Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 169 (2020) 606–614

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Creative virtual composer assistant based on the eBICA framework

Dmitry O. Kostenko, Ivan A. Mashtak, Danila D. Fomin, Darya V. Mashtak, Anastasia V. Razheva, Natalia V. Kim, and Alexei V. Samsonovich*

National Research Nuclear University "MEPhI" (Moscow Engineering Physics Institute), Kashirskoe Shosse 31, Moscow, 115409, Russian Federation

Abstract

The topic addressed here is the role of emotions in music creation. The fact that musical fragments can be evaluated on a number of emotional scales allows one to create a semantic map of fragments and to use this map as a guidance in music creation. The work presents results of the study of a virtual composer assistant based on the semantic map and on the graph of similarity (computed using *LibROSA*) of selected 158 musical fragments (*Apple Loops*). The assistant was built based on the eBICA cognitive architecture framework (Samsonovich, 2013, 2018). Participants were 16 NRNU MEPhI Undergraduate Students. Fragments were combined into compositions by participants guided by the assistant, and then evaluated on a number of scales by other participants. Four experimental conditions were compared: with and without guidance by the map, with and without guidance by the similarity graph. Results indicate that the usage of similarity data by the assistant significantly improves the quality of resulting compositions. At the same time, the effect of usage of the semantic map was found significant only in the absence of hints based on the similarity data. Nevertheless, obtained results support the idea of the semantic map significance in composer assistant's function. One of the next goals in this study will be to generate music automatically, following the dynamics of actual or expected human emotional states, e.g., during a videogame. In general, the advantage of this new technology is in its generality: it will have broad applications in many areas of artificial intelligence.

Keywords: *virtual actor; composer assistant; computational creativity; emotional intelligence; affective computing; cognitive architecture*

* Corresponding author. *E-mail address:* avsamsonovich@mephi.ru

1877-0509
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures.
10.1016/j.procs.2020.02.203

1. Introduction

Many studies have shown that music is inextricably linked to logic, and the process of music creation is related to execution of algorithms. Nevertheless, good music cannot be created or perceived adequately without emotional feelings [1]. Furthermore, it appears that music creation is not possible without engagement of personal feelings of the composer. There always should be an intimate emotional component in the process of music creation.

This work makes an attempt to further clarify this intuition about a connection between music creation and emotion. The idea is to use a semantic map of music fragments to design an assistant to the composer in the form of intelligent agent. The expected outcome is a personal assistant to the composer that improves human performance in music creation, specifically, when the composer attempts to continue a melody or to select a suitable accompaniment to a given melody. The assistant should maintain an emotional contact with the composer and choose the hints for melodies that are adequate not only to the current tune, but also to the current user's emotional state. At the same time, the assistant should be emotionally creative. Its suggestions should be nontrivial, yet positive and constructive.

In general, the topic of computational creativity used for computer-assisted music creation is very hot today [2]. In our work we selected an approach based on a weak semantic map [3,19]. To determine our map elements we need to take into account that every element should have an emotional assessment. Such map provides one approach to build cognitive architectures that will support our assistant and other creative agents, e.g., autonomous actors.

2. Materials and Methods - Basics

2.1. Background and procedures

At first, we recalled the basic theory of music. Especially we are interested in the notion of harmony. The general concept of harmony has an empirical nature and refers to the logical construction of sounds [4,5]. It means that not every combination of sounds is harmonic. In general, music consists of combinations of sounds. In this work, we shall limit our consideration to musical compositions representable as sequences of short musical fragments called here "loops". We shall regard loops as elements of a semantic map, the coordinates of which measure aesthetic feelings associated with the loops. One of our goals here was to construct and characterize this map (see also [6]).

The entire theory of musical perception is based precisely on the personal feelings of the composer and the listeners. The concept of sensation seems too subjective to be relied upon in an algorithm used to create music. Our first task associated with each given loop was to determine, what kind of sensations this loop will elicit in the listener. The answer was found empirically – by polling participants who enjoyed listening the loops.

In harmonic perception, there are two main points: the perception of the ladder structure of the chords and the perception of the very nature of the sound. They both contribute to the emotional loop perception. In our study, we used 158 Apple Loops selected from the GarageBand library.

In general, the melody emerges from the interaction of various forms of organization of musical sounds, primarily frets and rhythms. A common way to build a melody is by the choice of intervals within one or several keys. Determining intervals between tones is the process of creation of a musical composition. In our case, however, the process of composing was reduced to selection of a loop sequence using hints provided by the composer assistant, while all loops had the same tonality and compatible rhythms.

2.2. Participants

20 Undergraduate students from NRNU MEPhI Cybernetics Department of the age 18 to 25 participated in the study. 40% of them were women and 60% were men. All had Russian as their native language and all were fluent in English (the GUI was labeled in English, and the test questions were presented in Russian).

2.3. Algorithm for determining semantic distances between loops

To find similar loops we used *Python*, *LibROSA*. LibROSA is a Python package for music and audio analysis [7]. It provides the building blocks necessary to create music information retrieval systems. In designing LibROSA a few key concepts were prioritized. First, striving for a low barrier to entry for researchers familiar with MATLAB. It is useful, because we use MATLAB in our project. In particular, there was opted for a relatively flat package layout, and following scipy rely upon numpy data types and functions, rather than abstract class hierarchies. Second, the interfaces, variable names, and (default) parameter settings across the various analysis functions were standardized. Third, backwards compatibility against existing reference implementations is retained.

We download two files that need to be compared. Using LibROSA function we get estimates for 12 *mel-frequency cepstral coefficients (MFCCs)* [8].

Mel it is a unit of pitch based on the perception of that sound by our hearing organs. The amplitude-frequency response of the human ear does not resemble a straight line. Therefore, amplitude is not an exact measure of sound loudness. The pitch of a sound perceived by the human ear does not quite linearly depend on its frequency.

It is also necessary to clarify what is *Cepstrum*. Our speech is an acoustic wave that is emitted by the organ system and then converted into the vocal tract. Human speech apparatus can be represented as a set of generators of tones and noises, as well as filters.

Therefore, *MFCC* is a kind of representation of the signal spectrum energy [9]. We use the signal spectrum, which allows us to take into account the wave "nature" of the signal further. This spectrum is projected on a special mel-scale, allowing to select the most important frequencies for human perception. It is done for each time period.

After getting two arrays, we use *Dynamic Time Warping* algorithm [10]. It is used to find the greatest correspondence between time sequences. It is determined how two voice signals represent the same original spoken phrase. This algorithm helps to measure the distance between loops.

3. Building and Using Weak Semantic Map in the Assistant

The process of automated generation of a musical composition by the assistant cannot be based solely on the theory of music. The assistant should be creative, and should take into account emotional state of the composer [11]. Therefore, principles of its work should include elements of creativity that are sensitive to emotional appraisals of the created music and emotional assessment of the current mental state [12,20] of the composer, who guides the process.

Given these requirements, an element of the map should be a loop with the corresponding assessment. This database of combinations and their assessments (appraisals) will be used in the process of generation of a melody or an accompaniment. Using these appraisals, together with continuous evaluation of the emotional state of the human composer will help us to adjust the mood of the composition, matching composer's desires.

3.1. Semantic map creation

Ten students took a part in the map creation experiment. Participants were listening to the selected loops. Then they were asked to rank the loops. Using SurveyMonkey, participants gave grades form 1 to 9 to each loop on a set of scales, from which two characteristics were extracted using the Principal Component Analysis. The resultant characteristics were interpretable as valence (optimistic/pessimistic) and arousal (intense/calm). These characteristics were averaged over all participants and normalized to reduce their range of variability to the interval (-1, 1).

The result was a semantic map of loops that was used in the following experiments. This kind of a map corresponds to the notion of a weak semantic map, for which we refer the reader to another work [13]. In general, a semantic map is an embedding of a set of representations in semantic space. Semantic maps are ubiquitous in the brain and can naturally represent growing systems of values. Therefore, they should play a key role in emotionally guided cognition and learning.

3.2. Algorithm for hints generation by the assistant using the semantic map

Using the algorithm described in Section 2.3, we find for each track 75 most similar ones. In our GUI, these 75 tracks are represented by a green color of variable brightness. The more similar the track is to the original, the brighter the green color of that track is. The algorithm of hint generation can be outlined as follows.

Step 1. At first, we select a starting point. Then we choose the main direction of movement, limited to an axis direction. There are two axes on our map. Accordingly, the initial movement can occur in one of the four directions. The choice of direction is determined by the greatest absolute value of the current coordinates.

Step 2. We define n nearest points to the starting point along the selected direction. In our case, n is between n_{max} = 15 and n_{min} = 5. The value of n depends on the total number of loops and the number of loops found in the vicinity of the current location. If it is impossible to select at least n_{min} points, we jump to step 4.

Step 3. The next step is to choose m = 5 points from n. These points are the nearest to the starting one on the second axis. These five points are taken as the hint that the assistant will highlight on the semantic map. The process continues with Step 2, unless terminated by the user.

Step 4. If we could not determine at least five points in step 3, we must change the direction of movement. The direction changes by 90 degrees counterclockwise until at least five points can be selected. The process continues with Step 2, unless terminated by the user.

From the perspective of this work, the main function of the semantic map is to store and visualize the information about the loops database assessments. In control sessions, a random arrangement of loops within the square was used instead of the semantic map.

3.3. Experiment 1

In total, 12 participants took part in composition creation in this experiment. There were three scenarios – three kinds of sessions used in this experiment, all described below. Then 19 participants rated the resulting compositions on 3 scales described below.

Scenario 1. Participants composed music without any hints, from the assistant, using a random arrangement of loops on the square. The procedure is as follows. There is an assistant window in front of the user (Fig. 1). On the left of the window, there is a map with tracks, and on the right – the current state of the composition represented as a sequence of loops. Dots on the map correspond to loops. Coordinates of loops, however, are randomly generated: this map is not semantic.

Scenario 2. Participants composed music using hints that were generated based on the similarity graph only, without using the semantic map. The procedure is as follows. The assistant highlights dots on the map that represent the most similar loops to the last selected loop. The map is still not semantic: dot coordinates are random. The algorithm described above is used only to determine the similarity of the loops and to select most similar ones (Fig. 2).

Scenario 3. Participants composed music using hints that were generated based on the similarity and on the semantic map, as prescribed by the above algorithm. The procedure is as follows. The map becomes semantic. The hints will include not only an assessment of similarity, but also semantic affinity. The location of the points on the map will be very helpful (Fig.1 3).

Experimental paradigm. Participants with the help of the composer assistant implemented in Matlab create compositions. Each participant is required to create three compositions in each of the three scenarios. The order of scenarios was randomized and balanced for all participants, to exclude the primacy effect. After that, the created compositions were replayed in a random order for the audience of participants. Each composition was rated on the following three scales.

1) General aesthetic quality of the composition.
2) Integrity and internal consistency of the composition.
3) Consistency of transitions among fragments in the composition.

Research questions and data analysis. Statistics was computed over all rankings of all compositions. The two control groups in the experiment are Scenario 1 and Scenario 2. Comparing the first scenario to the second one helps us to understand whether the similarity graph is useful. Comparison of the second and third scenarios allows us to assess the additional impact of the semantic map on the resultant quality of music.

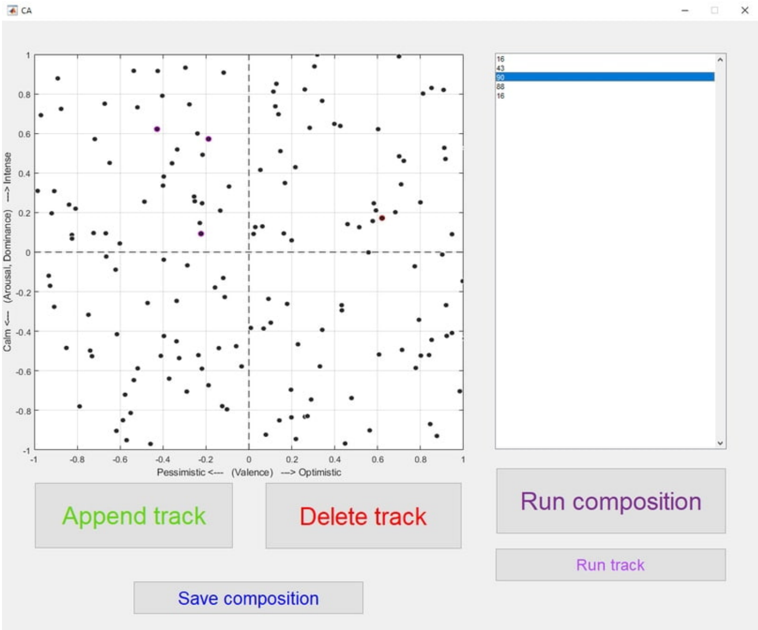

Fig. 1. GUI of the composer assistant application used to create music without receiving hints from the system. Arrangement of dots representing loops and their colors are random.

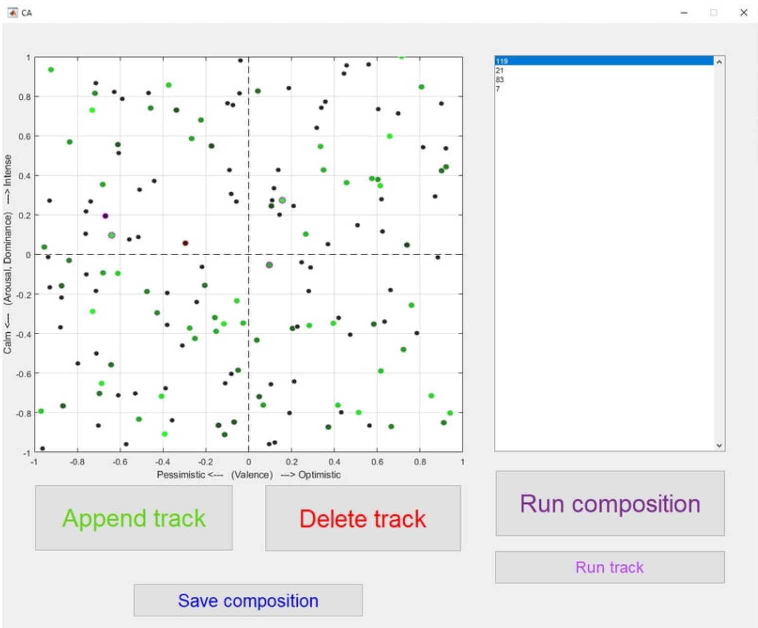

Fig. 2. GUI of the composer assistant application used to create music with hints based on the similarity of loops. Arrangement of dots representing loops is random.

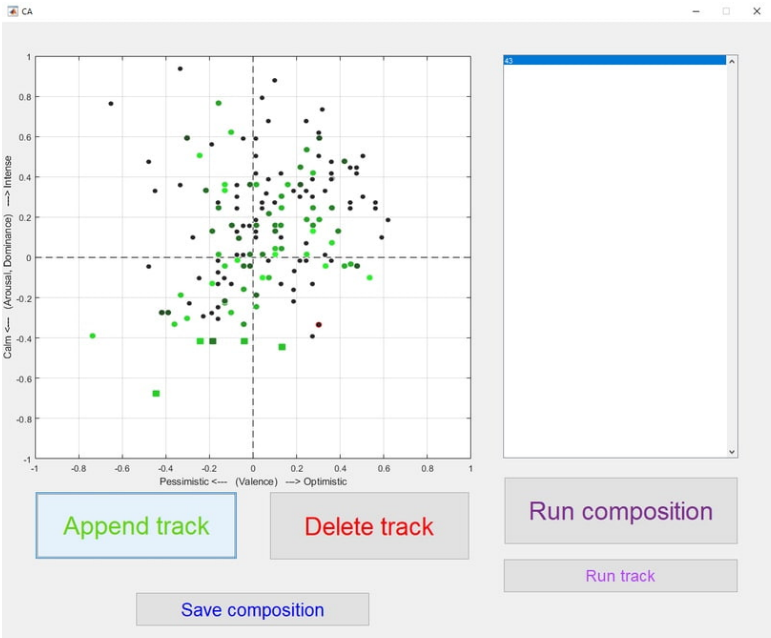

Fig. 3. GUI of the composer assistant application used to create music with hints based on both, the loop similarity and the semantic map. Arrangement of dots corresponds to the actual semantic map based on human rankings.

3.4. Experiment 2

In total, 8 participants took part in composition creation in this experiment. Two scenarios were used.

Scenario 1. Participants are creating music without any hints from the assistant. This scenario is the same as the first scenario used in Experiment 1 (described above: see Fig. 1). The map is random.

Scenario 2. Participants are creating music using hints generated based on the semantic map only. The map is semantic (Fig. 3). The hints offered by the assistant to the composer use semantic affinity of loops. The loop locations on the map may be also helpful to the composer.

Experimental paradigm. Participants with the help of the Matlab-implemented application create compositions. It is required that each participant creates three tracks using each scenario. The order of scenarions used by participants is individually randomized and balanced among participants. After that, the created compositions are played in a random order to the audience of participants. Each track is rated by all participants, except the author, on the same three scales that were used in Experiment 1 (described above).

Research questions and data analysis. Comparing the first scenario with the second helped us to understand whether the usage of semantic map affects characteristics of the created compositions.

4. Main Results and Analysis

4.1. Experiment 1

The outcome is summarized in Table 1. There are two control groups in the experiment, corresponding to Scenario 1 and Scenario 2. Comparing the first scenario with the second helped us to understand that the similarity score is useful and can be significant for the quality of the outcome (all differences between Scenarios 1 and 2 shown in Table 1 are significant).

At the same time, comparison of the second and third scenarios showed that at least in this case the usage of semantic map data did not allow participants to significantly improve the resultant characteristics of compositions.

Table 1. Summary of the mean values of rankings obtained in the first experiment.

Scenario	Aesthetic value	Integrity	Consistency
Scenario 1	2.7	2.5	2.3
Scenario 2	3.8	3.8	4.3
Scenario 3	3.6	3.3	3.5

4.2. Experiment 2.

The outcome of the second experiment is summarized in Table 2 and in Fig. 4. Comparison of the first and second scenarios showed the positive effect of semantic map usage on the qualities of resultant compositions in the absence of the similarity data.

Table 2. Summary of the values of rankings obtained in the second experiment.

Scenario	Aesthetic value mean, (stand.error)	Integrity	Consistency
Scenario 1	3.2 (0.2)	3.1 (0.2)	3.0 (0.2)
Scenario 2	3.8 (0.2)	3.4 (0.3)	3.6 (0.3)
P-value	0.0145	0.1694	0.0325

Results of the t-Test show that the improvement in both, the aesthetic value and the consistency of transitions, are significant ($p<0.015$ and $p<0.033$, respectively), while the difference in the integrity of individual compositions is not significant ($p>0.16$).

5. Discussion

Results of this study indicate that semantic maps can help a creative human composer to improve the quality of her/his work, when a map-based composer assistant is used. Because the architecture of this assistant based on the semantic map is consistent with the eBICA freamework [13,14,18], we can conclude that the eBICA framework provides a useful and efficient basis for the design of creative assistants, such as the composer assistant implemented and studied here.

An unexpected disappointment was the finding that the second and third scenarios in Experiment 1 did not differ significantly from each other in their metrics (Table 1), indicating that at least in this case the usage of semantic map data did not allow participants to make an improvement of their compositions qualities. This could happen for several reasons: e.g., semantic-map-based filtering during hint generation cold interfere with the filtering based on the similarity score. In other words, the presence of two hints at the same time confused the subjects, whereas just one hint based on the loop similarity alone worked better. This effect could be anticipated. In fact, it is eliminated by removal of the similarity information, when semantic map usage proves to be helpful (Table 2).

Another possible reason for the partially negative outcome of the first experiment could be that semantics-based loop arrangement results in a more compact distribution of dots, resulting in a poor separability of dots. These and similar factors can be excluded by the appropriate design of the assistant.

The work presented here continues our previous efforts in this direction [11,15]. Present findings support the idea of the semantic-map-based eBICA framework significance in composer assistant's function. Our next task is to generate music automatically, following the dynamics of actual or expected human emotional states, e.g., during a videogame. In such applications, emotional states can be registered using biometrics and behavioral data. In general, the advantage of this new technology is in its generality: it will have broad applications in many areas of Artificial Intelligence (AI).

Nowadays the sphere of AI is actively developing. The original spirit of the idea of AI was to replicate all the key intelligent functionality of the human mind in a machine. In the near future autonomous robots and intelligent agents as well as physical and virtual co-robots and personal assistants are expected to be an indispensable part in machine's collaborative work with humans [16]. It is expected that virtual agents and robots will integrate into human teams and work as general-purpose assistants to us. The use of them was found in almost all spheres of human activity. There is virtually no area in which man has not tried to create an automatic assistant.

However, the question of the role of socio-emotional intelligence in the successful interaction of the assistant and the human remains open. That is why the most promising approach here today is based on biologically inspired cognitive architectures that allows us to achieve emotional intelligence in a machine based on semantic maps [17,21].

In this study we have developed an approach to designing virtual creative assistants of a composer. For this purpose, we have constructed, described, analyzed and used semantic maps of the musical fragment (loops) and generated hints to the user. The semantic map is the central element in our assistant, as it is in the eBICA cognitive architecture. We hope that this new technology that we presented here will be useful in many other areas. We expect that it can help us to develop other artificial creative assistants beyond the area of music creation.

Acknowledgements

This work was supported by the Russian Science Foundation Grant # 18-11-00336. The authors are grateful to all participating MEPhI students and to the MEPhI Academic Excellence Project for providing computing resources and facilities to perform experimental data processing.

References

[1] Petrushin, V.I. (2008). *Musical Psychology*. Moscow, "Tricsta", 400 p.

[2] Tatar, K. and Pasquier, P. (2019). Musical agents: A typology and state of the art towards Musical Metacreation. *Journal of New Music Research*, 48 (1): 56-105.

[3] Samsonovich, A.V., Goldin, R. F., and Ascoli, G. A. (2010). Toward a semantic general theory of everything. *Complexity*, 15 (4): 12-18. DOI: 10.1002/cplx.20293.

[4] Tyulin, Yu.N. (1986). *The Teaching of Harmony* (Russian). Moscow, "Musica", 480 p.

[5] Ostrovsky, A. (1987). *Theory of Music* (Russian). Moscow, "Musica".

[7] McFee, B., Raffel, C., Liang, D., Ellis, P.W. (2015). *librosa: Audio and Music Signal Analysis in Python.*

[8] Mel-cepstral coefficients (MFCC) and speech recognition. Online resource. Retrieved on 04.10.2019 from https://habr.com/ru/post/140828/

[9] Librosa.feature.mfcc. Online resource. Retrieved on 10.10.2019 from https://librosa.github.io/librosa/generated/librosa.feature.mfcc.html?highlight=mfcc#librosa.feature.mfcc

[10] DTW algorithm. Online resource. Retrieved on 10.10.2019 from https://github.com/pierre-rouanet/dtw

[11] Buravenkova, Y., Yakupov, R., Samsonovich, A.V., Stepanskaya, E. (2018). Toward a virtual composer assistant. Procedia Computer Science, 123, pp. 553-561. DOI: 10.1016/j.procs.2018.01.084

[12] Samsonovich, A.V., De Jong, K. A., and Kitsantas, A. (2009). The mental state formalism of GMU-BICA. International Journal of Machine Consciousness 1 (1): 111-130. DOI: 10.1142/S1793843009000116.

[13] Samsonovich, A.V. (2018). On semantic map as a key component in socially-emotional BICA. *Biologically Inspired Cognitive Architectures*, 23, pp. 1-6. DOI: 10.1016/j.bica.2017.12.002

[14] Samsonovich, A.V. (2013). Emotional biologically inspired cognitive architecture. *Biologically Inspired Cognitive Architectures,* 6: 109-125.

[15] Voznenko, T.I., Samsonovich, A.V., Gridnev, A.A., Petrova, A.I. (2018). The principle of implementing an assistant composer. Studies in Computational Intelligence, 799: 300-304. Cham, Switzerland: Springer. DOI: 10.1007/978-3-030-01328-8_36.

[16] Chella, A., Lebiere, C.L., Noelle, D.C., and Samsonovich, A. V. (2011). On a roadmap to biologically inspired cognitive agents. Frontiers in Artificial Intelligence and Applications, vol. 233, pp. 453-460. Amsterdam: IOS Press. DOI: 10.3233/978-1-60750-959-2-453.

[17] Thorogood, M., Fan, J., and Pasquier, P. (2019). A framework for computer-assisted sound design systems supported by modelling affective and perceptual properties of soundscape. *Journal of New Music Research*, 48 (3): 264–280.

[18] Samsonovich, A.V. (2020). Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. Cognitive Systems Research, 60: 57-76. DOI: 10.1016/j.cogsys.2019.12.002

[19] Samsonovich, A.V. & Ascoli, G.A. (2013). Augmenting weak semantic cognitive maps with an "abstractness" dimension. Computational Intelligence and Neuroscience, Volume 2013, Article ID 308176. DOI: 10.1155/2013/308176

[20] Samsonovich, A.V., De Jong, K.A., Kitsantas, A., Peters, E.E., Dabbagh, N., & Kalbfleisch, M.L. (2008). Cognitive constructor: An intelligent tutoring system based on a biologically inspired cognitive architecture (BICA). Frontiers in Artificial Intelligence and Applications. Vol. 171, Issue 1, pp. 311-325.

[21] Tikhomirova, D.V. Chubarov, A.A., & Samsonovich, A.V. (2020). Empirical and modeling study of emotional state dynamics in social videogame paradigms. Cognitive Systems Research, 60: 44-56. DOI: 10.1016/j.cogsys.2019.12.001