclusion is that there is a substantial amount of overlap between the two tasks [music perception and imagery], and that internally creating a perceptual experience uses functionalities of normal perception." Thus, brain signals recorded while listening to a music piece could serve as reference data for a retrieval system in order to detect salient elements in the signal that could be expected during imagination as well. Furthermore, it has been shown that people are remarkably accurate at reproducing the tempo of a piece of music recalled from long-term memory [6]. This implies that there is little tempo variation between different recollections of the same song, which makes identifying and matching common patterns much easier. For imagined rhythms, a method for detecting temporal patterns has already been sketched in [3].

The potential of *music imagery*, i.e., "the process of deliberately imagining well-known music, by recreating the perceptual experience internally" [10, p. 3] as a possible paradigm for EEG-based brain-computer interface (BCI) design has very recently been investigated by Schaefer et al. They argue that "music is especially suitable to use here as (externally or internally generated) stimulus material, since it unfolds over time, and EEG is especially precise in measuring the timing of a response." This allows us to exploit temporal characteristics of the signal such as rhythmic information. Furthermore, music processing engages numerous distributed networks across the brain allowing us to extract musically relevant information from many electrode sites.

Unfortunately, the ability to classify purely imagined music fragments has not been examined. However, encouraging preliminary results were reported in [11] where four out of eight participants produced imagery that was classifiable (in a binary comparison) with an accuracy between 70% and 90% after 11 trials. Schaefer et al. obtained their results using generic classification techniques not specifically tailored to their problem. We suggest that the effectiveness of auditory neural decoding could be significantly improved by incorporating an MIR perspective and applying sophisticated signal processing and retrieval techniques. We also suggest that the success of such research could pave the way for exciting new MIR applications.

3. MIIR SCENARIOS

Many popular MIR scenarios could be transferred to the music imagery domain by regarding brain signals recorded during music perception and imagination as just another representation of music. One possible scenario would be: given an EEG signal, identify the corresponding music piece. This is similar to the EEG song identification experiment conducted by Schaefer et al. [11]. However, their method quickly becomes impractical for larger datasets and doesn't extend to music for which we have no corresponding brain activity. As an alternative approach, sophisticated indexing and perceptual fingerprinting techniques could be applied that have been developed in MIR for song identification. For the latter, the common objective is essentially to identify heavily distorted and degraded versions of the same recording [1], with popular approaches being "capable of quickly identifying a short segment of music captured through a cellphone microphone in the presence of foreground voices and other dominant noise, and through voice codec compression, out of a database of over a million tracks" [12]. Applied to music imagery, such techniques could be used to match an EEG recording of music imagery against a database of EEG recordings of music perception for the same song.

In some cases, we may wish to also time align the EEG signal to the corresponding audio. Here, MIR techniques for score following and score-performance / multimodal synchronization could be applied such as described by Müller et al. [7] by considering music imagery as another modality. Given a sufficient accuracy of the synchronization can be achieved, playing back the recognized fragments of recordings along to the imagination would be a natural way of giving feedback helping users to better control their imagination. Such immediate feedback could cause imagination and perception to resonate—helping the user to better concentrate on the one hand and the system on the other by a stronger input signal. If imagination and perception do not match, a dissonance effect can be expected, which also could be exploited to guide an adaptation of the retrieval model.

Additionally, we would want to develop MIIR technologies that could be used with music for which we have no corresponding brain data. Rather than modeling brain responses to individual pieces of music, we can find mappings between brain features and music features. For example, once a mapping between EEG and audio features has been learned, this trained model can be used to predict brain activity for a database of unheard music. This is the method used in the Nishimoto et al. paper on video reconstruction from fMRI [8]. These 'brain features' may be especially useful for music recommendation and music composition. In the case of music recommendation, a user could submit a query by thinking about the characteristics of the music he or she wishes to retrieve. In the case of composition, brain computer interfaces could be developed to render imagined music from brain activity. In fact, an adequately trained model could be used to predict brain activity for arbitrary music, after which these brain features may be used in any MIR application just as any other music feature.

4. CONCLUSIONS

Music Imagery Information Retrieval (MIIR) is a potential future sub-discipline of MIR, building upon recent advances in the fields of cognitive neuroscience and MIR with the goal to develop techniques that allow to query a MIR system by thinking about the music to be retrieved. Mind reading is undoubtedly an ambitious venture but, as outlined in this paper, computational neuroscientists have already made critical advances towards this goal. Regardless of the feasibility of the outlined MIIR scenarios, it seems clear the fields of MIR and cognitive neuroscience could both benefit largely from the insights gained by addressing this problem in an interdisciplinary collaboration. Furthermore, this provides a great chance to disseminate MIR techniques into the field of cognitive neuroscience, which may ignite other future collaborations.

5. REFERENCES

[1] Pedro Cano, Eloi Batlle, Ton Kalker, and Jaap Haitsma. A review of audio fingerprinting. *The Journal of VLSI Signal Processing*, 41:271–284, 2005.

[2] Michael Casey, Jessica Thompson, Olivia Kang, and Thalia Wheatley. Population codes representing musical timbre for high-level fMRI categorization of music genres *Machine Learning and Interpretation in Neuroimaging, Springer Lecture Notes on Artificial Intelligence (LNAI)*, 2012.

[3] Peter Desain. What rhythm do I have in mind? Detection of imagined temporal patterns from single trial ERP. In *Proc. of Int. Conf. on Music Perception and Cognition (ICMPC)*, page 209, 2004.

[4] Timothy L. Hubbard. Auditory imagery: empirical findings. *Psychol Bull*, 136(2):302–329, Mar 2010.

[5] Yune-Sang Lee, Petr Janata, Carlton Frost, Michael Hanke, and Richard Granger. Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI *NeuroImage*, 57(1):293–300, 2011.

[6] Daniel Levitin and Perry Cook. Memory for musical tempo: Additional evidence that auditory memory is absolute. *Attention, Perception, & Psychophysics*, 58:927–935, 1996.

[7] Meinard Müller, Frank Kurth, and Tido Röder. Towards an efficient algorithm for automatic score-to-audio synchronization. In *Proc. of 5th Int. Conf. on Music Information Retrieval*, 2004.

[8] Shinji Nishimoto, An T. Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L. Gallant. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies *Current Biology*, 21(19):10.1016/j.cub.2011.08.031, 2011.

[9] Brian N. Pasley, Stephen V. David, Nima Mesgarani, Adeen Flinker, Shihab A. Shamma, Nathan E. Crone, Robert T. Knight, and Edward F. Chang. Reconstructing speech from human auditory cortex. *PLoS Biol*, 10(1):e1001251, 01 2012.

[10] Rebecca S. Schaefer. *Measuring the mind's ear: EEG of music imagery*. PhD thesis, Radboud University Nijmegen, Nijmegen, The Netherlands, 2011.

[11] Rebecca S. Schaefer, Yvonne Blokland, Jason Farquhar, and Peter Desain. Single trial classification of perceived and imagined music from EEG. In *Berlin BCI Workshop*, 2009.

[12] Avery Wang. An industrial strength audio search algorithm. In *Proc. of 4th Int. Conf. on Music Information Retrieval*, 2003.