

А. В. КРАСНОСКУЛОВ

*Ростовская государственная консерватория им. С. В. Рахманинова, г. Ростов-на Дону, Россия
ORCID: 0000-0002-8468-8273, rostcons@yandex.ru*

ПРОЕКТ «Т|А»: ДУЭТ ЧЕЛОВЕКА И КОМПЬЮТЕРА

Цель проекта «Т|А» – исследовать потенциал компьютерной системы как «виртуального» музыканта-исполнителя («агента»), её возможности в создании и трансформации музыкального материала в художественных и процессуальных условиях, строящихся на перцепции и анализе звукового воплощения творческого замысла и внешнем выражении эмоций человека – «реального» исполнителя. Специально созданное для проекта программное обеспечение позволяет реализовать интерактивный дуэт «реального» и «виртуального» музыкантов, где последний воспринимает звучание партии исполнителя-человека и, используя генетический алгоритм, складывает в звуковой «ландшафт» собственной музыкальной партии. «Реальный» музыкант управляет процессом исполнения «агента» при помощи изменения выражаемых лицом эмоций. Каждая смена эмоционального состояния находит отражение в корректировке тембровых и реверберационных характеристик используемых компьютерной системой звуковых элементов и трансформации звучания всей «виртуальной» партии. Основываясь на научных трудах об особенностях корреляции звуков определённой высоты и/или тембра и вызываемых ими эмоциональных состояний, а также проведённых непосредственно в рамках проекта слуховых тестах, в статье приводится структура соотношения ключевых эмоций с частотными и пространственными параметрами звучания. На примере двух музыкальных композиций в работе даётся описание алгоритмических и творческих человеко-машинных процессов, а также обсуждаются проблемы и перспективы подобного интерактивного ансамбля.

Ключевые слова: человеко-машинное взаимодействие, распознавание эмоций, генетический алгоритм, интерактивная музыка.

ALEXEI V. KRASNOSKULOV

*Rostov State S. V. Rachmaninoff Conservatory, Rostov-on-Don, Russia
ORCID: 0000-0002-8468-8273, rostcons@yandex.ru*

THE “T|A” PROJECT: THE DUO OF THE HUMAN BEING AND THE COMPUTER

The aim of the “T|A” project is to research the potentials of the computer system as a “virtual” performing musician (“agent”), its possibilities in the creation and transformation of musical material in artistic and procedural conditions founded on the perception and analysis of sound manifestation of the artistic conception and the outer expression of emotions of the human being – the “real” performer. The software especially created for the project makes it possible to realize the interactive duo of the “real” and the “virtual” musicians, where the latter perceive the sound of the part of the human performer and by making use of the genetic algorithm assembles it into the sound “landscape” of its own musical part. The “real” musician directs the process of performance of the “agent” by means of change of the emotions expressed by the face. Each change of emotional state finds its reflection in the adjustment of the characteristics of timbre and reverberation utilized by the computer system of sound elements and the transformation of the sound of the entire “virtual” part. Basing itself on scholarly works about the peculiarities of the correlation of sounds of a particular pitch and/or timbre and the emotional states aroused by them, as well as the auditory tests carried out directly as part of the project, the article demonstrates a structure of correlation of the basic emotions with the frequency and spatial parameters of sound. On the example of two specific musical compositions this work gives a description of the algorithmic and creative processes of man and machine and also discusses the problems and perspectives of such kind of interactive ensemble.

Keywords: interaction between man and computer, discernment of emotions, genetic algorithm, interactive music.

правление музыкальным процессом в интерактивных дуэтах человека и компьютера, иначе говоря, «реального» и «виртуального» исполнителей, часто организуется при помощи различных жестов, сенсоров и т. п. (см.: [1]). Системы, использующие компьютерный анализ человеческой речи, выражений лица, поведенческих паттернов, становятся в последнее время более, чем когда-либо ранее, востребованными в различных сферах человеческой деятельности, в том числе – научной и творческой. Стремительный рост проектов, использующих нейронные сети как основу музыкального процесса «виртуального» исполнителя (или «агента»), связан, прежде всего, с ростом вычислительных мощностей и совершенствованием соответствующих алгоритмов. Распознавание эмоций, как и прочие новейшие программные разработки, даёт в руки «реального» исполнителя инструментарий, позволяющий ему координировать и кооперировать совместное человеко-машинное взаимодействие [9]. Используя потенциал распознавания, предоставляемый программным интерфейсом (Emotion API) «облачного» сервиса Microsoft Cognitive Services¹, а также исследованные ранее автором статьи возможности генетических алгоритмов в музыкальном контексте [2; 3], проект «Т|А» сфокусирован на изучении интерактивных композиций в ситуации детерминированного контроля над «виртуальным» исполнителем со стороны его ансамблевого партнёра – «реального» музыканта.

Ключевая особенность генеративного процесса в проекте – контроль над тембральными и реверберационными параметрами звуковых элементов, используемых «виртуальным» исполнителем в процессе создания музыкальной композиции. Изменение тембральных характеристик основано на взаимосвязях, которые существуют между различными эмоциональными состояниями и звуковысотными и динамическими свойствами конкретных звуков. Согласно ряду исследований [4; 5; 6; 10], определённые эмоции вызываются звуками фиксированного частотного и динамического диапазонов. В нашем случае важно, что подобная корреляция также существует между эмоциями и темброво сложными музыкальными звуками, в том числе темперированного строя. Упомянутые труды рассматривают десять эмоциональных «категорий» (соответственно: «счастливый»,

«печальный», «героический», «страшный», «комичный», «застенчивый», «романтический», «загадочный», «гневный», «спокойный»), однако для настоящего исследования актуальными оказались лишь четыре из них, поскольку только они совпадают с эмоциями, список которых формируется с использованием Emotion API (соответственно: «счастье», «печаль», «гнев» и «покой»). Для оставшихся четырёх эмоций («презрение», «страх», «отвращение» и «удивление») был разработан специальный слуховой тест, в котором приняла участие группа экспертов – профессиональных музыкантов – преподавателей и студентов РГК им. С. В. Рахманинова.

Для теста (аналогично тестам: [5]) были записаны звуки фортепиано (сэмплированного, с использованием звуковой библиотеки Native Instruments Akoustik Piano) в диапазоне от C контроктавы (C_3) до c пятой октавы (c^5). Слушатели выбирали, каким эмоциям соответствуют, на их взгляд, звуки той или иной высоты – в итоге определяя частотные диапазоны, наиболее ярко репрезентирующие соответствующие эмоциональные состояния. Изучение взаимосвязи между эмоцией и звуком, выраженное в смене его реверберационных характеристик, в контексте проблематики проекта также потребовало дополнительного исследования. Для второго слухового теста были подготовлены 3 отличных друг от друга звуковых элемента, каждый из которых воспроизводился многократно с использованием 24 заранее сформированных реверберационных пресетов (всего в тесте, таким образом, звучали 72 музыкальных фрагмента). В итоге для используемых в проекте 8 эмоций их взаимосвязи с частотными и реверберационными характеристиками звучания получились следующими:

Эмоция	Звуковысотность	Реверберация
Счастье	$c^2 - c^4$	Аудитория
Печаль	$C_3 - c^5$	Большой концертный зал
Презрение	$C_3 - c^3$	Ангар
Страх	$C_3 - H_3; c^4 - h^4$	Стадион
Отвращение	$c^1 - c^3$	Отсутствует
Удивление	$c^2 - c^5$	Гулкий коридор
Гнев	$C_3 - C_1$	Пустая комната
Покой	$c^1 - c^5$	Малый концертный зал

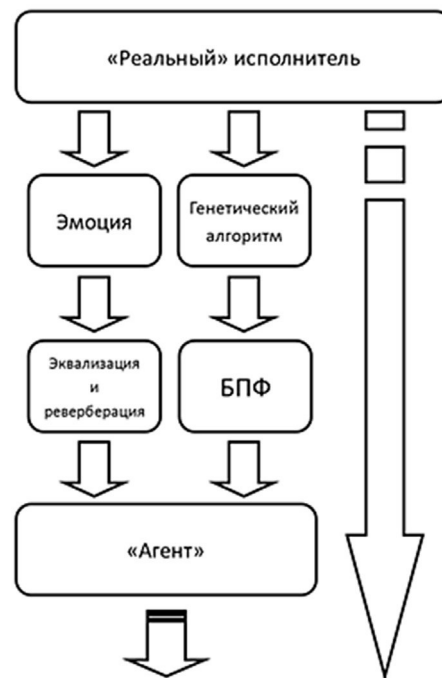
В процессе исполнения музыкальной композиции веб-камерой, направленной на лицо «реального» исполнителя, через равные промежутки времени делается «скриншот», который отправляется через интернет в «центр распознавания эмоций». Параллельно звучание «реального» исполнителя записывается с микрофона как звуковой поток и в виде последовательности MIDI-сообщений.

Захваченный микрофоном звук подвергается анализу с использованием быстрых преобразований Фурье (БПФ): полученные в результате значения амплитуды становятся фитнес-функцией генетического алгоритма [2; 3; 7]. Как следствие, исполнение «агента» никогда в точности не повторяет партию «реального» исполнителя, но непрерывно стремится «сымитировать» её. Кроме того, в рамках проекта предполагается, что все используемые «виртуальным» исполнителем звуковые элементы нейтральны до тех пор, пока не будут помещены в какой-либо эмоциональный контекст.

Распознанное в «облаке» изображение возвращается в виде списка из 8 эмоций, каждой из которых присвоено значение её «вероятности» (или «веса»). Таким образом, когда «реальный» исполнитель изменяет (осознанно или непроизвольно) выражение своего лица и сервис возвращает результат анализа очередного «скриншота» – «веса» всех эмоций также изменяются. Сразу же происходит смена звуковых паттернов (групп звуковых элементов, заранее определённых «реальным» исполнителем) в партии «агента», включая трансформацию каждого звука низкочастотным и высокочастотным фильтрами. Тип реверберации также меняется – глобально во всей партии «виртуального» исполнителя (см. схему 1).

Две музыкальные композиции – «Дукка»² и «Нгаурухое»³ – наглядно демонстрируют всю систему в действии. Для обеих композиций было создано 4 группы звуковых элементов: каждая группа содержит 84 звука одного определённого тембра, расположенных по полутонам хроматического звукоряда (C_3 – h^4). Во второй композиции, кроме того,

Схема 1



элементы одной из групп представляют собой не протяжённые «фразы», а краткие «мотивы», что даёт возможность генетическому алгоритму формировать из них достаточно сложные ритмические структуры. Звуковыми элементами различных групп (количество последних определяется предварительными установками) равномерно заполнялась хроматическая шкала от C контроктавы до h четвёртой октавы (всего 84 звука разной высоты).

Как отмечалось выше, звучание партии «реального» исполнителя после спектрального анализа становится фитнес-функцией генетического алгоритма: таким образом, результат очередной «эпохи» корректирует громкость каждого из звуков шкалы. При этом, в рассматриваемых композициях количество таких «эпох», как и вероятность «скрещиваний» и «мутаций», различно:

Композиция	Количество «эпох» (в секунду)	Вероятность «скрещивания» (в процентах)	Вероятность «мутации» (в процентах)
«Дукка»	20	50	70
«Нгаурухое»	6	30	80

Для установки реверберационных характеристик звучания в партии «агента» из списка эмоций выбиралась та, которая в текущий момент обладала наибольшим «весом». При этом, если в результате

распознавания наиболее «вероятным» определялось эмоциональное состояние «покоя», то такое значение игнорировалось, оставляя активным предыдущие «веса» эмоций. Это позволяло «реальному» исполнителю продлевать любое эмоциональное состояние – и, соответственно, выбранные музыкальные структуры, – простым сохранением «нейтрального» выражения лица.

Проект «Т|А» находится в процессе своего развития. Его дальнейший прогресс заключается в расширении возможностей акустической записи композиций в реальном времени и, главное, в переходе от «двух» к «ансамблю», в увеличении количества исполнителей – «реальных» и «виртуальных». На сегодняшний день основной проблемой остаётся сама система распознавания эмоций. Прежде всего, распознавание занимает – по меркам быстротечного музыкального процесса – слишком много времени, что связано как со скоростью работы

алгоритмов и самих нейронных сетей, так и латентностью сетевого обмена данными. В результате, в настоящий момент интервал между «скриншотом» лица исполнителя и получением итогового списка «весов» эмоций может достигать пяти и более секунд, что не позволяет «реальному» исполнителю максимально гибко управлять партией «агента». Не меньшие вопросы ставит значительная неточность распознавания: многие выражения лица, отчётливо понятные человеку, трактуются компьютером как «нейтральные» (эмоции «покоя»), что заставляет «реального» исполнителя гипертрофировать эмоции, излишне напрягая мышцы лица и даже гримасничая. Решение данного вопроса видится в создании и обучении собственной нейронной сети – не такой всеобъемлющей, как в используемом в данный момент сервисе, однако сфокусированной на точности и детальности выражений эмоций конкретных людей.

ПРИМЕЧАНИЯ

¹ Microsoft Cognitive Services: <https://www.microsoft.com/cognitive-services>.

² Композиция «Дукка»: <http://www.soundworlds.net/media/BigDukkaRiver.wav>.

³ Композиция «Hraupyrhое»: <http://www.soundworlds.net/media/Ngauruhoe.wav>.

ЛИТЕРАТУРА

1. Краснокулов А. В. Ансамблевое музицирование в цифровом мире // Музыкальное искусство в современном социуме: сб. науч. ст. Ростов н/Д.: Изд-во РГК, 2014. С. 278–288.
2. Краснокулов А. В. Эволюционное моделирование музыки: принципы, подходы, проблемы // Южно-Российский музыкальный альманах. 2016. № 1 (22). С. 24–30.
3. Краснокулов А. В. Эволюционные вычисления в интерактивной музыке // Вестник музыкальной науки. 2016. № 2 (12). С. 54–59.
4. Chau C.J., Mo R., Horner A. The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Sounds // Journal of the Audio Engineering Society. 2014. Vol. 62, no. 10, pp. 663–675.
5. Chau C.J., Mo R., Horner A. The Emotional Characteristics of Piano Sounds with Different Pitch and Dynamics // Journal of the Audio Engineering Society. 2016. Vol. 64, no. 11, pp. 918–932.
6. Chau C.J., Wu B., Horner A. Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones // Proceedings of the 40th International Computer Music Conference (ICMC). Michigan, 2014, pp. 982–989.
7. Evolutionary Computer Music. Miranda, Eduardo Reck; Biles, John Al (Eds.) London: Springer, 2007. 249 p.
8. Mo R., Wu B., Horner A. The Effects of Reverberation on the Emotional Characteristics of Musical Instruments // Journal of the Audio Engineering Society. 2015. Vol. 63, no. 12, pp. 966–979.
9. Winters R. M., Hattwick I., Wanderley M. M. Emotional Data in Music Performance: Two Audio Environments for the Emotional Imaging Composer // Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. URL: <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/41617/R.%20Michael%20Winters%20-%20Emotional%20Data%20in%20Music%20Performance%20-%20Two%20Audio%20Environments%20for%20the%20Emotional%20Imaging%20Composer.pdf?sequence=1>.

10. Wu B., Horner A., Lee C. Musical Timbre and Emotion: The Identification of Salient Timbral Features in Sustained Musical Instrument Tones Equalized in Attack Time and Spectral Centroid // *Proceedings of the 40th International Computer Music Conference (ICMC)*. Michigan, 2014, pp. 928–934.

Об авторе:

Краснокулов Алексей Владимирович, кандидат искусствоведения, заведующий кафедрой музыкальной звукорежиссуры и информационных технологий, профессор кафедры специального фортепиано, Ростовская государственная консерватория им. С. В. Рахманинова (344002, Ростов-на-Дону, Россия), **ORCID: 0000-0002-8468-8273**, rostcons@yandex.ru

REFERENCES

1. Krasnoskulov A. V. *Ansamblevoe muzitsirovanie v tsifrovom mire* [Ensemble Performance in a Digital World]. *Muzikal'noe iskusstvo v sovremennom sotsiуме: sb. nauch. st.* [The Art of Music in the Modern Society: A Compilation of Scholarly Articles]. Rostov-on-Don: Publishing House of the Rostov Conservatory, 2014, pp. 278–288.
2. Krasnoskulov A. V. *Evolyutsionnoe modelirovanie muzyki: printsipy, podkhody, problemy* [The Evolutional Modeling of Music: Principles, Approaches, Issues]. *Yuzhno-Rossiyskiy muzikal'nyy al'manakh* [South-Russian Musical Anthology]. 2016. No. 1 (22), pp. 24–30.
3. Krasnoskulov A. V. *Evolyutsionnye vychisleniya v interaktivnoy muzyke* [Evolutionary Computations in Interactive Music]. *Vestnik muzikal'noy nauki* [Herald of Musical Scholarship]. 2016. No. 2 (12), pp. 54–59.
4. Chau C. J., Mo R., Horner A. The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Sounds. *Journal of the Audio Engineering Society*. 2014. Vol. 62, no. 10, pp. 663–675.
5. Chau C. J., Mo R., Horner A. The Emotional Characteristics of Piano Sounds with Different Pitch and Dynamics. *Journal of the Audio Engineering Society*. 2016. Vol. 64, no. 11, pp. 918–932.
6. Chau C. J., Wu B., Horner A. Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones. *Proceedings of the 40th International Computer Music Conference (ICMC)*. Michigan, 2014, pp. 982–989.
7. *Evolutionary Computer Music*. Miranda, Eduardo Reck; Biles, John Al (Eds.) London: Springer, 2007. 249 p.
8. Mo R., Wu B., Horner A. The Effects of Reverberation on the Emotional Characteristics of Musical Instruments. *Journal of the Audio Engineering Society*. 2015. Vol. 63, no. 12, pp. 966–979.
9. Winters R. M., Hattwick I., Wanderley M. M. Emotional Data in Music Performance: Two Audio Environments for the Emotional Imaging Composer. *Proceedings of the 3rd International Conference on Music & Emotion (ICME 3)*. Jyväskylä, Finland, 11th–15th June 2013. URL: <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/41617/R.%20Michael%20Winters%20-%20Emotional%20Data%20in%20Music%20Performance%20-%20Two%20Audio%20Environments%20for%20the%20Emotional%20Imaging%20Composer.pdf?sequence=1>.
10. Wu B., Horner A., Lee C. Musical Timbre and Emotion: The Identification of Salient Timbral Features in Sustained Musical Instrument Tones Equalized in Attack Time and Spectral Centroid. *Proceedings of the 40th International Computer Music Conference (ICMC)*. Michigan, 2014, pp. 928–934.

About the author:

Alexei V. Krasnoskulov, Ph.D. (Arts), Head at the Department of Sound Engineering and Informational Technologies, Professor at the Piano Major Department, Rostov State S. V. Rachmaninoff Conservatory (344002, Rostov-on-Don, Russia), **ORCID: 0000-0002-8468-8273**, rostcons@yandex.ru