Designing symbiotic composing

Ilkka Kosunen* and Aleksander Väljamäe†

Tallinn University, Narva mnt 25, 10120 Tallinn, Estonia

(Received 30 September 2018, Accepted for publication 27 April 2019)

Keywords: Biofeedback, HCI, Music, Artificial intelligence, User modelling

PACS number: 43.60.Np, 43.64.Ri [doi:10.1250/ast.41.322]

1. Introduction

Computational creativity has been applied in the domains of music [1], visual arts [2], poetry and prose [3]. However, while many sophisticated algorithms exist for generating novel, creative themes and ideas, what computers still completely lack is the ability to judge the quality of their creations. That is, they lack the subjective aesthetic capabilities of humans. While computers can generate near infinite amount of melodies, text, and shapes of colors, they are completely unaware of the possible artistic values of their creations. A computer is like the monkey of the infinite monkey paradox: it has been said, that if a monkey were allowed to punch keyboard an infinite amount of time, eventually it would produce the collected works of Shakespeare. Similarly, modern computers can produce near infinite amount of raw “artistic material,” yet are incapable of picking up the meaningful pieces out of the overall noise; this is the problem symbiotic composing aims to solve: by combining the near infinite creative capability of computers with the aesthetic judgments of a human, the man and the machine can achieve an artistic symbiosis that eclipses what either could do separately. The symbiotic composing goes, however, beyond simple explicit feedback from the user and aims to use the tools of physiological computing to extract from psychophysiological recordings automatically both the conscious- and perhaps even unconscious- preferences for different aural patterns, as well as the affective and cognitive reactions to various musical parameters.

The symbiotic system consists of following parts:

- Physiological computing: the recording and processing of the psychophysiological signals such as electroencephalography (EEG), electrocardiography (ECG) or electrodermal activity (EDA) among other biosignals.
- Machine (deep) learning: the machine learning system that determines from the physiological input how the piece of music or audio was perceived, that is, the detection and quantification of the *aesthetic judgment* of the user.
- Computational Creativity: Creative artificial intelligence (AI) to generate new musical material, guided by the input from the machine learning (ML) system that provides real-time feedback on the human composer’s response to different musical ideas.

In the following section, we will first briefly introduce each of the three components of the symbiotic composing system followed by a description and simple example on how they could be combined to realize the symbiotic composing vision. Finally, we discuss on limitations of current systems and the future directions on how symbiotic composing can evolve.

2. Background

2.1. Physiological computing

Physiological computing is a type of Human-Computer Interaction (HCI) where the interaction is based on measuring the user’s physiological signals and then adapting the system to these signals, usually in real-time [4]. Ordinary HCI is highly asymmetric as the computer can present very complex audiovisual feedback to the user while the user is usually limited to the low bandwidth provided by mouse and keyboard [5].

Physiological computing also allows the computer access to the cognitive and affective states of the user, thus facilitating *affective computing* [6]. In the context of acoustics, this would provide a real-time and temporarily highly precise (in the range of milliseconds) view into how the user perceives a change in the aural environment. In the context of this paper, we are primarily interested in how different implicit reactions to audio stimuli could be detected from the physiological signals. That is, we are interested in detecting the aesthetic judgment as well as the affective reaction of the user automatically.

In our previous work we have shown that it is possible to detect various cognitive and affective states, such as boredom [7] and basic emotions like anger and sadness [8], directly from the physiological signals. We have also shown that it is possible to detect whether given stimulus was relevant to the user [9,10] and even whether the content was funny [11]. Now the aim is to do the same for the aesthetic judgment of aural stimuli.

2.2. Machine learning

Psychophysiological recordings generate a very complex data set of multiple time-series at different sampling frequencies, each of which can have very different interpretation. Various machine learning methods have been used to classify affective states from physiological signals successfully, including K-nearest neighbors [12], Linear Discriminant analysis [13], Naive Bayes [14] and many others (see [15] for review). However, the utilization of the modern deep

*e-mail: ilkka.kosunen@gmail.com

†e-mail: aleksander.valjamae@gmail.com

learning methods [16] show promise to bring a new level of performance to the analysis of physiological signals as compared to the classical, or “shallow,” methodologies [17,18]. We aim to utilize these modern methods to detect the aesthetic judgment of the aural stimuli.

2.3. Computational creativity

Computational creativity has been applied in the domains of music [1], visual arts [2], poetry and prose [3]. Most computational creativity systems are based on the idea that the computer is given large corpus of existing material, such as Bach’s chorals or Shakespeare’s sonnets, and then the machine learning algorithm tries to build a statistical model on how often different notes or letters appear together. Once the model is trained, it can then be sampled to produce “new” works of art. However, what these systems lack is a way for the computer to tell whether these new creations were successful or not, that is, the computer lacks the human aesthetic judgment; this is what we aim to change by allowing the computer to “borrow” the human composer’s “aesthetic taste” by directly measuring the psychophysiological responses the human composer has to the creative ideas of the computer.

2.4. Dual-process model and musical emotions

When working on the symbiotic composing system, we would rely on seminal work by Juslin and Västfjäll [19] where six psychological mechanisms are described through which music might induce emotions: brain stem reflex, evaluative conditioning, emotional contagion, visual imagery, episodic memory and musical expectancy. The six mechanisms represent a complex interplay between cognitive and emotional processing, with some being more related to automatic, emotional responses to loudness (brain stem reflex), while others requiring cognitive analysis of musical structure (musical expectancy). Interacting with artificial composer takes place on two levels, both explicitly (via composition) and implicitly (via physiological reactions). This closely follows the dual-process theory and fast vs. slow thinking described by [20]. The use of the dual-process model in creativity-enhancing interfaces has been already discussed in [21]. In our work, we specifically focus on implicit interaction between human and computer during symbiotic creation.

3. Example scenario: A symbiotic piano

In this section, we outline how a symbiotic composing system would work. As a basis, we use an idea similar to the *continuator* [22] which was a system that allowed a human performer to have a musical dialog with the machine: the musician would play a theme, and the Continuator would respond with a musical variation. The continuer had access only to the music played by the human participant, but in the symbiotic system, there would be the additional channel of psychophysiological data that the system can use to annotate the musical idea. That is, when the human player is improvising something the computer can also determine which parts the human was particularly satisfied with, and which parts they might have felt did not work so well. The system can then use this additional information when generating its musical variation. Furthermore, when playing its variation, the computer can again record how well it was perceived by the human and update its model accordingly.

Fig. 1 The generic workflow of a symbiotic composing system.

Then the human composer would consider the computer’s idea and use it as a starting point for the next round of iteration (see Fig. 1). In the following, we detail each step of the symbiotic loop.

- (1) The human composer initiates the composing process by improvising on a MIDI keyboard. The system records two streams of data:
 - Stream A (music): MIDI.
 - Stream B (physiology): EDA, ECG, EEG.
- (2) Stream B goes through pre-processing and feature extraction after which the two streams are synchronized and forwarded to the learning module.
- (3) The learning module generates a model based on the MIDI information and adjusted by the corresponding physiological data: if the human composer showed affective or cognitive response during a certain part of their improvisation, the model would weight that part as more important.
- (4) The artificial composer will then sample from the model build in the previous step a new musical variation and play it back to the human composer.
- (5) The system will observe the human composer’s reactions to the new musical variation and again update its model based on the human composer’s reactions.
- (6) Human composer listens to the version of the system. If he is satisfied, the final version is stored, and the process ends. Otherwise, the system loops back to point 1.

3.1. Extensions

In the previous section, we described how the symbiotic system could work in a classic example of a composer writing music alone on a piano. However, the same principles can be utilized in many other ways. For example, it has been shown that such *interactive reflexive musical systems* like the Continuator [22] and the symbiotic piano can be beneficial for early childhood music and movement education [23]. They utilize mirroring, turn-taking, imitation, reflexive listening and co-regulation that help the child in the construction of a “musical self” [24].

The system could also easily be changed to include one or more audience members: instead of observing the physiolog-

ical responses of the human composer, the system would instead monitor the audience members and create variations that seem most pleasing to them.

The symbiotic system is not only limited to composing music: other aural parameters can also be optimized using similar iterative, physiologically adapted loop. For example, an audio engineer could work on the equalization, mixing and sound effects.

3.2. Detecting the aesthetic judgment

At the core of the symbiotic system is the computer system's ability to detect and quantify the aesthetic response of the human composer to a given audio stimulus. In the previous work, the neural correlates of aesthetics have been studied both with EEG [25,26] as well as with peripheral physiology [27]. While we plan to utilize this information when planning the data pre-processing and feature extraction from the physiological signals, our approach will more closely resemble the end-to-end type of machine learning that has been proven successful in providing computers with human-like capabilities [28]. In other words, we aim to conduct empirical user studies to have participants listen to aural stimuli and label them, and the use machine learning methods to extract the correlates of aesthetic judgment. One interesting question will be whether these correlates are universal or individual, that is, does the system need to be trained separately for each user.

4. Discussion and future work

In this paper, we described the idea of a symbiotic composing system and how it would be designed. While we concentrated on the topic of music composition, the same design could also be used for a variety of other topics/applications where sensing human physiology in real-time reveals the cognitive and emotional state of a user and therefore assist the symbiotic creative process between man and machine:

- assisted film editing;
- photography and cinematography where, e.g., the shooting process is guided by a symbiotic system that is aware of director's attention;
- live interactive theatre performances, such as neuro-theatre, where implicit reactions of actors and viewers could be combined into a united biological system or android theatre where robots and humans co-exist;
- adaptable virtual acoustics software that would allow for finding new types of aural architecture;
- symbiotic computing can be also seen as a research tool studying basic principles of human cognition and emotion, e.g., emotional responses to music, using iterative process.

While this work is still very much work-in-progress, we would like the opportunity to discuss the particularities of the symbiotic design with the community to both improve our design as well as better explain the complexities involved in neuroadaptive creativity. For example, symbiotic composing imply that physiological reactions of the human part become know an important component of the whole system. Till recently, it is mostly drama actors and other representatives of performing arts domain for who central nervous system is

literally a tool. Wide adoption of symbiotic systems would then reinforce implicit interaction channel between humans and machines.

As a first step, we aim to collect data in co-operation with composers and music students to explore the correlates between musical ideas and physiological responses. This basic research allows us to more clearly explore the physiological aspects of aesthetics. In the following face we then aim to build a classifier that allows in real-time classify a musical idea in a way that facilitates and guides the creative, generative model.

References

- [1] F. Pachet, "A joyful ode to automatic orchestration," *ACM Trans. Intell. Syst. Technol.*, Special Issue on Intelligent Music Systems and Applications (2016).
- [2] S. Colton, "The painting fool: Stories from building an automated painter," in *Computers and Creativity* (Springer, Berlin, Heidelberg, 2012), pp. 3–38.
- [3] R. P. Pérez and M. Sharples, "MEXICA: A computer model of a cognitive account of creative writing," *J. Exp. Theor. Artif. Intell.*, **13**, 119–139 (2001).
- [4] S. H. Fairclough, "Fundamentals of physiological computing," *Interact. Comput.*, **21**, 133–145 (2008).
- [5] L. J. Hettinger, P. Branco, L. M. Encarnacao and P. Bonato, "Neuroadaptive technologies: Applying neuroergonomics to the design of advanced interfaces," *Theor. Issues Ergon. Sci.*, **4**, 220–237 (2003).
- [6] R. Picard, *Affective Computing* (MIT Press, Cambridge, MA, 1997).
- [7] D. Giakoumis, A. Vogiannou, I. Kosunen, K. Moustakas, D. Tzovaras and G. Hassapis, "Identifying psychophysiological correlates of boredom and negative mood induced during HCI," *Proc. B-Interface 2010*, pp. 3–12 (2010).
- [8] O. Barral, I. Kosunen, T. Ruotsalo, M. M. Spap, M. J. Eugster, N. Ravaja, S. Kaski and G. Jacucci, "Extracting relevance and affect information from physiological text annotation," *User Model. User-Adapt. Interact.*, **26**, 493–520 (2016).
- [9] O. Barral, M. J. Eugster, T. Ruotsalo, M. M. Spap, I. Kosunen, N. Ravaja, S. Kaski and G. Jacucci, "Exploring peripheral physiology as a predictor of perceived relevance in information retrieval," *Proc. 20th Int. Conf. Intelligent User Interfaces*, pp. 389–399 (2015).
- [10] M. J. Eugster, T. Ruotsalo, M. M. Spap, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci and S. Kaski, "Predicting term-relevance from brain signals," *Proc. 37th Int. ACM SIGIR Conf. Research & Development in Information Retrieval*, pp. 425–434 (2014).
- [11] O. Barral, I. Kosunen and G. Jacucci, "No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment," *ACM Trans. Comput.-Hum. Interact. (TOCHI)*, **24**(6), p. 40 (2018).
- [12] G. Rigas, C. D. Katsis, G. Ganiatsas and D. I. Fotiadis, "A user independent, biosignal based, emotion recognition method," in *International Conference on User Modeling* (Springer, Berlin, Heidelberg, 2007), pp. 314–318.
- [13] J. Blechert, M. Lajtman, T. Michael, J. Margraf and F. H. Wilhelm, "Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information," *Biomed. Sci. Instrum.*, **42**, 136–141 (2006).
- [14] J. Zhai and A. Barreto, "Stress detection in computer users through non-invasive monitoring of physiological signals," *Biomed. Sci. Instrum.*, **42**, 495–500 (2006).

- [15] D. Novak, M. Mihelj and M. Munih, “A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing,” *Interact. Comput.*, **24**, 154–172 (2012).
- [16] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” *Nature*, **521**(7553), 436 (2015).
- [17] S. Jirayucharoensak, S. Pan-Ngum and P. Israsena, “EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation,” *Sci. World J.*, **2014**, 627892, 10 pages (2014).
- [18] Y. R. Tabar and U. Halici, “A novel deep learning approach for classification of EEG motor imagery signals,” *J. Neural Eng.*, **14**(1), 016003 (2016).
- [19] P. N. Juslin and D. Västfjäll, “Emotional responses to music: The need to consider underlying mechanisms,” *Behav. Brain Sci.*, **31**, 559–575 (2008).
- [20] D. Kahneman and P. Egan, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York, 2011).
- [21] R. Tubb, “Creativity, dual process theory, and the navigation of music solution-space,” *Proc. 7th AISB Symp. Computing and Philosophy*, 4 pages (2014).
- [22] F. Pachet, “The continuator: Musical interaction with style,” *J. New Music Res.*, **32**, 333–341 (2003).
- [23] A. R. Addressi, M. Maolli and F. Anelli, “The MIROR platform for young children’s music and dance creativity: Reexive interaction meets body-gesture, embodied cognition, and laban educational dance. Perspectives,” *J. Early Music Mov. Assoc.*, **10**, 9–17 (2015).
- [24] A. R. Addressi, “Developing a theoretical foundation for the reflexive interaction paradigm with implications for training music skill and creativity,” *Psychomusicology*, **24**, 214–230 (2014).
- [25] L. Höfel and T. Jacobsen, “Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes?” *Int. J. Psychophysiol.*, **65**, 20–31 (2007).
- [26] M. Müller, L. Höfel, E. Brattico and T. Jacobsen, “Aesthetic judgments of music in experts and laypersons: An ERP study,” *Int. J. Psychophysiol.*, **76**, 40–51 (2010).
- [27] E. Wassiliwizky, S. Koelsch, V. Wagner, T. Jacobsen and W. Menninghaus, “The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles,” *Soc. Cogn. Affect. Neurosci.*, **12**, 1229–1240 (2017).
- [28] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski and S. Petersen, “Human-level control through deep reinforcement learning,” *Nature*, **518**(7540), 529 (2015).

Ilkka Kosunen has spent more than a decade in developing all kinds of real-time adaptive physiological computing systems ranging from first person shooter games to neuroadaptive virtual reality meditation systems. He has published over 30 scientific articles on biofeedback, neuroadaptive systems, brain-computer interfaces and related topics.

Aleksander Väljamäe is an assoc. professor in physiological computing at Human-Computer-Interaction group, Tallinn University, Estonia. His current psychophysiology research concerns how audiovisual media influence humans on perceptual, cognitive and emotional levels, with particular stress on well-being and new media applications (neurocinema, neurotheatre).