Psychomusicology: Music, Mind, and Brain
2012, Vol. 22, No. 2, 168–174

0275-3987/12/$12.00 DOI: 10.1037/a0031014

Probing Neural Mechanisms of Music Perception, Cognition, and Performance Using Multivariate Decoding[1]

Rebecca S. Schaefer
University of Edinburgh

Shinichi Furuya
Hannover University of Music, Drama, and Media

Leigh M. Smith
iZotope, Inc., Cambridge, Massachusetts

Blair Bohannan Kaneshiro
Stanford University

Petri Toiviainen
University of Jyväskylä

Recent neuroscience research has shown increasing use of multivariate decoding methods and machine learning. These methods, by uncovering the source and nature of informative variance in large data sets, invert the classical direction of inference that attempts to explain brain activity from mental state variables or stimulus features. However, these techniques are not yet commonly used among music researchers. In this position article, we introduce some key features of machine learning methods and review their use in the field of cognitive and behavioral neuroscience of music. We argue for the great potential of these methods in decoding multiple data types, specifically audio waveforms, electroencephalography, functional MRI, and motion capture data. By finding the most informative aspects of stimulus and performance data, hypotheses can be generated pertaining to how the brain processes incoming musical information and generates behavioral output, respectively. Importantly, these methods are also applicable to different neural and physiological data types such as magnetoencephalography, near-infrared spectroscopy, positron emission tomography, and electromyography.

Keywords: machine learning, classification, music neuroscience

Supplemental materials: http://dx.doi.org/10.1037/a0031014.supp

Music is acoustic information with complex temporal and spatial features. Research into perception and cognition of multifaceted aspects of music aims to decode the information from neural signals elicited by listening to music. Music performance, on the other hand, entails the encoding of musical information to neural commands issued to the muscles. To understand the neural processes underlying music perception, cognition, and performance, therefore, researchers face issues of extracting meaningful information from extremely large data sets with regard to neural, physiological, and biomechanical signals. This is nontrivial in light of recent technological advances in data collection, which can lead to a potentially overwhelming amount of data. The supervised and unsupervised methods of machine learning are powerful tools for uncovering unseen patterns in these large data sets. In this way, not only can the means of specified conditions be compared, but data-driven methods are used to uncover sources of informative variance in the signals. Moreover, machine learning allows for quantitative evaluation of individual differences in music perception and performance.

In this article, we introduce key features of machine learning and highlight some examples of their use on a range of data types. After reviewing basic concepts and terminology, we discuss dimensionality reduction and the impact of the choice of algorithm. We then turn to data types we judge to be most relevant to the neural processing of music. For an audio waveform, it is possible to elucidate the most perceptually informative part of the signal, by determining which aspects of the signal are most salient or useful to the brain in determining specific characteristics of the sound. In the same way, it is possible to uncover neural representations of musical attributes such as rhythm and harmony in a data-driven way by applying supervised and unsupervised learning to single-

Rebecca S. Schaefer, Institute for Music in Human Social Development, University of Edinburgh, Edinburgh, United Kingdom; Shinichi Furuya, Institute for Music Physiology and Musicians' Medicine, Hannover University of Music, Drama, and Media, Hannover, Germany; Leigh M. Smith, iZotope, Inc., Cambridge, Massachusetts; Blair Bohannan Kaneshiro, Center for Computer Research in Music and Acoustics, Stanford University; Petri Toiviainen, Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland.

Photos and brief biographies of each author are available at http://dx.doi.org/10.1037/a0031014.supp

Correspondence concerning this article should be addressed to Rebecca S. Schaefer, Reid School of Music, Alison House, 12 Nicolson Square, Edinburgh EH8 9DF, UK. E-mail: r.schaefer@ed.ac.uk

[1] This article is based on a symposium held at the joint ESCOM8/ICMPC12 meeting in Thessaloniki, Greece, 2012.

trial electroencephalography (EEG) or functional MRI (fMRI) data. Finally, machine learning methods are also useful in behavioral research, allowing characterization of fundamental patterns of movements that use a large number of joints and muscles during musical performance.

Methods

Although we do not aim to give a complete overview of these methods here (for a detailed review of machine learning methods for brain imaging, see Lemm, Blankertz, Dickhaus, & Müller, 2011), we introduce a number of key aspects of machine learning as they pertain to music cognition research.

Basic Terminology

Machine learning (or *statistical learning*) involves uncovering meaningful patterns in collections of observations, often with the goal of *classifying*, or categorizing, the observations in some way. *Observations* are instances of data—for example, audio excerpts, EEG or fMRI responses, or motion capture data. Each observation is described by qualitative or quantitative descriptors that make up its *feature vector*. *Labels* (or *classes*) specify the stimulus, task, or state, depending on the focus of the research. Machine learning can be divided into supervised and unsupervised learning tasks. In *supervised learning*, a model is fitted to labeled observations, and then used to predict labels of new observations. The *classification rate*, also known as *classifier accuracy*, or *F-measure*, is the percentage of correctly labeled observations, often separated into *recall*, the percentage correctly classified, and *precision*, the percentage of those correctly classified divided by the number of classifications predicted (Van Rijsbergen, 1979). Supervised learning is predictive—attempting to correctly label a future event—in contrast to descriptive tasks, such as averaging-based analyses, where the goal is to summarize or characterize a set of observations whose labels are already known. Labels of observations for supervised learning are known a priori; in contrast, labels of observations for *unsupervised learning* are not known in advance. Here, observations are clustered, and the clusters are then used to define labels a posteriori.

Classification is a core function of machine learning (Hastie, Tibshirani, & Friedman, 2009) and data mining (Witten & Frank, 2005). Supervised learning tasks use classifier training and test sets. The *training set* is the collection of labeled observations used to build the model, and the *test set* is the collection of unlabeled observations (or observations whose labels are withheld) on which the model is tested. Although the classification rate is the most commonly reported metric of classifier performance, further insight can be gained by analyzing the *confusion matrix* (or *conditional probability matrix*), which shows how many observations having label *i* (rows) were given label *j* (columns). As a particular classification rate could result from a variety of confusion matrices (as shown in Figure 1), confusion matrices elaborate on accuracy by giving an indication of distance between the responses. These distance measures can then be compared with measures of distance between the stimuli, tasks, or states that produced them. Responses that are more distinct from one another will classify with greater accuracy, whereas responses that are similar will have a greater tendency to be confused with one another by the classifier.

Dimensionality Reduction

A concern in classification is the risk of misclassifying unseen data due to overfitting a model to training data in a high-dimensional feature space, thereby basing subsequent classification on irrelevant noise or measurement outliers. This phenomenon is referred to as the *curse of dimensionality* (Tan, Steinbach, & Kumar, 2006). Following preprocessing, therefore, a data set is typically subjected to *dimensionality reduction* (or *data reduction*), which uses either unsupervised or supervised approaches to extract a small number of features by removing irrelevant, redundant, and noisy information. Unsupervised approaches include principal component analysis (PCA), independent component analysis (ICA), non-negative matrix factorization (NMF), and factor analysis (FA). PCA and ICA are the most commonly used dimensionality reduction methods, which create new variables (i.e., compo-

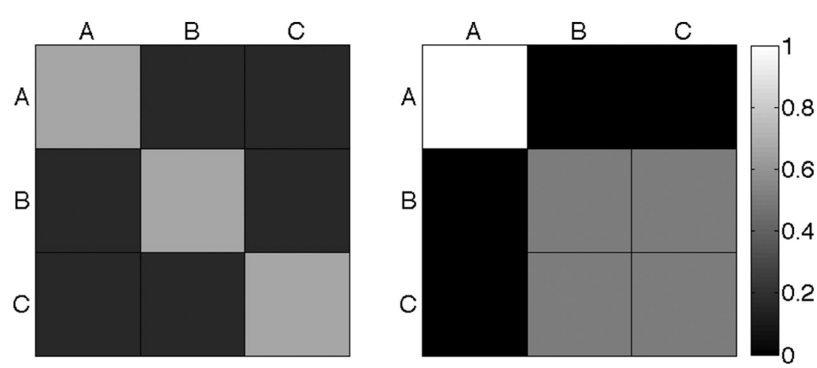

Figure 1. Two confusion matrices with identical accuracy but different performance for a three-class task. Both matrices correspond to accuracies of 66.7%. Left: The monochromatic diagonal and monochromatic off-diagonal indicate that the classifier correctly labeled observations from each category with equal accuracy, and additionally mislabeled observations with equal confusion between the other two classes. Right: Observations from category A were always correctly labeled, while the classifier was unable to distinguish well between categories B and C.

nents) according to criteria of maximizing variance explained or maximizing statistical independence between variables, respectively. Each component can be expressed as a linear combination of original variables. A small set of the derived components often accounts for a large portion of variance of the original variables; thus, these techniques allow for projecting a data set into a low-dimensional subspace. Note that ICA can also be used for identification and removal of spontaneous noise and signals unrelated to brain signals such as eye movements and cardiac activity. By contrast, when the class (or label) information is available, supervised approaches, such as linear discriminant analysis (LDA), are usually more effective than unsupervised ones such as PCA and ICA for dimensionality reduction to allow better performance of the subsequent classification. LDA computes an optimal projection to a lower-dimensionality hyperspace by simultaneously minimizing within-class variance and maximizing between-class variance, thus achieving maximum class discrimination. This therefore not only decreases dimensionality of a data set, but also increases accuracy in subsequent classification tasks.

Classifiers

The various algorithms proposed for classification differ in their *inductive bias* (intrinsic learning capability) and therefore suitability to a particular classification task or problem domain (Alpaydin, 2009). Classification algorithms include *probabilistic* (or *generative*) models (e.g., Bayesian networks) that make classifications based on joint probability estimations of label likelihood given evidence (test data); and *discriminative* models (e.g., LDA, support vector machines) that directly compute multidimensional hyperplanes between classes given the feature sets of training data. Historical variants include decision trees and artificial neural networks, which can offer nonlinear discriminants; however, the justification for increases in complexity and training iterations often depends on the problem domain. Examples of unsupervised classifiers include k-means clustering, self-organizing maps (Alpaydin, 2009) and projection into spaces while enforcing a sparsity criterion such as NMF (for more detailed information, see Hastie et al., 2009).

In many classification algorithms, feature data can be *discrete* or *continuous*, whereas some methods require normalization across features before classification, typically with respect to the mean feature values of the training data. In addition, classifier output can be discrete (choosing among a set number of classes) or continuous (choosing a point on a continuum—sometimes referred to as a *regression* problem). A single resulting label is produced by *hard classification*, choosing from the label taxonomy; in contrast, *soft classification* may produce several candidate labels for a single instance of test data. In both cases, accompanying these chosen labels may be a likelihood measure of each label's accuracy, which can be directly computed with probabilistic models or estimated by the distance from the discriminating hyperplane. *Cross-validation* is one commonly used technique to better estimate prediction error in classification (Hastie et al., 2009). For an S-fold cross-validation task, data are first randomized and partitioned into S subsets. A total of S classifications are then performed. For fold i, subset Si will be used as the test (validation) set, whereas the remaining $S - 1$ subsets will be used as the training set to build the model for that fold. The reported rate is the mean rate across the S classifications.

Applications

Although classification methods can be used for a range of data, we will briefly review four data types that we believe are most relevant to investigating the neural mechanisms of music processing.

Audio

Human listeners demonstrate remarkable skill in identifying specific characteristics (such as timbre, genre, or other qualities) of individual sounds with little data (e.g., short duration of audio stimuli, see for instance Krumhansl, 2010). This demands efficient coding of the spectro-temporal behavior of sound for classification, identification, and interpretation. Machine classification of audio signals—often termed *auto-tagging*—has been applied to simulate human listener capabilities such as identifying musical instrument (Smith, Pope, Leboeuf, & Tjoa, 2012), musical genre, mood (Knox, Mitchell, Beveridge, & MacDonald, 2011) and vocal gender identification, identifying performers from performances (Saunders, Hardoon, Shawe-Taylor, & Widmer, 2008; Stamatatos & Widmer, 2005), and auditory scene detection. Such labels are suitable for applications in automating content interpretation and retrieval, mixing, and other signal processing. Additionally, by identifying the source of meaningful variance in specific aspects of audio signals (which is not possible with univariate methods), we may infer what information the brain uses in perceptual processing.

Spectral and temporal audio features that are commonly used for classification include spectral behavior of the waveform over short time windows using perceptually informed critical bands such as bark or Mel frequency scales (Zwicker & Fastl, 1999) to calculate a spectral envelope (Logan, 2000). A related set of features are *spectral moments* which are analogous to statistical moments (Herrera-Boyer, Klapuri, & Davy, 2006). Low-level temporal features, such as attack slope and temporal centroid (Peeters, 2004), are often computed along with musically specific features such as rhythm patterns, pitch chromagrams, and bass-line pitch detection (Pampalk, Rauber, & Merkl, 2002; Müller, Ellis, Klapuri, & Richard, 2011). These features are thought to contribute to label distinction; however, the exact contribution of each feature to that classification decision is determined from supervised learning against training data sets. Machine learning techniques such as support vector machines (Burred & Peeters, 2009) and Adaboost (Hastie et al., 2009) have been applied to produce classification labels given a selected taxonomy of labels. In contrast, a recent viable alternative to this approach using unsupervised learning from high-dimensional sparse feature representations instead of predetermined features is presented by Nam, Herrera, Slaney, & Smith (2012). The performance of machine learning methods in classification of audio demonstrates the relative contribution of bottom-up, signal-derived features and data-oriented classification processes to human cognition. Such demonstrations then sharpen the delineation of the contribution of top–down, expectation-based processes in human auditory cognition.

EEG

EEG measures electrical signals of the brain. This recording modality yields high-resolution time courses for multiple (often highly correlated) channels spread over the scalp, leading to a high number of data points for a limited number of electrodes. Single-trial analyses are quite commonly used with EEG (or magnetoencephalography, MEG) data in the field of brain-computer interfacing (BCI), where covert mental actions are decoded in real-time to be used as symbolic communication or to drive a device (van Gerven et al., 2009). Traditional methods do not allow for this, as they need many trials to create averages that cancel out irrelevant signals. The most common tasks investigated in this field are movement imagery (e.g., Pfurtscheller, Brunner, Schlögl, & Lopes da Silva, 2006) and selective visual attention (e.g., Fazel-Rezai et al., 2012); however, other tasks such as tactile attention, visual imagery, mental navigation, and mental arithmetic are also used. By using subjective rhythm, or effortfully imagining metrically patterned accents on a metronome stimulus of identical sound events, a more musical task has also shown promise in the BCI domain (Vlek, Schaefer, Gielen, Farquhar, & Desain, 2011a). Only recently, however, has classification been applied to EEG with the goal of investigating music cognition. In this way, the neural representations of timbre (Bohannan, Terasawa, Arnardottir, Perreau Guimaraes, & Suppes, 2010), tonal expectation (Kaneshiro, Berger, Perreau Guimaraes, & Suppes, 2012), chord changes (Sturm, Curio, & Blankertz, 2010), musical emotion (Lin, Wang, Wu, Jeng, & Chen, 2007, 2010), mechanisms of subjective accenting (Vlek, Schaefer, Gielen, Farquhar, & Desain, 2011b), imagined rhythmic patterns (Desain, 2004), and imagined and perceived natural music (Schaefer, Perreau Guimaraes, Desain, & Suppes, 2008, Schaefer, Farquhar, & Desain, 2011) have been investigated. In most cases, an informative feature is identified through data reduction techniques, and classification is performed on only that feature. This procedure thus includes an initial phase in which a hypothesis is formulated (feature selection), and then tested (classification). In picking features to use for classification, previous work has included the time course (or event-related single trial) as well as the power in frequency bands. As described earlier, not only do these methods allow estimation of the distance between classes in terms of neural activation, but the widespread interindividual differences in music processing can also be investigated.

fMRI

Functional MRI (fMRI) allows for spatially accurate (~1 mm) localization of brain activity. The temporal resolution, however, is inferior (~1 s) to that obtained with other brain imaging methods, such as EEG or MEG. Consequently, fMRI data typically comprise a high number (~10^5) of voxel time series, each containing a relatively low number (~10^2) of data points. Therefore, in contrast to EEG and MEG data, which typically contain a large number of data points in a small number of channels, the features-to-observations ratio in fMRI is extremely high. This calls for careful measures, such as dimensionality reduction, regularization, and cross-validation, to avoid overfitting (Hansen et al., 1999; Cox & Savoy, 2003; Lemm et al., 2011). A further characteristic of fMRI data is that it does not directly measure the neural response to a stimulus, but rather provides an indirect measure by measuring the change in blood flow (blood-oxygen-level-dependent contrast; BOLD). The respective *hemodynamic* delay is of the order of 5 s and must be taken into account in the classification task. Machine learning offers an alternative to statistical methods that are vulnerable to multiple comparison problems, which are particularly problematic for data sets of this size.

Although fMRI data have been subjected to various classification tasks in the visual domain (Cox & Savoy, 2003; Thirion et al., 2006), in the domain of music such studies are, as yet, almost nonexistent (with some notable exceptions, such as Abrams et al. (2011), who located brain areas specific to music and speech processing, and Lee, Janata, Frost, Hanke, & Granger (2011), who investigated brain areas related to upward or downward melody contours). In a more naturalistic setting, Toiviainen et al. (2012) performed classification of fMRI data obtained from participants listening to natural music stimuli, and obtained a classification accuracy rate that was significantly above chance level. To provide an example, Figure 2 displays the first eigenvector obtained from LDA on fMRI data evoked by excerpts of baroque and bebop music. This corresponds to the direction in the feature space along which there is maximal separation between the two classes. Before classification, PCA was used to reduce the dimensionality of the fMRI data to 10, which was found to yield maximal classification performance in a four-category classification task (blues, baroque, bebop, pop) with leave-one-out cross-validation. As can be seen, listening to baroque was associated with increased activation in parts of bilateral middle temporal gyrus, medial superior frontal gyrus, and middle and posterior bilateral cingulate gyrus (blue areas), while listening to bebop resulted in increased activation in bilateral superior temporal lobe, right rolandic operculum, right thalamus, and right precentral gyrus (red areas, for a color version of this figure, please see the Supplementary Materials).

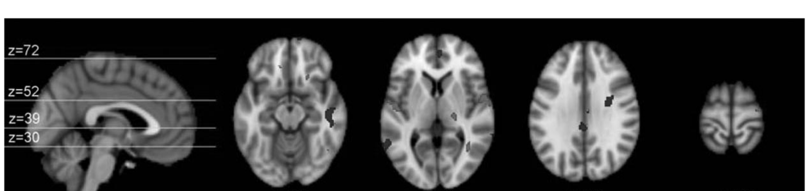

Figure 2. Areas contributing maximally ($p < .01$) to the Linear Discriminant Function between fMRI responses to baroque and bebop music. Blue and red indicate areas with high activation during listening to baroque and bebop. A color version of this figure is available as supplemental material at: http://dx.doi.org/10.1037/a0031014.supp.

Motion Capture

Body movements in musical performance provide invaluable information for understanding neural mechanisms underlying a variety of skillful movement productions. Recordings require many sensors to reflect the large number of joints/muscles in our motor system. Movement data collected with a motion capture system, data glove, or electromyography therefore have an inherently high dimensionality. Multivariate and cluster analyses allow for identifying a small set of fundamental movement patterns obscured in these vast data sets—key features characterizing the skilled motor behavior of musicians.

A key issue in music performance is to characterize the rich repertoire of movements needed to play a variety of music ranging from Johan Sebastian Bach to Gyorgy Ligeti. One possible approach to this question is to find a small number of fundamental movement patterns that are common across various tone sequences. However, this exploratory analysis is difficult to perform with conventional hypothesis-driven statistics. Instead, a combined use of PCA and unsupervised cluster analysis shows better performance. A data set that consists of time-varying joint kinematics during the playing of a number of melodies yields a small set of distinct patterns of finger joint coordination that are common across the sequences, describing a dimensionality reduction of complex motor behaviors in music performance (Furuya, Flanders, & Soechting, 2011; Furuya & Soechting, 2012). PCA on a data set of finger joint kinematics elicited by transcranial magnetic stimulation over the primary motor cortex can also allow for a distinct set of joint covariation patterns, which indeed differed between pianists and nonmusicians (Gentner et al., 2010).

Another issue is individual differences in the movement organization across players in music performance (Dalla Bella & Palmer, 2011; Furuya, Aoki, Nakahara, & Kinoshita, 2012). For example, an infinite number of ways of changing the upper-limb movements exist for adjusting acoustic features of music such as loudness and tempo. To classify the interindividual differences across players, multiple regression and cluster analysis provide useful information (Furuya et al., 2012). Multiple regression can identify a coefficient that represents a change in the kinematics of each joint in relation to the acoustic variable for each player. An unsupervised cluster analysis classifies a feature vector that consists of the derived coefficient at multiple joints at the upper-extremity across all players. These analyses can group players according to similarity of variation of the interjoint coordination in relation to the acoustic variable. A possible medical application can be found in a study that examined pianists (Furuya et al., 2012), where groups of players were identified, between whom the muscular load during playing differed substantially, which indicates that specific muscles with high risk of playing-related injury can be identified based on movement kinematics of individual players. This emphasizes the pedagogical and clinical importance of addressing individual differences in body movements by using classification techniques.

Conclusion

We have introduced some key concepts of machine learning methods and their application to different types of data. We assert that machine learning provides valuable tools for investigating perception, cognition, as well as production of music. Not only does it provide a solution to dealing with large data sets, but it also provides a method of localizing informative content in a data-driven way, thereby creating a hypothesis and testing it within the same data set (but on unseen data). This informative content can be used in different ways, either to determine the distance between classes, but also to assess commonality between classes or even people, for instance by training a classifier on one data set and testing it on a different data set, thus uncovering unseen patterns shared by the two. Examples include classifying between tasks, as Vlek et al. (2011b) did by training a classifier on heard stimuli and using this classifier for imagined stimuli; or classifying between subjects, as Schaefer et al. (2011) demonstrated for common representations between listeners at the single-trial level for heard musical fragments. Additionally, machine learning methods have recently been suggested to have great potential for investigating social interaction in two-brain data sets (Konvalinka & Roepstorff, 2012).

It is important to consider that it is often not easy to interpret the patterns found by machine learning methods in a neural context. Classification can be—and often is—used as a blind method delivering a specific source of information that is relevant to the categories that are investigated. Often weight patterns are shown as research results while they may mean not much without further interpretation. Alternatively, some aspect of the experiment design, data collection or analyses may produce a category-irrelevant effect which for some reason varies with the different classes. It should also be noted that classifiers differ in their interpretability, as by examining the trained classifier it is difficult to infer how it has ended up with the particular classification. For instance, LDA is more explicit (the canonical discriminant functions indicate the directions in the hyperspace along which the class separation is maximal), whereas for example, the support vectors produced by SVM do not offer easy interpretation, and hence the learning is much more implicit.

In spite of the strong potential of machine learning techniques for the analysis of large data sets, decreasing the number of variables during experimental design and feature selection is still highly recommended in order to obtain better classification results. A selection of simpler classifiers such as nearest-neighbor algorithms may also help to prevent overfitting. The use of correctly implemented cross-validation is also instrumental in judging the extent of overfitting in specific folds or data partitions. Regularization approaches (e.g., ridge regression, LASSO, elastic net) can help to avoid overfitting even without decreasing dimensionality.

As these methods gain interest in the field of cognitive neuroscience (see a recent special journal issue on the topic edited by Haynes, 2011), resources are becoming increasingly available. Researchers can readily perform the described analyses using commercial software (e.g., MATLAB) and open-source software (e.g., R, Weka, Octave, EEGLAB, LIBSVM). To learn more about classification techniques, several online courses are available (e.g., Stanford Online Machine Learning Course). These materials provide music researchers with tools to extract meaningful information from the increasingly large data sets currently being acquired.

Although only four data types were discussed here, these methods are potentially useful in analyzing a range of other neural and physiological measurements such as MEG, near-infrared spectroscopy (NIRS), positron emission tomography (PET), and electromyography (EMG), or even non physiological music cognition data (such as the rich song metadata currently available through the Internet (see for instance Wu, Lin, Chen, & Jeng, 2008; Lin, Yang, & Chen, 2011). For all these data types, machine learning methods offer a way to identify informative content from large data sets in many different contexts, and in ways that are often neglected in more conventional analyses.

References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex, 21,* 1507–1518. doi:10.1093/cercor/bhq198

Alpaydin, E. (2009). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press.

Bohannan, B., Terasawa, H., Arnardottir, S., Perreau Guimaraes, M., & Suppes, P. (2010). Perceptual and neural representation of the temporal attribute of timbre. In: S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition* (p. 491), Seattle, WA.

Burred, J. J., & Peeters, G. (2009). An adaptive system for music classification and tagging. In: *Proceedings of the International Workshop on Learning the Semantics of Audio Signals (LSAS),* Graz, Austria.

Cox, D. D., & Savoy, R. (2003). Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage, 19,* 261–270. doi:10.1016/S1053-8119(03)00049-1

Dalla Bella, S., & Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS One, 6, e20518.* doi:10.1371/journal.pone.0020518

Desain, P. (2004). What rhythm do I have in mind? Detection of imagined temporal patterns from single trial ERP. In: S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition* (p. 209). Evanston, IL, Adelaide, Australia: Causal Productions.

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: Current challenges and emerging trends. *Frontiers in Neuroengineering, 5, 14*. doi:10.3389/fneng.2012.00014

Furuya, S., Aoki, T., Nakahara, H., & Kinoshita, H. (2012). Individual differences in the biomechanical effect of loudness and tempo on upper-limb movements during repetitive piano keystrokes. *Human Movement Science, 31,* 26–39. doi:10.1016/j.humov.2011.01.002

Furuya, S., Flanders, M., & Soechting, J. (2011). Hand kinematics of piano playing. *Journal of Neurophysiology, 106,* 2849–2864. doi:10.1152/jn.00378.2011

Furuya, S., & Soechting, J. (2012). Speed invariance of independent control of finger movements in pianists. *Journal of Neurophysiology, 108,* 2060–2068. doi:10.1152/jn.00378.2012

Gentner, R., Gorges, S., Weise, D., aufm Kampe, K., Buttmann, M., & Classen, J. (2010). Encoding of motor skill in the corticomuscular system of musicians. *Current Biology, 20,* 1869–1874. doi:10.1016/j.cub.2010.09.045

Hansen, L. K., Larsen, J., Nielsen, F., Strother, S., Rostrup, E., Savoy, R., . . . Paulson, O. (1999). Generalizable patterns in neuroimaging: How many principal components? *Neuroimage, 9,* 534–544. doi:10.1006/nimg.1998.0425

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York, NY: Springer Verlag. doi:10.1007/978-0-387-84858-7

Haynes, J. D. (2011). Editorial: Multivariate decoding and brain reading: Introduction to the special issue. *Neuroimage, 56,* 385–386. doi:10.1016/j.neuroimage.2011.03.057

Herrera-Boyer, P., Klapuri, A., & Davy, M. (2006). Automatic classification of pitched musical instrument sounds. In: A. Klapuri & M. Davy (Eds.), *Signal processing methods for music transcription*. New York, NY: Springer Verlag. doi:10.1007/0-387-32845-9_6

Kaneshiro, B., Berger, J., Perreau Guimaraes, M., & Suppes, P. (2012). An exploration of tonal expectation using single-trial EEG classification. In: E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiades (Eds.), *Proceedings of the 12th International Conference on Music Perception & Cognition* (pp. 509–515). Thessaloniki, Greece.

Knox, D., Mitchell, L., Beveridge, S., & MacDonald, R. (2011). Acoustic analysis and mood classification of pain-relieving music. *Journal of the Acoustical Society of America, 130,* 1673–1682. doi:10.1121/1.3621029

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience, 6, 215*. doi:10.3389/fnhum.2012.00215

Krumhansl, C. L. (2010). Plink: "Thin Slices" of music. *Music Perception, 27,* 337–354. doi:10.1525/mp.2010.27.5.337

Lee, Y. S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage, 57,* 293–300. doi:10.1016/j.neuroimage.2011.02.006

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage, 56,* 387–399. doi:10.1016/j.neuroimage.2010.11.004

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., & Chen, J.-H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering, 57,* 1798–1806. doi:10.1109/TBME.2010.2048568

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., & Chen, J.-H. (2007). Multilayer perceptron for EEG signal classification during listening to emotional music. In: *Proceedings of the IEEE International Region 10 Conference (TENCON′07)* (pp. 1–3). Taipei, Taiwan.

Lin, Y.-C., Yang, Y.-H., & Chen, H. H. (2011). Exploiting online music tags for music emotion classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 7S,* Article 26.

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In: *Proceedings of the ISMIR International Symposium on Music Information Retrieval*. Plymouth, MA.

Müller, M., Ellis, D. P. W., Klapuri, A., & Richard, G. (2011). Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing, 5,* 1088–1110. doi:10.1109/JSTSP.2011.2112333

Nam, J., Herrera, J., Slaney, M., & Smith, J. (2012). Learning sparse feature representations for music annotation and retrieval. In: *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*. Porto, Portugal.

Pampalk, E., Rauber, A., & Merkl, D. (2002). Content-based organization and visualization of music archives. In: L. A. Rowe, B. Mérialdo, M. Mühlhäuser, K. W. Ross, & N. Dimitrova (Eds.), *Proceedings of the 10th ACM International Conference on Multimedia*. Juan les Pins, France.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the cuidado project*. Technical report. Paris, France: IRCAM.

Pfurtscheller, G., Brunner, C., Schlögl, A., & Lopes da Silva, H. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks, *Neuroimage, 31,* 153–159. doi:10.1016/j.neuroimage.2005.12.003

Saunders, C., Hardoon, D., Shawe-Taylor, J., & Widmer, G. (2008). Using string kernels to identify famous performers from their playing style. *Intelligent Data Analysis, 12,* 425–440.

Schaefer, R. S., Farquhar, J., & Desain, P. (2011). Name that tune: Decoding music from the listening brain. *Neuroimage, 56,* 843–849. doi:10.1016/j.neuroimage.2010.05.084

Schaefer, R. S., Perreau Guimaraes, M., Desain, P., & Suppes, P. (2008). Detecting imagined music from EEG. In: K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition*. Sapporo, Japan.

Smith, L. M., Pope, S. T., Leboeuf, J., & Tjoa, S. (2012). Automated classification of music genre, sound objects, and speech by machine learning. In: E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiades (Eds.), *Proceedings of the 12th International Conference on Music Perception & Cognition* (p. 943). Thessaloniki, Greece.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence, 165,* 37–56. doi:10.1016/j.artint.2005.01.007

Sturm, I., Curio, G., & Blankertz, B. (2010). Single-trial ERP analysis reveals unconsciously perceived structures of music. In: *Proceedings of the TOBI workshop*. Graz, Austria.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston, MA: Pearson Education.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *Neuroimage, 33,* 1104–1116. doi:10.1016/j.neuroimage.2006.06.062

Toiviainen, P., Alluri, V., Brattico, E., Nielsen, A. H., Dohn, A., Wallentin, M., & Vuust, P. (2012). I can read your mind: Inverse inference in musical neuroinformatics. In: E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiades (Eds.), *Proceedings of the 12th International Conference on Music Perception & Cognition*. Thessaloniki, Greece.

van Gerven, M., Farquhar, J., Schaefer, R. S., Vlek, R., Geuze, J., Nijholt, A., . . . Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering, 6, 041001*. doi:10.1088/1741-2560/6/4/041001

Van Rijsbergen, C. V. (1979). *Information retrieval* (2nd ed.). London, Boston, MA: Butterworth.

Vlek, R. J., Schaefer, R. S., Gielen, C. C., Farquhar, J. D. R., & Desain, P. (2011a). Sequenced subjective accents for brain–computer interfaces. *Journal of Neural Engineering, 8, 036002*. doi:10.1088/1741-2560/8/3/036002

Vlek, R. J., Schaefer, R. S., Gielen, C. C., Farquhar, J. D., & Desain, P. (2011b). Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology, 122,* 1526–1532. doi:10.1016/j.clinph.2011.01.042

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.

Wu, T.-L., Lin, Y.-P., Chen, J.-H., & Jeng, S.-K. (2008). Movie themes playlist generation through gaps across emotion loci and curve of GAEL versus preference. In: K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition* (p. 655). Sapporo, Japan.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). New York: Springer-Verlag.

Received August 23, 2012
Revision received October 18, 2012
Accepted October 31, 2012 ■