NEUROIMAGING METHODS FOR MUSIC INFORMATION RETRIEVAL: CURRENT FINDINGS AND FUTURE PROSPECTS

Blair Kaneshiro
Center for Computer Research in Music and Acoustics
Stanford University, Stanford, CA, USA
blairbo@ccrma.stanford.edu

Jacek P. Dmochowski
Department of Psychology
Stanford University, Stanford, CA, USA
dmochowski@gmail.com

ABSTRACT

Over the past decade and a half, music information retrieval (MIR) has grown into a robust, cross-disciplinary field spanning a variety of research domains. Collaborations between MIR and neuroscience researchers, however, are still rare, and to date only a few studies using approaches from one domain have successfully reached an audience in the other. In this paper, we take an initial step toward bridging these two fields by reviewing studies from the music neuroscience literature, with an emphasis on imaging modalities and analysis techniques that might be of practical interest to the MIR community. We show that certain approaches currently used in a neuroscientific setting align with those used in MIR research, and discuss implications for potential areas of future research. We additionally consider the impact of disparate research objectives between the two fields, and how such a discrepancy may have hindered cross-discipline output thus far. It is hoped that a heightened awareness of this literature will foster interaction and collaboration between MIR and neuroscience researchers, leading to advances in both fields that would not have been achieved independently.

1. INTRODUCTION

Since its inception, music information retrieval (MIR) has been characterized as an interdisciplinary and multifaceted field, drawing from such diverse domains as information science, music, computer science, and audio engineering to explore topics ranging from indexing and retrieval to musical analysis and user studies [22, 24]. The field has become increasingly collaborative over time, and cross-disciplinary output has grown [33].

However, one field that has yet to establish itself as a definitive sub-discipline of MIR is that of neuroscience. Recent papers by Aucouturier and Bigand [6, 7] have highlighted the challenges faced by MIR researchers attempting to publish in cognitive science and neuroscience journals, pointing out that MIR approaches have occupied at best a marginal or incidental role in that literature. The authors cite as a main obstacle a fundamental lack of interest, or understanding, from the cognitive science/neuroscience community. At the same time, the few brain-based MIR studies published to date [16, 40, 52] have emphasized application over background, potentially leaving readers lacking sufficient introduction to the imaging technique and brain response of interest. As things currently stand, the fields of MIR and neuroscience operate largely independently, despite sharing approaches and questions that might benefit from cross-disciplinary investigation.

In an effort to begin reconciling these two fields, the present authors—whose backgrounds collectively span music, neuroscience, and engineering—present a review of studies drawn from the music neuroscience literature and examine their relevance to MIR research. While such a review will not immediately resolve the significant philosophical issues described above, it may perhaps open a window between the two disciplines by highlighting shared approaches and potential collaborations while acknowledging differences in aims and motivations. Envisioned outcomes are twofold: First, that MIR researchers may find, in brain responses, a new setting to apply analysis techniques already developed for other types of data; and second, and more importantly, that heightened awareness of this literature will increase collaborations between MIR and neuroscience researchers, advancing both fields and leading to the formation of a robust cross-discipline.

Since a review of the entire literature on music and neuroscience would be beyond the scope of this paper, we narrow the present focus to approaches that align closely with MIR applications. For rigor, we include only peer-reviewed papers, though interested readers are encouraged to visit other venues—including but not limited to ICMPC, SMPC, and late-breaking ISMIR proceedings—for a wealth of additional ideas and findings. The primary focus here is on EEG, though behavioral and fMRI studies will be touched upon as appropriate.

The remainder of this paper is structured as follows. First, we evaluate the suitability of various neuroimaging modalities for MIR research (§2). We then review three neuroimaging approaches used in music research (§3) and consider how these methods, and others, might be used for MIR research (§4). We conclude with a discussion of diverging objectives between the two fields, and opportunities for future cross-disciplinary research (§5).

2. NEUROIMAGING METHODS FOR MIR

Neuroimaging is the use of magnetic, electrical, hemodynamic, optical, or chemical means to measure activity in the central nervous system, most often the cerebral cortex (Table 1). The central idea behind bridging neuroimaging with MIR is that music is encoded by the brain, and thus can be "read out" or decoded using imaging techniques. In order to exploit this idea, it would be advantageous to track neural activity at the temporal resolution of music (i.e., milliseconds), which necessitates the use of techniques that provide direct electromagnetic measures of neural activity. While techniques measuring hemodynamic responses, such as functional magnetic resonance imaging (fMRI), provide superb spatial resolution that can indirectly probe neural activation on a millimeter scale and elucidate the functional brain networks recruited to process music, the sluggishness of these responses makes them less likely to play a role in MIR.

	EEG	MEG	ECoG	fMRI	DTI
Temporal Resolution	high	high	high	low	NA
Spatial Resolution	low	low	high	high	high
Invasiveness	low	low	high	low	low
Mobility/Portability	high	low	low	low	low
Field of View	large	large	small	large	large
Expense to Operate	low	high	NA	high	high

Table 1. Characteristics of neuroimaging techniques frequently used in music and auditory research. Adapted from Mehta and Parasuraman [39].

On the other hand, electroencephalography (EEG) and magnetoencephalography (MEG) provide millisecond temporal resolution that can in principle be used to infer properties of the stimuli evoking encephalographic responses. EEG and MEG consist of sensors placed at or near the scalp surface that detect mass superpositions of activity in the cerebral cortex. The signal-to-noise ratio (SNR) of EEG/MEG is inherently low, typically on the order of -20 dB. However, as activity is usually collected over a spatial aperture consisting of tens or hundreds of sensors, multivariate approaches can be used to derive spatial filters that will enhance the desired signal while suppressing the noise. The limitation of EEG/MEG is low spatial resolution that results from a spatial smoothing of the evoked signal and renders it difficult to localize the underlying source. In order to achieve fine resolution in both space and time, electrodes can be placed directly on the cortical surface, an invasive practice that is feasible only in the case of neurological disease where it is known as electrocorticography (ECoG), which has been recently employed to study processing of music [47, 48, 55]. Note, however, that in the context of MIR, precise spatial localization is likely not a fundamental requirement. All of the above techniques refer to imaging the function of the brain; methods that measure the connections among brain areas, such as diffusion tensor imaging (DTI), have also been used in the context of music research (e.g., [38]).

In order to feasibly integrate neuroimaging with MIR, a form of imaging that is inexpensive, noninvasive, and finely temporally resolved is required. For these reasons, our primary focus in the present paper is on EEG, which represents the most promising modality for bridging neural responses with MIR. Moreover, EEG offers a whole-brain field of view that allows for studying the interaction of distributed brain areas during musical processing.

3. APPROACHES OF INTEREST

In this section we review three approaches that may prove useful for MIR. The first is an early-latency response generated by the auditory brainstem, while the latter two involve longer latency cortical responses.

3.1 The Frequency-Following Response

The frequency-following response (FFR) is an early-latency subcortical response generated by the auditory brainstem less than 10 msec after an auditory stimulus occurs. It is a sustained, phase-locked response that oscillates at the same frequency as an auditory stimulus to such an extent that the stimulus can be "played back" from an average of many trials of the brain response [25].

The FFR is typically recorded from a single electrode at the vertex of the head, plus reference and ground electrodes. The response is averaged over many stimulus presentations, and is usually analyzed in the frequency or time-frequency domain. The FFR has an especially low SNR; therefore, FFR experiments require on the order of hundreds or thousands of stimulus presentations. The frequency range of interest for this response is primarily under 1,000 Hz, and studies presented here generally use complex, synthesized stimuli with fundamental frequencies no greater than 300 Hz. An introduction to the response and technique can be found in the 2010 tutorial by Skoe and Kraus [51], and recent findings pertaining to music are summarized in a 2013 review by Bidelman [9].

Despite being an early, low-level auditory response, the FFR has been found to show effects of learning-based neural plasticity. Its involvement in the music literature grew out of speech studies that compared subcortical responses of speakers of tone languages, such as Mandarin and Thai, to those of English speakers. These studies showed that FFRs to certain pitch-varying phonemes and phoneme-like stimuli were more robust in the tone-language speakers than in the English speakers, pointing to experience-dependent processing enhancements [29–32, 56]. Trained musicians, who possess a complementary type of pitch expertise, became a population of interest in generalizing these findings. For example, a study by Wong et al. [62] showed that musicians exhibited more robust encoding of Mandarin phonemes than did nonmusicians, despite not being tone-language speakers.

The first study to investigate the FFR specifically in response to musical stimuli was a 2007 study by Musacchia and colleagues [41]. Here, musicians' enhanced subcortical encoding of speech and musical stimuli presented in audio, visual, and audiovisual modalities could be identified in both the time and frequency domains of the brain

response. Subsequent studies have investigated encoding of musical intervals by musicians and nonmusicians [34], as well as encoding of music by both musicians and Mandarin speakers [10, 11].

Musical characteristics of the stimuli have also been found to modulate the strength of the FFR. A 2009 study by Bidelman and Krishnan [12], revealed enhanced encoding for consonant versus dissonant musical intervals. The authors later found a similar effect in responses to pleasant (major/minor) versus unpleasant (augmented/diminished) triads [13]. It should be noted that these results cannot be merely a reflection of the acoustical properties of the stimuli, as the consonant and dissonant intervals are interleaved (e.g., the dissonant tritone lies between the consonant P4 and P5), as are the constituent intervals (major and minor thirds) comprising the different types of musical triads.

3.2 Single-Trial EEG Classification

We now move from the auditory brainstem to the cerebral cortex, where responses begin roughly 50 msec after stimulus onset and are typically recorded from between 32–256 electrodes arranged across the surface of the scalp at regular intervals, often by means of a cap or net. Cortical responses are generally analyzed in a lower frequency range than FFRs, usually below 50 or 60 Hz.

Cortical EEG research has a long history of univariate analysis. Readers may be familiar with time-averaged event-related potential (ERP) studies, which focus on amplitudes and latencies of particular waveform peaks from selected electrodes. Some recent studies have taken a different approach to EEG analysis by classifying single trials of the brain response. The goal in this case is to correctly predict, from the brain response, which stimulus the participant was experiencing (see Blankertz et al. [15] for an introduction and tutorial). This multivariate approach enables data from multiple electrodes and time points to be analyzed at once. Classification of neuroimaging data has a longer history in fMRI (as multi-voxel pattern analysis [43]) than in EEG; however, the overarching methodology lends itself well to extracting stimulus- or task-relevant components out of noisy, high-dimensional EEG data, as is done with other types of data used in music research [50].

The first single-trial EEG classification study focusing on musical stimuli was published in 2011 by Schaefer and colleagues [49]. They found that brain responses to seven short excerpts of naturalistic music [1] from a variety of genres could be classified significantly above chance. More recently, Stober et al. recorded EEG responses from East African listeners who heard twelve Western and twelve East African rhythms, and used deep-learning techniques to predict both the rhythm family of a stimulus (2-class problem) as well as the individual rhythm (24-class problem) from the EEG [52]. The prediction task of EEG classification has also extended beyond characterizing the stimuli to labeling listeners' emotional states—for example, in response to music videos [28] and musical excerpts [16].

[1] The term "naturalistic music" is used to refer to ecologically valid musical material as opposed to controlled, synthesized stimuli.

A brain-computer interface (BCI) is often cited as a general application of single-trial EEG classification [14]. In a musical context, a successful BCI would enable a user to communicate mentally by selectively interacting with an ongoing musical stimulus. Studies by Vlek and colleagues showed that subjective (mentally imposed) metrical accents on a beat sequence could be detected in the EEG response [60], and that a classifier trained upon responses to perceived accents could be used to detect the imagined accents [61]. In a recent EEG study by Treder et al. [58], also working toward BCI application, listeners were played polyphonic musical stimuli wherein each stream produced intermittent "oddball" musical events, and attended to just one of the streams. The authors leveraged the fact that the brain responds differently to attended oddball auditory stimuli than to unattended oddballs, and classified brain responses to just the oddball events in the music in order to identify the attended stream.

3.3 Tracking Temporal Dynamics of Acoustical Features

Certain music cognition studies have drawn explicitly from MIR techniques, utilizing acoustical features developed specifically for music analysis [59]. These studies use short-term (e.g., spectral flux, spectral centroid) and long-term (e.g., musical mode, pulse clarity) acoustical features, computationally extracted from musical stimuli, as a basis for quantitatively comparing stimuli with responses.

A 2010 behavioral study by Alluri and Toiviainen [1] set the foundation for this approach in the music cognition literature. The authors formulated perceptual scales suitable for assessing timbre of naturalistic music, and then linked human ratings of short musical excerpts to the excerpts' constituent short-term acoustical features. Subsequent fMRI studies used a refined set of short-term features, as well as long-term features, to characterize their musical stimuli. Alluri and colleagues identified brain regions whose fMRI time series correlated with those of the acoustical features of a tango piece [2], and later predicted brain activations from the features of a variety of musical excerpts [3]. A 2014 study by Toiviainen and colleagues took the inverse approach, predicting acoustical features from fMRI-recorded responses to Beatles songs [57].

Acoustical feature representation has also been studied in ongoing EEG. In contrast to relatively short epochs used in FFR and classification analysis, ongoing-EEG epochs can span many minutes, and are thus well suited to the analysis of responses to longer musical excerpts such as songs [17]. A 2013 study by Cong and colleagues used the same stimulus and long-term acoustical features as the 2012 Alluri study [2] in an ongoing-EEG paradigm, decomposing the EEG response into temporally independent sources using Independent Component Analysis (ICA), and then identifying sources whose frequency content corresponded to the time courses of the acoustical features [17]. More recently, Lin and colleagues also used EEG ICA sources to link ongoing-EEG responses to musical mode and tempo in shorter musical excerpts [36].

4. MIR APPLICATIONS

In the previous section, we reviewed three approaches used to study brain responses to music: The FFR, which directly encodes the pitch of an auditory stimulus, and two analysis techniques used for classifying and characterizing cortical responses. We will now discuss MIR applications of neuroimaging data. We consider the relevance of each approach to MIR research and assess the added value of analyzing the brain response—over analyzing, for example, the auditory stimulus directly.

4.1 Transcription

The FFR is unique among the auditory responses presented here in that it directly reflects the stimulus. As described above, the FFR has been used primarily as a measure of encoding. To date, its robustness has been the main attribute of interest, reflecting effects of expertise (tone-language speaker or musician) and stimulus properties (musical consonance or pleasantness) in the brain response.

The FFR could prove to be a powerful transcription tool; to our knowledge, this application has not yet been explored. From an MIR perspective, there would be little added value in transcribing responses to the simple musical stimuli used in the FFR studies described here (mostly monophonic, sometimes intervals or triads—see §3.1), as transcription could be easily accomplished directly from the audio. However, selective attention has been found to enhance FFR amplitudes for simultaneously presented speech stimuli [26, 35]; therefore, future research could study this topic further using musical stimuli, for example to extract a melody from polyphonic music—an open topic in audio MIR research, but something a human can accomplish effortlessly. Though FFRs to imagined sounds have yet to be confirmed, an FFR-based transcription system of this kind would certainly open another exciting and novel avenue for future research.

As described above, FFR studies typically involve up to thousands of stimulus repetitions due to low SNR. Therefore, signal-processing techniques that could efficiently extract the FFR out of the EEG—perhaps by recording the response from a montage of multiple electrodes, analogous to the use of multiple microphones in a source-separation scenario—would provide a useful resource for more flexible experiment design, and provide a critical step toward FFR-based transcription.

4.2 Tagging and Annotation

Characterizing musical attributes and listener responses is a recurring goal in MIR research, and has also been explored in EEG research [28, 37, 40]. In their 2010 paper, Alluri and Toiviainen [1] draw explicit connections between their proposed approach and the use of acoustical features in computational systems for music categorization. Along these lines, the acoustical feature following approach used in neuroimaging studies could extend beyond the prediction of the features from the brain response (as in [57]), toward a global prediction of musical genre from combinations of these features over time, as is done in audio-based genre classification.

Interestingly, a fine-grained temporal representation of acoustical musical features in the brain response has yet to be explored using noninvasive imaging techniques. While short-term acoustical features were used in the behavioral and fMRI studies discussed above (§3.3), they were averaged or downsampled to match the length of the behavioral stimuli (1.5 seconds) or the sampling rate of fMRI (0.45–0.5 Hz) [1–3, 57]. At the same time, the studies using EEG—arguably the best modality for investigating representation of short-term acoustical features—considered only long-term acoustical features in their analysis [17,36]. It may be the case, too, that neurally encoded features of music do not correspond exactly to the hand-crafted acoustical features discussed here; therefore, feature-learning approaches could also prove useful for connecting temporally resolved stimulus features to the brain response, whether to study feature processing and representation, or to develop an annotation tool.

Single-trial EEG classification could also be applied to this problem. Of the classification studies discussed here, only one used naturalistic music as stimuli [49]; the others used rhythmic patterns [52, 60, 61] or short events segmented from an ongoing stimulus [58]. One possibility for future MIR application could be to classify responses to a larger set of naturalistic musical excerpts to build, for example, a classification model that surpasses excerpt-level specificity and instead predicts genre, mood, or other global attributes from responses to new musical excerpts.

4.3 Predicting Large-Scale Audience Preferences

Brain responses can also be used to model listener preferences. This topic has been explored to some extent in the music neuroscience literature (e.g., [4]). However, to accomplish a widespread application of this goal—for example, in a neuromarketing setting [5]—would require that responses of the experimental sample generalize beyond that sample to a large-scale measure of success, such as sales of a product or ratings collected from the general public [8].

Recent studies have successfully used brain responses from a small sample to predict large-scale audience preferences. In a 2012 study, Berns and Moore collected fMRI responses and subjective ratings from participants who listened to a set of unknown songs. The authors then tracked the sales of the songs over the next three years and found a brain region whose activity correlated significantly with eventual song popularity [8]. Recent studies by Falk et al. [23] and Dmochowski et al. [21] showed that large-scale success of television commercials could be predicted from fMRI and EEG responses, respectively. In all three of these studies, the brain responses of the experimental sample correlated more strongly with large-scale measures of popularity and success than they did with self-reported preferences of that same sample. These findings lend credence to the theory that brain responses provide objective measures of preference, and that generalizations may be drawn from these responses with greater validity than sub-

Company	Product	Application	Website	Features
Emotiv	EPOC	commercial	http://emotiv.com/	fixed montage, wireless, iOS, Android
NeuroSky	MindWave, MindSet	commercial	http://neurosky.com/	fixed montage, wireless, iOS, Android
EGI	Avatar	research	http://avatareeg.com/	flexible montage and sensors, wireless, Android
Grass	Comet	research	http://www.grasstechnologies.com/	flexible montage and sensors
Neuroelectrics	StarStim	research	http://www.neuroelectrics.com/	flexible montage and sensors, wireless, stimulation

Table 2. Selected portable/mobile EEG systems.

jective ratings from a small experimental sample—even the very sample providing the brain responses. Therefore, MIR researchers may find brain-based measures of preference or success to be a useful channel of information in predicting or modeling large-scale music popularity.

4.4 Portable/Mobile EEG

While not an application per se, another area of growing interest in neuroscience involves portable and mobile EEG systems. It should be noted that nearly all of the studies reviewed here were conducted in controlled laboratory settings; thus, the listening experiences of the experimental participants likely did not reflect their experiences of music in everyday life. However, a number of commercial- and research-grade systems have come to market over the past decade (Table 2), and have recently begun to gain traction in the scientific literature as valid data-acquisition tools.

In an MIR context, a 2013 study by Morita et al. used the NeuroSky MindSet to assess mental states in response to music [40], and the 2014 study by Stober and colleagues (§3.2) used a portable Grass system for data collection [52]. Other recent scientific publications report real-time 3D imaging implementations using wireless EEG with a smartphone interface built using Emotiv equipment [44, 54], and a 2014 study by De Vos and colleagues showed that usable single-trial auditory responses could be recorded from a custom portable apparatus, also built off of the Emotiv system [18, 19]. The adoption of such methodologies by the scientific community presents an opportunity for MIR researchers to study music consumption and music processing in real-world listening situations [36].

5. DISCUSSION

In this review, we have surveyed neuroimaging techniques that can be used in MIR research, and highlighted a number of potential research topics spanning the two fields. Why, then, have collaborations not flourished to date?

One answer may emerge from a consideration of fundamental motivational differences between the two fields. Neuroscience, by definition, is the study of the brain; therefore, the thrust of much neuroscientific research is to gain an understanding of brain functioning underlying processing of various stimuli, including music. As a result, experiment design, data analysis, and interpretation of results will tend toward this goal, even when analysis involves decoding or prediction of stimulus or response features. A useful perspective on this topic is provided by Naselaris and colleagues [42], who characterize encoding versus decoding approaches used in fMRI research: Encoding approaches assess variations in neural space in response to variations in stimulus space, or perhaps seek to predict the brain response from the stimulus. Decoding, on the other hand, seeks to predict information about the stimulus from the brain response. In a neuroscientific setting, both approaches are used to map stimulus features to responses in order to better understand brain processing.

This objective is clearly evident in the studies reviewed above (§3). The FFR, providing arguably the most decodable brain signal, is used primarily to study neural encoding of auditory stimuli. One outcome of single-trial classification is the identification of temporal and spatial EEG components that best discriminate or differentiate stimuli or stimulus categories. The acoustical feature studies also focused upon identifying brain areas whose activity covaried with the stimuli, and not specifically on transcription. Of the approaches described above, perhaps only the BCI-focused EEG classification studies are purely application-based, with system performance taking priority over an exploration of the underlying neural processing—though an understanding of the latter is often a design consideration in the development of a high-performing BCI system.

MIR, on the other hand, tends to be a more application- and goal-oriented field [7]. For MIR researchers, then, brain data may serve more as a medium through which information about music may be recovered, than as the fundamental object of investigation. This disparity in what the brain, and brain data, represent in the overall goal of the research may be partly responsible for the lack of connection and collaboration between the two fields to date.

Another likely hinderance to the incorporation of neuroscientific techniques in MIR is access to data. Historically, researchers have had to acquire their own data, which requires access to equipment as well as domain-specific expertise in experiment design and data collection. Following that, data preprocessing and analysis can require significant signal-processing proficiency to extract stimulus-related information from noisy EEG recordings, especially for the single-trial and ongoing-EEG approaches discussed above. Luckily, the global scale of music neuroscience research now underway should provide many opportunities for collaboration, whereby MIR researchers may bypass some of the above steps if they wish. In addition, the creation of publicly available repositories of neuroimaging data has become a recent area of focus in the fMRI community [45, 46], and the EEG community is following suit (music-related EEG datasets include Koelstra et al.'s DEAP [27] and Stober et al.'s OpenMIIR [53]). Such public datasets, as well as open-source analysis packages such as EEGLAB [20], can facilitate cross-disciplinary research

even in the absence of formal collaborations.

While the fields of MIR and neuroscience have yet to form a strong connection, there exist many opportunites for collaboration that could advance both fields. It is hoped that the studies and ideas presented in this review will prove useful to both MIR researchers and neuroscientists. It is likely that the two fields will take some time to grow closer; therefore, MIR output using neuroscientific data may not immediately reach the neuroscientific audience (nor should it be intended to). Even so, we hope that a greater knowledge of neuroscientific approaches and findings will spark the interest of MIR researchers and lead to future intersections between these two exciting fields.

6. ACKNOWLEDGMENTS

This research is supported by the Wallenberg Network Initiative: Culture, Brain, Learning. The authors thank Jonathan Berger, Anthony Norcia, Jorge Herrera, and four anonymous ISMIR reviewers for their helpful suggestions and feedback on this paper.

7. REFERENCES

[1] V. Alluri and P. Toiviainen. Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception: An Interdisciplinary Journal*, 27(3):223–242, 2010.

[2] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677–3689, 2012.

[3] V. Alluri, P. Toiviainen, T. E. Lund, M. Wallentin, P. Vuust, A. K. Nandi, T. Ristaniemi, and E. Brattico. From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. *NeuroImage*, 83(0):627–636, 2013.

[4] E. Altenmüller, K. Schürmann, V. K. Lim, and D. Parlitz. Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, 40(13):2242–2256, 2002.

[5] D. Ariely and G. S. Berns. Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience*, 11(4):284–292, 2010.

[6] J. J. Aucouturier and E. Bigand. Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition. In *ISMIR*, pages 397–402, 2012.

[7] J. J. Aucouturier and E. Bigand. Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, 41(3):483–497, 2013.

[8] G. S. Berns and S. E. Moore. A neural predictor of cultural popularity. *Journal of Consumer Psychology*, 22(1):154–160, 2012.

[9] G. M. Bidelman. The role of the auditory brainstem in processing musically relevant pitch. *Frontiers in psychology*, 4:264, 2013.

[10] G. M. Bidelman, J. T. Gandour, and A. Krishnan. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of cognitive neuroscience*, 23(2):425–434, February 2011.

[11] G. M. Bidelman, J. T. Gandour, and A. Krishnan. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and cognition*, 77(1):1–10, October 2011.

[12] G. M. Bidelman and A. Krishnan. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience*, 29(42):13165–13171, 2009.

[13] G. M. Bidelman and A. Krishnan. Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport*, 22(5):212–216, March 2011.

[14] B. Blankertz, G. Curio, and K. R. Müller. Classifying single trial EEG: Towards brain computer interfacing. In *Advances in Neural Information Processing Systems*, pages 157–164, 2002.

[15] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, 56(2):814–825, 2011.

[16] R. Cabredo, R. S. Legaspi, P. S. Inventado, and M. Numao. An emotion model for music using brain waves. In *ISMIR*, pages 265–270, 2012.

[17] F. Cong, V. Alluri, A. K. Nandi, P. Toiviainen, R. Fa, B. Abu-Jamous, L. Gong, B. G. W. Craenen, H. Poikonen, M. Huotilainen, and T. Ristaniemi. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. *IEEE Transactions on Multimedia*, 15(5):1060–1069, August 2013.

[18] M. De Vos, K. Gandras, and S. Debener. Towards a truly mobile auditory braincomputer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, 91(1):46–53, 2014.

[19] S. Debener, F. Minow, R. Emkes, K. Gandras, and M. De Vos. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, 49(11):1617–1621, 2012.

[20] A. Delorme and S. Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[21] J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra. Audience preferences are predicted by temporal reliability of neural processing. *Nature communications*, 5:4567, 2014.

[22] J. S. Downie. Music information retrieval. *Annual Review of Information Science and Technology*, 37(1):295–340, 2003.

[23] E. B. Falk, E. T. Berkman, and M. D. Lieberman. From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychological Science*, 23(5):439–445, 2012.

[24] J. Futrelle and J. S. Downie. Interdisciplinary communities and research issues in music information retrieval. In *ISMIR*, pages 215–221, 2002.

[25] G. C. Galbraith, P. W. Arbagey, R. Branski, N. Comerci, and P. M. Rector. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, 6(17):2363–2367, November 1995.

[26] G. C. Galbraith, S. M. Bhuta, A. K. Choate, J. M. Kitahara, and T. A. Mullen. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport*, 9(8):1889–1893, June 1998.

[27] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, January 2012.

[28] S. Koelstra, A. Yazdani, M. Soleymani, C. Mhl, J. S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras. Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. In *Brain Informatics*, volume 6334 of *Lecture Notes in Computer Science*, pages 89–100. Springer Berlin Heidelberg, 2010.

[29] A. Krishnan, G. M. Bidelman, and J. T. Gandour. Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research*, 268(12):60–66, 2010.

[30] A. Krishnan, J. T. Gandour, and G. M. Bidelman. The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, 23(1):81–95, 2010.

[31] A. Krishnan, J. T. Gandour, G. M. Bidelman, and J. Swaminathan. Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*, 20(4):408–413, March 2009.

[32] A. Krishnan, Y. Xu, J. Gandour, and P. Cariani. Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1):161–168, 2005.

[33] J. H. Lee, M. C. Jones, and J. S. Downie. An analysis of ISMIR proceedings: Patterns of authorship, topic, and citation. In *ISMIR*, pages 58–62, 2009.

[34] K. M. Lee, E. Skoe, N. Kraus, and R. Ashley. Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience*, 29(18):5832–5840, 2009.

[35] A. Lehmann and M. Schönwiesner. Selective attention modulates human auditory brainstem responses: Relative contributions of frequency and spatial cues. *PloS one*, 9(1):e85442, 2014.

[36] Y. P. Lin, J. R. Duann, W. Feng, J. H. Chen, and T. P. Jung. Revealing spatio-spectral electroencephalographic dynamics of musical mode and tempo perception by independent component analysis. *Journal of NeuroEngineering and Rehabilitation*, 11(1), 2014.

[37] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, July 2010.

[38] P. Loui, D. Alsop, and G. Schlaug. Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience*, 29(33):10215–10220, 2009.

[39] R. K. Mehta and R. Parasuraman. Neuroergonomics: A review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, 7(889), 2013.

[40] Y. Morita, H. H. Huang, and K. Kawagoe. Towards music information retrieval driven by EEG signals: Architecture and preliminary experiments. In *IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, pages 213–217, June 2013.

[41] G. Musacchia, M. Sams, E. Skoe, and N. Kraus. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40):15894–15898, 2007.

[42] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant. Encoding and decoding in fMRI. *NeuroImage*, 56(2):400–410, 2011.

[43] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby. Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9):424–430, 2006.

[44] M. K. Petersen, C. Stahlhut, A. Stopczynski, J. E. Larsen, and L. K. Hansen. Smartphones get emotional: Mind reading images and reconstructing the neural sources. In *Affective Computing and Intelligent Interaction*, volume 6975 of *Lecture Notes in Computer Science*, pages 578–587. Springer Berlin Heidelberg, 2011.

[45] R. A. Poldrack, D. M. Barch, J. Mitchell, T. Wager, A. D. Wagner, J. T. Devlin, C. Cumba, O. Koyejo, and M. Milham. Toward open sharing of task-based fMRI data: The OpenfMRI project. *Frontiers in Neuroinformatics*, 7(12), 2013.

[46] R. A. Poldrack and K. J. Gorgolewski. Making big data open: Data sharing in neuroimaging. *Nature neuroscience*, 17(11):1510–1517, November 2014.

[47] C. Potes, P. Brunner, A. Gunduz, R. T. Knight, and G. Schalk. Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *NeuroImage*, 97:188–195, 2014.

[48] C. Potes, A. Gunduz, P. Brunner, and G. Schalk. Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. *NeuroImage*, 61(4):841–848, 2012.

[49] R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain. Name that tune: Decoding music from the listening brain. *NeuroImage*, 56(2):843–849, 2011.

[50] R. S. Schaefer, S. Furuya, L. M. Smith, B. B. Kaneshiro, and P. Toiviainen. Probing neural mechanisms of music perception, cognition, and performance using multivariate decoding. *Psychomusicology: Music, Mind, and Brain*, 22(2):168–174, 2012.

[51] E. Skoe and N. Kraus. Auditory brain stem response to complex sounds: A tutorial. *Ear and hearing*, 31(3):302–324, June 2010.

[52] S Stober, D. J. Cameron, and J. A. Grahn. Classifying EEG recordings of rhythm perception. In *ISMIR*, pages 649–654, 2014.

[53] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn. Towards music imagery information retrieval: Introducing the OpenMIIR dataset of EEG recordings from music perception and imagination. In *ISMIR*, 2015.

[54] A. Stopczynski, C. Stahlhut, J. E. Larsen, M. K. Petersen, and L. K. Hansen. The smartphone brain scanner: A portable real-time neuroimaging system. *PloS one*, 9(2):e86733, 2014.

[55] I. Sturm, B. Blankertz, C. Potes, G. Schalk, and G. Curio. ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Frontiers in Human Neuroscience*, 8(798), 2014.

[56] J. Swaminathan, A. Krishnan, and J. T. Gandour. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, 19(11):1163–1167, July 2008.

[57] P. Toiviainen, V. Alluri, E. Brattico, M. Wallentin, and P. Vuust. Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage*, 88(0):170–180, 2014.

[58] M. S. Treder, H. Purwins, D. Miklody, I. Sturm, and B. Blankertz. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *Journal of neural engineering*, 11(2):026009, April 2014.

[59] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.

[60] R. J. Vlek, R. S. Schaefer, C. C. A. M. Gielen, J. D. R. Farquhar, and P. Desain. Sequenced subjective accents for brain-computer interfaces. *Journal of neural engineering*, 8(3):036002, June 2011.

[61] R. J. Vlek, R. S. Schaefer, C. C. A. M. Gielen, J. D. R. Farquhar, and P. Desain. Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology*, 122(8):1526–1532, 2011.

[62] P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, and N. Kraus. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4):420–422, April 2007.