

First Demonstration of a Musical Emotion BCI

Scott Makeig, Grace Leslie, Tim Mullen, Devpratim Sarma,
Nima Bigdely-Shamlo, and Christian Kothe

Swartz Center for Computational Neuroscience, Institute for Neural Computation,
University of California San Diego, 9500 Gilman Drive, La Jolla, California, USA 92093-0559
smakeig@ucsd.edu

Abstract. Development of EEG-based brain computer interface (BCI) methods has largely focused on creating a communication channel for subjects with intact cognition but profound loss of motor control from stroke or neurodegenerative disease that allows such subjects to communicate by spelling out words on a personal computer. However, other important human communication channels may also be limited or unavailable for handicapped subjects -- direct non-linguistic emotional communication by gesture, vocal prosody, facial expression, etc.. We report and examine a first demonstration of a musical 'emotion BCI' in which, as one element of a live musical performance, an able-bodied subject successfully engaged the electronic delivery of an ordered sequence of five music two-tone bass frequency drone sounds by imaginatively re-experiencing the human feeling he had spontaneously associated with the sound of each drone sound during training sessions. The EEG data included activities of both brain and non-brain sources (scalp muscles, eye movements). Common Spatial Pattern classification gave 84% correct pseudo-online performance and 5-of-5 correct classification in live performance. Re-analysis of the training session data including only the brain EEG sources found by multiple-mixture Amica ICA decomposition achieved five-class classification accuracy of 59-70%, confirming that different voluntary emotion imagination experiences may be associated with distinguishable brain source EEG dynamics.

Keywords: BCI, affective computing, emotion, music, ICA.

1 Introduction

Cognitive neuroscience now recognizes that the human mind and brain do indeed have parallel perceptual and communication channels for rational / objective versus emotional / affective awareness [1]. Emotional expression and communication with others is viewed as being strongly linked to health and sense of well-being – however, until recently direct emotional communication has not been a major part of computer-based communications. The field of affective computing [2] has arisen recently to address the challenge of incorporating affective communication into information and communication technologies. Various measures can be used to track related physiological responses [3], such as electromyography (EMG), blood volume pressure (BVP), and galvanic skin response (GSR). Some researchers [4] have defined basic emotions (such as fear, anger,

sadness, happiness, disgust, surprise) based on facial expressions, among others. These and other measures are now being incorporated into consumer products, so far mainly for video gaming.

Linking emotion, EEG and music. This demonstration project was based on the assumption that spontaneous emotional associations of a musically sensitive and acculturated listener with a given musical interval are to a significant degree stimulated by properties of the musical interval itself and by a web of associations common across a musical culture. Relationships between the perceived affective character of the intervals and the harmonic ratios underlying them have long been examined [5]. Here, five whole-ratio (just) frequency ratio combinations of two low cello tones were used to create five continuous drone sounds with differing affective characters.

Although earlier efforts to differentiate emotional reactions and states via average event-related potential (ERP) measures have had limited success [6], recent efforts in these directions using more adequate measures of larger parts of the recorded electroencephalographic (EEG) dynamic information have proved more successful. As an example, real-time recognition of emotions from EEG was recently demonstrated by Liu et al. [9] using fractal dimension analysis. Recently as well, Onton and Makeig [7] reported orderly changes in the spectral character of source-resolved high-density (248-channel) EEG activity during imagination of 15 different emotions using a method of guided imagery [8]. The dynamic state differences were stable across imagination periods of 1-5 min. Further, separable features of both brain and scalp/neck muscle activities were linked to the nature of the imagined emotion.

Most previous attempts to use EEG signals to drive musical sound have attempted to sonify the EEG directly, as first reported in 1934 by Adrian and Matthews though Alvin Lucier's 1965 *Music for Solo Performer* is widely considered the first actual EEG-based musical performance. More recently, adaptation of a standard P300 BCI speller by Grierson [10] allowed users to produce a note sequence by selectively attending to symbols on a computer display. Others have attempted passive sonification of cognitive state [11], hoping that feelings engendered by the resulting soundscapes would in turn affect participant EEG. In a recent offline study, Lin et al. [12] were able to recognize music-induced emotions using Support Vector Machines and spectral EEG features. However, the production of explicitly feeling-dominated musical expression via a brain-computer interface is still largely unexplored.

A Musical emotion BCI. These results suggest the possibility of creating brain-computer interface (BCI) systems that communicate a user's feelings non-verbally, for instance via affective musical communication. So far, the nascent field of BCI systems based on EEG signals has focused on providing subjects with profound loss of motor control with the ability to perform binary or smooth prosthetic device control [13] and to communicate by spelling-out words [15] or actuating musical tones [10].

Here, we provide an account and further *post hoc* analysis of a first demonstration of the potential feasibility of an EEG-based BCI system that attempted to directly express, by musical sound production, the feeling states of its operator. We describe an experimental musical production produced for and performed at the Fourth International BCI Meeting in Pacific Grove, California. In this piece, a subject wearing an EEG cap, the 'brainist' (TM), contributed to the live musical performance

by a violinist (SM), flautist (GL), and cellist by imaginatively *re-experiencing* the emotions or affective qualities that he had spontaneously associated with five musical two-tone combinations delivered to him, during previous training sessions, under computer control as continuous, unvarying, two-tone ‘drone’ sounds. During the live performance, an online EEG classifier first attempted to identify which emotional or affective state the brainist intended to convey, then initiated production of the corresponding drone sound. As this drone sound continued, the three musicians performed an original music composition composed (SM) for the occasion ‘over’ it.

Following a more detailed account of this performance, we report re-analysis of a part of the training session EEG data collected from the brainist subject using a new method for constructing BCI classification models that may be resolved into spatially localizable brain and non-brain source features amenable to neuroscientific interpretation. Finally, we discuss the potential for both brain and non-brain information to be used in emotion BCI applications.

2 Methods

Training protocol. Following two preliminary subject training sessions, four BCI training sessions (Sessions A to D) were recorded over three days. In the two preliminary training sessions, nine different musical ‘drone sounds,’ each comprised of two recorded, pitch adjusted cello tones, were presented to the subject who was asked to spontaneously imagine emotional associations with each one. During the four BCI training sessions, this number was reduced to five drone sounds for which five associated musical pieces were composed and rehearsed out of brainist hearing.

In all sessions, the subject sat with eyes closed in a comfortable chair facing three loudspeakers. The first section of pre-recorded audio instructions asked the subject to fully relax into an emotionally neutral state. Next, the subject was asked to listen to the first musical drone sound, while imagining it to be a human emotional expression (for instance, an expressive sigh or moan). The subject was asked to attempt to empathize with the human they imagined to be the source of the sound, while also paying attention to their own somatic sensations associated with their experience. The latter suggestion was made to create a somatic feedback loop stabilizing and prolonging the subject’s empathetic experience.

The subject was asked to press a hand-held button once when he began to experience a definite feeling or emotion he spontaneously imagined as associated with the drone sound, to then attempt to strengthen and maintain this feeling for as long as possible, and to press the button a second time when the experienced emotion waned. After the second button press, the drone sound faded out and another recorded instruction asked the subject to return to their previous relaxed, emotionally neutral state in preparation for hearing and imaginatively emotionally experiencing the next drone sound. EEG recordings of the BCI training sessions were retained for model development and *post hoc* analysis, each of which comprised a sequence of extended (and not further partitioned) blocks, where each block comprised continuous data associated with the subject’s experience of an imagined empathetic feeling.

The BCI training sessions A and B were recorded on a single day in the laboratory to serve as calibration data for subsequent real-time testing. Session C was recorded for additional pseudo-online testing; Session D was the BCI training session used to train the model used in the dress rehearsal and live performance.

Performance protocol. The live performance protocol represented a verifiable attempt to demonstrate an emotion BCI operating under constraints of time and social pressure, while (hopefully) delivering a satisfying musical experience for the musical performers and audience of roughly 200 BCI researchers. The brainist sat in a comfortable chair at stage center wearing a high-density EEG cap. Right and left of the stage, elevated speakers presented a recorded audience introduction and subject instructions. These included a brief (few word) description of the feeling the brainist had used to describe the feeling he associated with each drone sound following training sessions. These descriptions were, respectively: “*uncertain, quiet, shy and sensitive*” (for the 5/4 frequency ratio, musical just major third); “*frustrated, sullen and angry*” (45/32, just tritone); “*hopeful longing*” (15/8, just major seventh), “*at peace and surrounded by love*” (3/2, perfect fifth), and “*triumphant, grandiose and exultant*” (2/1, perfect octave). These few-adjective cues constituted, in effect, the musical 'score' for the brainist to realize by re-capturing and experiencing the intended feeling, thereby initiating the associated drone sound that began each musical piece.

A few seconds after the affective cue, the BCI computer began processing brainist EEG signals until a sufficiently certain classification decision was reached. At that point, the previously associated drone sound was produced by the computer through a speaker facing the brainist. The task of the three musicians was to recognize the selected drone sound and then to play the piece written to accompany it. (In the event the computer made an unintended interval selection, the musicians would have needed to quickly bring the unexpected score to the front on their music stands; fortunately, during the performance this was never necessary). The BCI-selected drone sound continued playing throughout the performance of the musical selection, then was terminated by the BCI operator. A photograph of the performance scene is available at http://sccn.ucsd.edu/events/just_carmel10.html.

Music production and EEG recording. The drone sounds initiated by BCI classification were based on a recording of a cellist playing a series of long bowed notes on the open ‘G’ string. The upper notes were tuned to so-called “just” ratios (identified above) of the lower-note frequency (near 98 Hz). A Max/MSP software patch seamlessly looped playback of the BCI-selected musical intervals. Five (1-3 min) pieces were composed (by SM) in twentieth-century chamber music styles intended to convey feelings and harmonies compatible with each of the five drone intervals. The brainist heard these pieces only at the final rehearsal and performance, in each case after the BCI classification was complete and with thus no effect on the outcome. In all sessions, EEG was recorded from 128 scalp channels via a Biosemi ActiveTwo system (Amsterdam) at a sampling rate of 512 Hz with 24-bit resolution. In training sessions, additional peri-ocular electrodes were placed at the right and left mastoids, at the outer corner of the right eye, and below the mid line of the left eye.

Online BCI learning and classification. For live performance, a classifier was trained on previously recorded training session data using a rather conventional BCI approach, then was applied online with additional application-specific post-processing. The training data used for the performance (lasting in all 35 minutes) contained five blocks (102s +/- 12s each), one per class of emotion beginning 5 s after the first subject button press for each drone, and ending 5-s before the second press; 2-s time windows with 1.5-s overlap were extracted from these blocks and used as training trials, yielding a total of 1512 training trials. The data in each window were notch-filtered between 55-65 Hz and band pass filtered between 8-200 Hz, as suggested by the recent report of Onton and Makeig [7] in which broadband high-frequency (30-200 Hz) EEG power was found to be informative for classification of imagined emotions. Subject-specific spatial filters were then learned using the Common Spatial Patterns (CSP) method ([15] and references therein). Since standard CSP operates on only two classes, here a CSP contrast was learned separately for each pair of classes, yielding $nchoosek(5,2)=10$ CSP pair contrasts each comprised of six spatial filters. Subsequent log-variance feature extraction and classifier training was done separately for each CSP solution. As classifiers we used Linear Discriminant Analysis (LDA) with shrinkage regularization, using an analytically derived regularization parameter. The total number of features used across classifiers was 60.

During online operation, incoming EEG was classified every 200 ms using the most recent 2-s data window. This window was spatially filtered using the CSP-derived filters, and spectrally filtered as in the calibration phase. Log-variance features were then extracted and passed to the respective binary LDA classifier, whose gradual outputs were mapped onto per-class (pseudo-)probabilities. The probabilities assigned to each class were summed across classifiers according to a 1-vs-1 voting scheme. Multiple successive outputs were aggregated and averaged in a growing window. A classification decision was reached (within at most 45 s) when the estimated probability of a class exceeded a threshold that was lowered at a constant rate from 1 to 0, allowing for a quick prediction in clear cases and more lengthy accumulation of sufficient evidence in other cases. To prevent any musical selection from being selected twice in live performance, the admissible classes were those intervals that had not been played before. Thus, the online classification was effectively five-class for the first interval, then four-class, and so on.

Refined post-hoc analysis. While the emotion detector used in live performance and described above could (and did) make sufficiently accurate predictions on new data, it was not clear to what extent its performance relied on measures of brain versus non-brain source activities. Exclusion of non-brain data is less important in BCI applications for healthy users (e.g., in gaming or other HCIs), but is of practical interest when considering users who lack muscle control. During a post-hoc analysis, calibration Session A was used for advanced model calibration and Session B for model testing, as both were measured on the same day using the same electrode montage. Since for technical reasons data from the online performance was not stored, it could not be included in the post-hoc analysis.

Since each EEG channel measures a linear superposition of signals from sources distributed across brain, head and environment, it is not generally possible to interpret

single sensor signals as a measure of the activity of a distinct cortical source. This limitation can be lifted or at least minimized when spatial filters are optimized to recover source signal components that are mutually statistically independent. Following data pre-processing and automated artifact rejection using the default pipeline for Independent Component Analysis (ICA) [16] in our open-source BCILAB toolbox [17], we employed a recently-developed extension of ICA, Adaptive Mixture ICA (Amica) [18], to derive a set of maximally independent source signals, as a mixture of multiple (here six) full-rank signal decompositions (each decomposition with different, possibly overlapping, temporal support).

The components of each model (here 92 each) were then visually screened for clear brain components (see [7]) indicated by the close resemblance of their cortical maps to the projection of a single equivalent dipole or cortical patch. Non-brain (muscle, heart, eye) component processes were also identified, based on temporal and spectral properties, and eliminated, leaving a total of 38 brain component processes. A single (or in some cases dual-symmetric) equivalent current dipole model was fit to each brain component using a four-shell spherical head model. All components were localized within or on the periphery of the brain volume and above the neck or lower head region that contributes the majority of electromyographic (EMG) artifacts.

Next, trial epochs and features were extracted from the unmixed continuous multi-component signal, separately for each Amica decomposition. Epoch extraction was analogous to the original analysis: the continuous unmixed data were low-pass filtered below 90 Hz, sub-sampled to 180 Hz and then high-pass filtered above 2 Hz using a causal minimum-phase FIR filter. From these data, 3-s windows overlapping by 2.5 s were extracted and the discrete Fourier power spectral densities in each window were used as features, giving 1,183 trials in total each with 10,260 spatio-spectral features.

On these data, a weighted l_1 -regularized logistic regression classifier (realized by 1-vs-1 voting of binary classifiers) selected a sparse, best-classifying subset of spatially- and spectrally-localized features. The trials were weighted according to the temporal support of the respective underlying Amica model, and individual features were standardized similarly to the first-order model introduced in [19] and further weighted using a (here 0/1-valued) masking of brain vs. non-brain components.

During pseudo-online evaluation, the data were then causally pre-processed and mapped to per-model features as described above, and the classifiers for each model were applied to yield a discrete probability distribution over the five possible outcomes. The final probabilities were obtained as a weighted average of the classifier outputs under each Amica model, where the weight was the total probability of the respective model under the training dataset (a measure of the model's relative temporal support in the data).

3 Results

Initial offline analysis results. To determine the method to be used for the live performance, the across-session prediction accuracies of a variety of methods – CSP (see Section 2), Spectrally weighted CSP, and an implementation of Independent Modulators [7] – were assessed using data from Sessions A and B. The CSP-based classification gave the best across-session performance, reaching a single-time

window, between-session classification accuracy of 84% (chance level 20%), so was chosen for all real-time analyses. A subset of its spatial filters is shown in Fig. 2(a).

Live performance. In the concert performance, the BCI classification selected the brainist- (and composer-) intended drone sound in all five cases, though the BCI classification monitor showed that one of the selections was nearly mis-classified. BCI performance level (>70%) was thus as expected from numerical experiments on the data from the four pilot sessions. Anecdotally, in the rehearsal performance the intended first interval (major third, described by the brainist as ‘uncertain, quiet, shy, and sensitive’) was substituted by its near opposite (tritone; ‘frustrated, sullen, and angry’); the brainist later said he actually did feel frustrated at the time by the presence of unwelcome noise in the rehearsal hall.

Post-hoc data analysis. In a pseudo-online analysis, the classifier described above achieved a single-time window, between-session classification accuracy of 59%-70% when trained on Session A and applied on Session B, depending on the length of the time window (windows up to 15 s in length were tested, with longer windows giving better results). The model used approximately 1/3 of all features. Figure 1 shows a smoothed (5-s moving-average window) time course of the predicted probability distribution over the 1-s time windows of this session. The fitted dipole locations for each component in Figure 2(b)-(c) included across-models clusters of near-identical component processes in or near primary and lateral visual cortex, somatomotor cortex, dorsolateral prefrontal cortex, superior parietal cortex, middle temporal gyrus, and anterior and posterior cingulate cortex. For each brain source, we identified the coordinates of equivalent dipoles and corresponding Brodmann Area designations of the nearest gray matter using the Talairach Daemon. Figure 2(e) lists these results sorted in descending order by average absolute classifier weight of each independent component (IC). The learned spectral weights of the classifiers for these components (Fig. 2(d)) show a clear focus on alpha band amplitudes of many components, as well as sensitivity to high-frequency broadband (HFB) activity in some somatomotor and occipital components, as in previous emotion imagination results in [7].

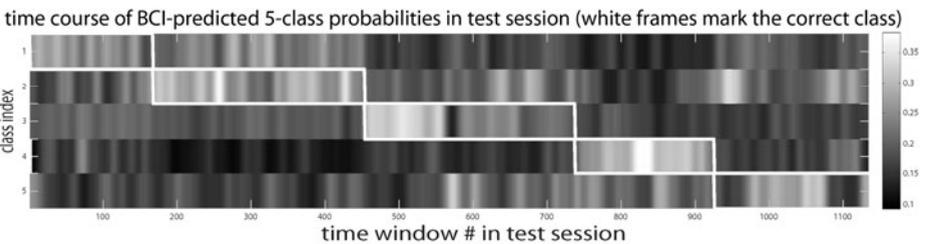


Fig. 1. Predicted probability distribution across test session (B) time windows in the post-hoc analysis. The ordinate separates the five possible outcomes of the classification. Light shades indicate high probability of the respective outcome; dark shades indicate low probability. White rectangles mark the five (true) conditions the subject was asked to imaginatively experience during the five training sessions.

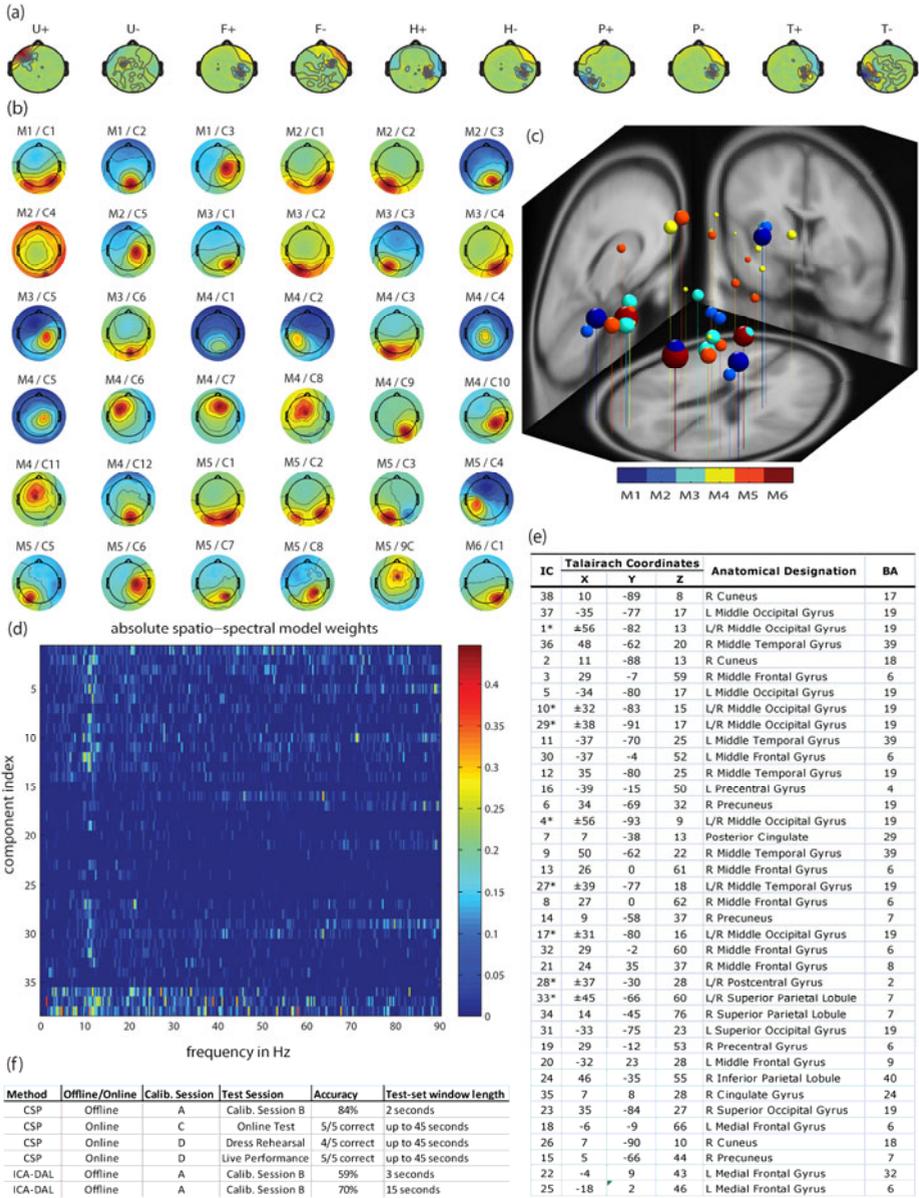


Fig. 2. Original (a) and post-hoc (b-e) classifier structure. (a) Highest-weighted CSP filters per emotion class (letter matching the leading description word), from both ends of the eigenvalue spectrum (+/-). (b) Scalp projections of leading 36 (of 38) selected independent components, grouped by Amica model. (c) Equivalent dipole locations for components in (b), surface area proportional to total weight under the classifier and colored by Amica model. (d) Absolute spectral weights for the selected components and frequencies, order corresponding to (b) when read from left to right, top to bottom. (e) Anatomical labels for equivalent dipole locations, sorted by descending total classifier weight. (f) Overview of BCI analyses and obtained results.

4 Discussion

The classification accuracy of both the CSP classifier and the ICA classifier are remarkably high, given that the chance level was 20% in both cases. The successful live performance produced its intended result of demonstrating the potential feasibility of a direct emotion BCI, here in the form of a system that used a learned vocabulary of musical sounds to express feeling states of the ‘brainist’ subject. Such a system might be usable by paralyzed users, by users limited in emotional expression, or in ordinary social settings to augment emotional communication.

Using ICA decomposition methods, we separated the scalp data into physiologically localizable brain and non-brain source process features weighted by a classifier. Localization of the involved IC sources implicated a number of anatomical regions known to be involved in visual and somatomotor imagery, self-reflection, emotion and music processing. The predominance of visual cortical areas among localized sources is not surprising given the use of visual imagery as reported by the subject. Alpha power modulation of ICs localized to bilateral occipital cortex has also been linked to changes in music structure (mode/tempo) as well as emotional responses to music [20]. Some dependence on sources localized in or near premotor cortex (MFG), precentral gyrus, and postcentral gyrus (IC 28) might also be expected since the subject was asked to pay attention to somatic sensations associated with the emotional experience, and reported significant somatomotor imagery associated with his emotional state. The precuneus has been implicated in episodic memory (including those related to the self), visuospatial processing and imagery, self-referential processing, and is thought to be the core hub of the “default mode (resting) network” [21]. Changes in regional blood flow as well as theta- and alpha-band power modulation of ICs localized to precuneus have also been linked to musical dissonance and major/minor mode distinctions [20] also tied to musical frequency ratios.

5 Conclusion

We have demonstrated the potential feasibility of a novel emotion-classification and augmented emotional communication system via a live musical performance in which EEG-based BCI classification played an artistic role, intended to suggest the development of BCI technology to enable or augment direct non-verbal emotional communication. In *post hoc* analysis of the training session data we successfully applied a novel method to learn ICA source-resolved BCI models that allow use of anatomical source constraints (here to eliminate non-brain sources) and can be interpreted in terms of localizable cortical dynamics.

Acknowledgments. We thank an anonymous donor and The Swartz Foundation (Old Field, NY) for gifts supporting this research, Theresa Vaughan for encouraging our efforts, and Julie Onton for key contributions to our emotion research.

References

1. Adolphs, R.: Cognitive Neuroscience of Human Social Behaviour. *Nat. Rev. Neurosci.* 4, 165–178 (2003)
2. Picard, R.: *Affective Computing*. MIT Press, Cambridge (1997)

3. Cacioppo, J.T., Bernston, G.G., Larsen, J.T., Poehlmann, K.M., Ito, T.A.: The psychophysiology of emotion. In: Lewis, R., Haviland, J.M. (eds.) *The Handbook of Emotion*, pp. 119–142. Guilford Press, New York (1993)
4. Ekman, P., Friesen, W.V.: *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto (1978)
5. Makeig, S.: *Theory of Interval Affect*. Monograph submitted for the Master's Thesis in Music, University of South Carolina, Columbia, SC (1979)
6. Olofsson, J., Nordin, S., Sequeira, H., Polich, J.: Affective picture processing: An integrative review of ERP findings. *Biolog. Psycholog.* 77, 247–265 (2008)
7. Onton, J., Makeig, S.: High-frequency broadband modulation of electroencephalographic spectra. *Frontiers In Human Neurosciences* 3, 61 (2009)
8. Bonny, H.L., Savary, L.M.: *Music and your mind: listening with a new consciousness*. Harper & Row, New York (1973)
9. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based Human Emotion Recognition and Visualization. In: *Proc. 2010 Int. Conf. on Cyberworlds, Singapore* (2010)
10. Grierson, M.: Composing with brainwaves: minimal trial P300b recognition as an indication of subjective preference for the control of a musical instrument. In: *Proc. Int. Cryog., Mat. Conf., Seoul, Korea* (2008)
11. Hinterberger, T., Baier, G.: Parametric orchestral sonification of EEG in real time. *IEEE Multimedia* 12(2), 70–79 (2005)
12. Lin, Y.P., Wang, C.H., Wu, T.L., Jeng, S.K., Chen, J.H.: EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine. In: *ICASSP 2009*, pp. 489–492 (2009)
13. Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T.M., Pawelzik, H., Schalk, G., McFarland, D.J., Birbaumer, N., Wolpaw, J.R.: Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology* 64(10), 1775–1777 (2005)
14. Pfurtscheller, G., Müller-Putz, G.R., Scherer, R., Neuper, C.: Rehabilitation with Brain-Computer Interface Systems. *Computer* 41, 58–65 (2008)
15. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.-R.: Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Mag.* 25, 41–56 (2008)
16. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* 7, 1129–1159 (1995)
17. Delorme, A., Mullen, T., Kothe, C., Acar, Z.A., Bigdely-Shamlo, N., Vankov, A., Makeig, S.: EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG/MEG processing. In: *Computational Intelligence and Neuroscience: Special Issue on Academic Software Applications for Electromagnetic Brain Mapping Using MEG and EEG* (2011)
18. Palmer, J.A., Makeig, S., Kreutz-Delgado, K., Rao, B.D.: Newton Method for the ICA Mixture Model. In: *Proc. 33rd IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV*, pp. 1805–1808 (2008)
19. Tomioka, R., Müller, K.-R.: A regularized discriminative framework for EEG analysis with application to brain-computer interface. *Neuroimage* 49(1), 415–432 (2010)
20. Lin, Y.P., Duann, J.R., Chen, J.H., Jung, T.P.: Electroencephalographic dynamics of musical emotion perception revealed by independent spectral components. *Neuroreport* 21(6), 410–415 (2010)
21. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3), 186–198 (2009)