Proceedings of the Fifteenth AAAI Conference on Artificial
Intelligence and Interactive Digital Entertainment (AIIDE-19)

Adaptive Game Soundtrack Generation Based on Music Transcription

Munir Makhmutov
1, Universitetskaya Str., Innopolis, 420500, Russia
m.makhmutov@innopolis.ru

Abstract

This research aims to bring new prospects for adaptive game soundtrack generation, which theoretically can increase the involvement of players into the gaming process. To investigate the problem of adaptive soundtrack generation, human-centered evaluation has to be done. Furthermore, an important part of this research is a collaborative soundtrack generation with the participation of composers and artificial intelligence, which is able to decrease expenses of game developing companies. For soundtrack customization, this work suggests a joint generation of music by human and artificial intelligence and its further intuitive modification during the game based on in-game actions.

It is evident that intuitiveness in games helps players during their gameplay experience. Intuitiveness is the ability to understand things without help or guidelines. Consistency between the players' actions and their expectations about changes inside the game makes games intuitive. Intuitiveness can be achieved by the application of a good design avoiding "Norman's doors" described in "The Design of Everyday Things" (Norman 2013). This work focuses on soundtrack intuitiveness in games. Custom changes in games make them more suitable for player needs. One of such individual approaches is the use of music modified based on player's actions. Allowing users to choose their own melody for a game can be tricky because the audio part of the game has a big impact on game perception. Therefore, the usage of fully custom songs chosen by a player can spoil the gameplay experience. Customization of a game soundtrack can be realized in a lighter and smoother way. The player can prepare or choose a melody, which can be modified by algorithms for the game settings in an intuitive manner. Also, the choice of initial music can be done by game developers in case of necessity. Therefore, it is necessary to understand what makes the soundtrack intuitive. Recent research shows that the feature allowing players to choose own music for car radio in Grand Theft Auto series slightly increased the immersion into the game (Williams 2018). Theoretically, intuitively modified own music can increase the immersion even more.

The motivation for this research is to provide an opportunity for musicians for partial game soundtrack generation. Most likely, it will bring a new experience to those players, because they will feel that the game was partially created by them, and it will make the game more involving. It is obvious that most of the players are far from music composing. That is why the game should check the expediency to use given input melodies and reject the melody in case of necessity. Even good input melody may require some rethinking and additional ideas, which can be delegated to artificial intelligence. However, the algorithm should adhere to the author's main musical idea.

Methodology

This work considers the music transcription, music generation and adaptive real-time music modification inside the video game. General view of a process is shown in figure 1.

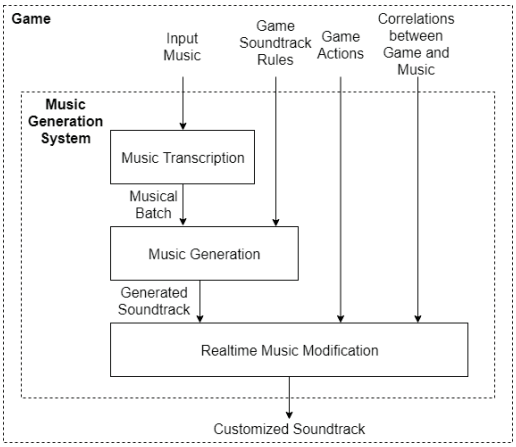

Figure 1: Soundtrack Generation

The first part is uploading a music record inside the game. To make it possible to be modified the recorded music has to be transcribed into music notation. Automatic music transcription allows increasing the speed of notating the music, which is helpful for musicians in the process of songwriting. The current state of automatic music transcription allows re-

trieving monophonic melodies with high accuracy (Sterian, Simoni, and Wakefield 1999). Polyphonic music transcription is still not very accurate at the time of writing this paper (Nakamura et al. 2018), (Sigtia, Benetos, and Dixon 2016), (Benetos and Dixon 2012). For higher accuracy monophonic music transcription is chosen for this research.

After the melody transcription, artificial intelligence should analyze the input and decide if it is possible to generate the rest part of music based on it. In case of positive decision evolutionary algorithm is used for soundtrack generation. This algorithm based on game settings should modify initial melody. The game settings can describe the tempo, the mood and many other aspects of music. And only then other musical batches should be generated by an evolutionary algorithm. This process can include several evolutionary algorithms: for initial melody, for each new batch and for the music in general. Fitness function for music generator is a set of mathematical formulas, which is different for all instruments. Evolutionary algorithms were used for game soundtrack generation in MetaCompose (Scirea et al. 2016), (Scirea et al. 2018) and resulting aesthetically pleasing music proves that it is a promising approach.

The last part is the game during which the soundtrack generated on the previous step has to be intuitively modified in real-time. For the understanding of correlations between the game actions and music features the human-centered evaluation has to be done. Different combinations of game and music factors can be tested by playtesting of serious video games. The games developed for this research are serious games as they have other goals other than pure entertainment. According to (Djaouti, Alvarez, and Jessel 2011), serious game designers use people's interest in video games to capture their attention for a variety of purposes that go beyond pure entertainment. In this paper, authors describe the so-called G/P/S model, that allows classifying serious games. The game combinations which affect positively on most of the players can be considered as combinations with intuitive correlations. Any player's action or game surrounding can be considered as a game factor excluding the music in this research. Theoretically, after all these steps the game will be able to produce the customized soundtrack.

Achieved Results

A monophonic music transcription system has been developed as an initial part of the research (Figure 2). The music that has been recorded acts as input for the "Music Analysis" block. which is then trimmed in order to detect tempo, time signature, bar amount, notes and their duration. Finally, and accordingly to the processed data, the tool generates a MIDI file for further music generation. Based on this part of research the paper is written (Makhmutov et al. 2017).

Currently for finding intuitive correlations between game and soundtrack two small serious video games are developed. The first game considers correlations between game factors such as surrounding colours, common types of bonuses (increasing and decreasing of speed, multiplication of scores, etc.), number of health points and between the tempo of a soundtrack. Hypothetically these game factors can have a correlation with soundtrack tempo.

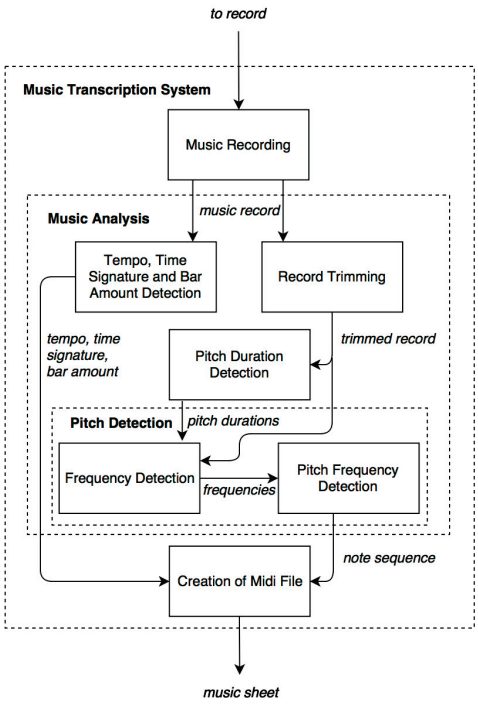

Figure 2: Music Transcription (Makhmutov et al. 2017)

According to (Spenwyn, Barrett, and Griffiths 2010) gamblers risk more during fast tempo songs with red light surroundings. In the next research (Stark, Saunders, and Wookey 1982) it was demonstrated that gamblers within a red light environment gambled money more often and selected riskier odds than gamblers under blue light. Most likely, the red light illumination with fast music tempo has to make players more risky and blue light illumination with slower music tempo has to make them less risky.

It is also proved that the tempo of music listened by drivers affects their speed (Brodsky 2001). Faster tempo leads to faster driving, and lower tempo leads to slower driving. This research was made using PC-controlled simulations. Therefore, our assumption is that the speed of moving in games (or speed bonuses) can have a strong correlation with soundtrack tempo, which also has to be tested.

The playtesting process has been started and the results so far demonstrate that on average people agree that some soundtrack changes look intuitive. Based on the results of experiments with the first game it was decided to design the second game to test more dependencies between game and music such as key, instruments, volume, note duration, etc.

Conclusions

According to the plan of research the first part related to music transcription is finalized and the third part related to real-time soundtrack modification is in progress. Implementation of evolutionary algorithms for music generation is in progress. Furthermore, the accuracy of music transcription is in the progress of evaluation through FMEA.

References

Benetos, E., and Dixon, S. 2012. A shift-invariant latent variable model for automatic music transcription. *Computer Music Journal* 36(4):81–94.

Brodsky, W. 2001. The effects of music tempo on simulated driving performance and vehicular control. *Transportation research part F: traffic psychology and behaviour* 4(4):219–241.

Djaouti, D.; Alvarez, J.; and Jessel, J.-P. 2011. Classifying serious games: the g/p/s model. In *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*. IGI Global. 118–136.

Makhmutov, M.; Brown, J. A.; Mazzara, M.; and Johard, L. 2017. Momos-mt: mobile monophonic system for music transcription: sheet music generation on mobile devices. In *Proceedings of the Symposium on Applied Computing*, 543–549. ACM.

Nakamura, E.; Benetos, E.; Yoshii, K.; and Dixon, S. 2018. Towards complete polyphonic music transcription: Integrating multi-pitch detection and rhythm quantization. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 101–105. IEEE.

Norman, D. 2013. *The design of everyday things: Revised and expanded edition*. Constellation.

Scirea, M.; Togelius, J.; Eklund, P.; and Risi, S. 2016. Metacompose: A compositional evolutionary music composer. In *International Conference on Computational Intelligence in Music, Sound, Art and Design*, 202–217. Springer.

Scirea, M.; Eklund, P.; Togelius, J.; and Risi, S. 2018. Evolving in-game mood-expressive music with metacompose. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, 8. ACM.

Sigtia, S.; Benetos, E.; and Dixon, S. 2016. An end-to-end neural network for polyphonic piano music transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24(5):927–939.

Spenwyn, J.; Barrett, D. J.; and Griffiths, M. D. 2010. The role of light and music in gambling behaviour: An empirical pilot study. *International Journal of Mental Health and Addiction* 8(1):107–118.

Stark, G.; Saunders, D.; and Wookey, P. 1982. Differential effects of red and blue coloured lighting on gambling behaviour. *Current Psychology* 2(1-3):95–99.

Sterian, A.; Simoni, M.; and Wakefield, G. H. 1999. Model-based musical transcription. In *ICMC*. Citeseer.

Williams, D. 2018. Repurposing music according to individual preferences for personalized soundtracks. In *Emotion in Video Game Soundtracking*. Springer. 105–114.