Connection Science
Vol. 21, Nos. 2–3, June–September 2009, 85–88

EDITORIAL
Trends and perspectives in music cognition research and technology

This special issue on *Music, Brain, & Cognition* aims to shed light on some of the key issues in current and future music research and technology. Cognitive musicology was envisaged by Seifert (1993) and Leman (1994) to be composed from diverse disciplines such as brain research and artificial intelligence striving for a more scientific understanding of the phenomenon of music. One and a half decades following the special issue on *Music and Creativity* in Connection Science, edited by Griffith and Todd (1994), this issue, again, demonstrates how the horizons in the field have continued to expand. In recent years, computational neuroscience has attracted great aspirations, exemplified by the silicon retina (Chow et al. 2004) and the ambitious Blue Brain Project that aims at revolutionising computers by replacing their microcircuits by models of neocortical columns (Markram 2006). Research activity in auditory neuroscience, applied to music in particular, is catching up with the scientific advances in vision research. Shamma (2001) proposed that the same neural processing takes place for the visual as well as for the auditory domain. Other researchers suggested biologically inspired models specific to the auditory domain; e.g., Smith and Lewicki (2006) decomposed musical signals into gammatone functions that resemble the impulse response of the basilar membrane measured in cats.

The fast advancement of the brain computer interface (Blankertz et al. 2004) and brain-imaging methodology such as the electroencephalogram has further encouraged music research. Brain imaging grants access to music-related brain processes directly rather than circuitously via psychological experiments and verbal feedback by the subjects. A lot of experimental work in auditory neuroscience has been performed, in particular exploring the innate components of music abilities. In developmental studies of music, magnetoencephalograms have been used to study fetal music perception (Eswaran et al. 2002). Mismatch negativity in newborns has shown how babies discriminate pitch, timbre, and rhythm (Stefanics et al. 2007). A summary of electroencephalogram research in music leads Koelsch and Siebel (2005) to a physiologically inspired model composed of modules, e.g., for gestalt formation and structure building where the special features of the model are the feedback connections enabling structural reanalysis and repair.

We may assume a functional and physiological separation between sequential processing (related to musical syntax and grammar), Broca's area (Maess, Koelsch, Gunter, and Friederici 2001), and the processing of timing information, right temporal auditory cortex and superior temporal gyrus (Peretz and Zatorre 2005). Sequential processing can be seen as statistical learning (Saffran, Johnson, Aslin, and Newport 1999), in this case, learning Nth order transition sequences. The idea of statistical learning has been anticipated by Leibniz (1712): 'Music is the hidden mathematical endeavour of a soul unconscious it is calculating'. On the other hand, timing information is closely related to movement planning and kinematics. The relation between timing aspects of

ISSN 0954-0091 print/ISSN 1360-0494 online

DOI: 10.1080/09540090902734549
http://www.informaworld.com

music and movement is emphasised by the concept of mirror neurons. A mirror neuron would not only be active when the individual is articulating themselves vocally but also when the individual is observing or listening to another individual articulating a communicative sound. Prather, Peters, Nowicki, and Mooney (2008) have found neurons in the sparrow's forebrain that establish an auditory-vocal correspondence. Models of musical timing are often based on oscillators, Fourier transform, or autocorrelation.

Adaptability is an important topic on the agenda of roadmaps for the development of music technology. Adaptability helps transferring knowledge to new situations, users, or music styles. During a computer-assisted musical performance, a human performer may create new ideas (introduce new motifs, rhythms, harmonies, use new playing techniques). His mechanic companion would be required to identify the novelty and reply directly, not waiting until the musical conversation has ended. In music information retrieval, solutions have been developed to solve specialised tasks, to detect particular genres such as Cuban Son or hip-hop, to identify keys such as major and minor, or to distinguish between a 3/4 and a 4/4 meter. But would such a system be useful to identify a Cuban Hip-hop cross-over in a Blues scale and in 4/5 meter? How is the perception of a musical event influenced by the context of previous musical development and high-level structure? Active learning has been suggested to aid the user in rapidly organising collections of sounds according to the user's own criteria (Adiloğlu, Anniés, Henrich, Paus, and Obermayer 2008).

The general success of Bayesian networks inspired cognitive science as well, developing models of concept learning, inference, and surprise (Itti and Baldi 2005; Tenenbaum, Griffiths, and Kemp 2006; Purwins et al. 2008). Bayesian networks have proven to be an approach well suited to address some of the most vital phenomena in music, such as beat, expectation, attention, tension, interestingness, and surprise (Huron 2006). Maess et al. (2001) have identified neurological correlates of harmonic expectation. Fox, Rezek, and Roberts (2007) suggested a variational Bayesian system for beat and rhythm recognition that exploits prior knowledge in semi-improvised systems. Browne and Fox (2009) extracted a global tension structure from a piece and used it for automated composition.

The presented papers in this special issue show some of the variety of recent and different approaches tackling the prominent issues in music cognition and technology.

Abdallah and Plumbley use time-varying information measures and relate them to the perception of structure and to aesthetic goodness. Their work goes beyond simply measuring information content intrinsic to a piece of music. They suggest a subjective theory of response incorporating in their model the experience of the subject perceiving the music. The predictive information is defined as the Kullback–Leibler divergence between the predictive distribution knowing and not knowing the present. Thereby, events with high information gain can be identified. In addition, points of surprise and structural boundaries are detected.

Hazan et al. build a system for generation of musical expectation that operates on music in audio data format. The auditory front-end segments the musical stream and extracts both timbre and time description. In an initial bootstrap phase, an unsupervised clustering process builds up and maintains a set of different sound classes. The resulting sequence of symbols is then processed by a multi-scale technique based on N-grams. Model selection is performed during a bootstrap phase via the Akaike information criterion.

Pérez Sancho et al. focus on automatic genre classification by exploring the symbolic approach, bringing to music cognition some technologies, like the stochastic language models, already successfully applied to text categorisation. They explore some genres and sub-genres among popular, jazz, and academic music investigating the limit of using harmonic information under these models. Pérez-Sancho et al. continue to present and discuss results at different levels of the genre hierarchy for the techniques employed.

Durrant et al. employ both a general linear model and a support vector machine (SVM) to classify stimulus conditions (tonal or atonal) on the basis of the BOLD signal of novel data, and the prediction performance is evaluated. A more detailed assessment of the SVM performance reveals that the SVM is successfully identifying voxels relevant to the classification, and it is this that allows it to perform well in the classification task in spite of very noisy data and stimuli that involve higher order cognitive functions and considerably inter-subject variation in neural response.

Bharucha suggests a hierarchical scheme that builds key from frequency spectra through pitch, pitch class, and chords by self-organisation. It is shown how tonal centres and pitch salience emerge. The model accounts for fusion and pattern completion if sufficient context is available. A qualitative evaluation is performed in the light of various psychological experiments. It is discussed how fusion plays a role in the formation of pitch and harmony.

Coath et al. apply a biophysically motivated model of auditory salience to onset and beat detection. After being filtered by an auditory front-end, the transients are enhanced by calculating the skewness. The onset transient activity is further processed by a filter that maximises correlation with the formative sounds. A continuous wavelet transform together with a weighting according to perceptually plausible beats, yields an oscillogram of rhythmic frequencies, resulting in a beat detector.

Almansa and Delicado use functional principal component analysis (FPCA) to investigate expressive timing in performances of Schumann's *Träumerei*. FPCA proves to capture the performance's tempo characteristics as well as agogic subtleties (phrase structure, fermata). Deviations from the mean tempo curve are decomposed into prominent eigenfunctions. Each eigenfunction represents an expressive timing mode, with the eigenvalues indicating the prominence of these expressive features. This yields a dimension reduction of the performance data enabling clustering according to performance styles.

Hoover and Stanley design a model of music composition – more general: of musical creativity – that is based on the functional relationship between instrumental parts in ensemble playing. The system is built on a compositional pattern producing network. The user selects a pattern for the next generation of an evolutionary cycle. Thereby the network structure incrementally grows. In addition, a so-called conductor, a hidden function, endows the drum track with a beat-relative contour on its own, independent from the musical partner instruments. Examples are given for automatic percussion accompaniment, provided by an ensemble of salient melodic and harmonic instruments.

Finally, Paiement et al. present a model that is capable of predicting and generating melodies using a combination of Bayesian networks, clustering, rhythmic self-similarity and a special representation of melody. The method exploits the self-similarity of a piece and the dyadic organisation of its rhythmic structure. Then the occurring distances between rhythmical patterns are clustered. The continuation of a melody is predicted conditioned on the chord root, chord type, and Narmour group of recent melodic notes.

Acknowledgements

The emergence of this special issue has been supported by EPSRC project Le Strum (EP-D063612-1), the European projects CLOSED (FP6-NEST-PATH, contract 29085), and EmCAP (FP6-IST, contract 013123). One editor (H.P.) received a Juan de la Cierva scholarship from the Spanish Ministry of Education and Science. We would like to thank the special editorial board, consisting of Klaus Obermayer, Eduardo Reck Miranda, Xavier Serra, and John Shawe-Taylor. We owe a great thanks to the highly competent reviewers that have provided elaborated reviews often of original scientific value of their own.

References

Adiloğlu, K., Anniés, R., Henrich, F.F., Paus, A., and Obermayer, K. (2008), 'Geometrical Approaches to Active Learning', in *Autonomous Systems Self-Organization, Management, and Control*, eds. B. Mahr and S. Huanye, Netherlands: Springer, pp. 11–19.

Blankertz, B., Müller, K.R., Curio, G., Vaughan, T., Schalk, G., Wolpaw, J., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., Schroder, M., and Birbaumer, N. (2004), 'The BCI Competition 2003: Progress and Perspectives in Detection and Discrimination of EEG Single Trials', *IEEE Transactions on Biomedical Engineering*, 51 (6), 1044–1051.

Browne, T.M., and Fox, C. (2009), 'Global Expectation-Violation as Fitness Function in Evolutionary Composition', in *Applications of Evolutionary Computing, Lecture Notes in Computer Science* (vol. 5484), Berlin: Springer, pp. 538–546.

Chow, A.Y., Chow, V.Y., Packo, K.H., Pollack, J.S., Peyman, G.A., and Schuchard, R. (2004), 'The Artificial Silicon Retina Microchip for the Treatment of Vision Loss From Retinitis Pigmentosa', *Archives of Ophthalmology*, 122, 460–469.

Eswaran, H., Preissl, H., Wilson, J.D., Murphy, P., Robinson, S.E., Rose, D., Vrba, J., and Lowery, C.L. (2002), 'Short-term Serial Magnetoencephalography Recordings Offetal Auditory Evoked Responses', *Neuroscience Letters*, 331 (2), 128–132.

Fox, C., Rezek, I., and Roberts, S.J. (2007), 'Drum'n'Bayes: On-line Variational Inference for Beat Tracking and Rhythm Recognition', in *Proceedings of the ICMC*.

Griffith, N., and Todd, P.M. (1994), 'Special Issue Music and Creativity', *Connection Science*, 6 (2–3), 247–280.

Huron, D. (2006), *Sweet Anticipation*, Cambridge, MA: The MIT Press.

Itti, L., and Baldi, P. (2005), 'A Principled Approach to Detecting Surprising Events in Video', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 631–637.

Koelsch, S., and Siebel, W.A. (2005), 'Towards a Neural Basis of Music Perception', *Trends in Cognitive Sciences*, 9 (12), 578–584.

Leibniz, G.W. (1712), 'Letter to Christian Goldbach, April 17, 1712', in *Epistolae Ad Diversos* (Vol. 1), ed. C. Kortholt, Leipzig: Breitkopf, p. 240.

Leman, M. (1994), 'Schema-Based Tone Center Recognition of Musical Signals', *Journal of New Music Research*, 23, 169–204.

Maess, B., Koelsch, S., Gunter, T.C., and Friederici, A.D. (2001), 'Musical Syntax is Processed in Broca's Area: An MEG Study', *Nature Neuroscience*, 4, 540–545.

Markram, H. (2006), 'The Blue Brain Project', *Nature Reviews Neuroscience*, 7, 153–160.

Peretz, I., and Zatorre, R. (2005), 'Brain Organization for Music Processing', *Annual Review of Psychology*, 56 (1), 89–114.

Prather, J.F., Peters, S., Nowicki, S., and Mooney, R. (2008), 'Precise Auditory-Vocal Mirroring in Neurons for Learned Vocal Communication', *Nature*, 451, 305–310.

Purwins, H., Grachten, M., Herrera, P., Hazan, A., Marxer, R., and Serra, X. (2008), 'Computational Models of Music Perception and Cognition II: Domain-Specific Music Processing', *Physics of Life Reviews*, 5, 169–182.

Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999), 'Statistical Learning of Tone Sequences by Human Infants and Adults', *Cognition*, 70 (1), 27–52.

Seifert, U. (1993), *Systematische Musiktheorie und Kognitionswissenschaft* (Vol. 69 of *Orpheus-Schriftenreihe zu Grundfragen der Musik*), Bonn: Verlag für systematische Musikwissenschaft.

Shamma, S. (2001), 'On the Role of Space and Time in Auditory Processing', *Trends in Cognitive Sciences*, 5, 340–348.

Smith, E.C., and Lewicki, M.S. (2006), 'Efficient Auditory Coding', *Nature*, 439, 978–982.

Stefanics, G., Haden, G., Huotilainen, M., Balazs, L., Sziller, I., Beke, A., Fellman, V., and Winkler, I. (2007), 'Auditory Temporal Grouping in Newborn Infants', *Psychophysiology*, 44, 697–702.

Tenenbaum, J.B., Griffiths, T.L., and Kemp, C. (2006), 'Theory-Based Bayesian Models of Inductive Learning and Reasoning', *Trends in Cognitive Sciences*, 10 (7), 309–318.

Hendrik Purwins[a,b] and David R. Hardoon[c]
[a] *Music Technology Group*,
Universitat Pompeu Fabra, Barcelona;
[b] *Neural Information Processing Group*,
Technische Universität Berlin;
[c] *Centre for Computational Statistics and Machine Learning*,
University College London