Towards an Emotion-Driven Adaptive System for Video Game Music

Manuel López Ibáñez[1], Nahum Álvarez[2], and Federico Peinado[1]

[1] Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
c/Profesor José García Santesmases 9, 28040 Madrid, Spain
manuel.lopez.ibanez@ucm.es, email@federicopeinado.com
[2] Artificial Intelligence Research Center
National Institute of Advanced Industrial Science and Technology
2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan
nahum.alvarez@aist.go.jp

Abstract Perfectly adaptive music is a long dreamed goal for game audio designers. Although numerous systems have been developed in both academic and industrial contexts, currently there is no unified method for producing this type of content — not even an agreement on input variables, output variations or assessment criteria. Our research aims to create an audio system for video games that improves the experience by adapting environmental music to emotions associated with the ongoing narrative. This system combines short audio tracks of pre-designed music in real time, using player behavior and emerging feelings as main cues. In this paper, we identify some of the issues that dynamic music faces through the study of current adaptive and procedural techniques, and describe how our sound system architecture attempts to solve them. In the end, we claim that a sonification method like the proposed one could improve player engagement by adapting music to each game situation, driven by a solid focus on emotional storytelling.

Keywords: Interactive Digital Storytelling · Affective Computing · Player Experience · Sound Design · Procedural Content Generation

1 Introduction

Designing sound and music for video games has always been a task with an intrinsic difficulty. As in many other interactive experiences, sound designers and musicians cannot predict all possible outcomes, so video game music should be *dynamic*, that is, not designed with a fixed timeline in mind but aimed towards a large set of possible player actions, reactions and in-game situations.

Thus, it makes sense to create music that is not strictly responsive (as in a rhythm game) but "automatically adaptive" in some form, looking for an improved experience for the player, where ambient sound around him would fit his circumstances. Such capability moves audio production tasks closer to software system design, not being an isolated task for the musician anymore, but having a much tighter relationship with the overall game architecture.

So as to consider most issues and difficulties that current audio systems need to tackle, in the next section we review scientific literature on adaptive and procedural sound design, as well as on the role of emotions in player experience.

In section 3 we identify the main issues of dynamic and procedural music composition and the solutions proposed until now. Section 4 contains a description of our system architecture, which aims to improve and unify current methods for adapting music in narrative video games, taking into account basic emotional changes of the story as a reference. This is illustrated with a dialogue between an in-game character and the player.

The last section outlines preliminary conclusions about improvements on player engagement that could be achieved with this emotion-driven adaptive system for dynamic music.

2 Adaptive music and basic emotions

Recently, an increased interest in developing new models for the understanding of auditory perception has been developed. For example, Asutay et. al. [1] provide an evaluation system which considers the emotional meaning attached to a sound as an important factor when identifying audio, along with signal characteristics.

Inger Ekman [2] also states that designing sound for video games has to take into account the non-linear nature of the media, so as to improve the emotional response achieved during a gaming session. This author considers there is a "readiness to act" at the core of every emotion, and consequently, emotions prepare a player for certain types of actions. Thus, sound in video games should be centered on guiding these possible actions and adapting to them, instead of following a linearity of any kind. This "readiness to act" can also add coherence to emotions felt during gameplay.

One of our claims is that dynamic music design for games has to consider the emotional nature of audio, so a need arises: to have a simple taxonomy which allows to classify basic emotions felt by players during a standard gaming session. Paul Ekman designed a taxonomy including 5 basic emotions [3], widely used today. "Surprise", though not considered a basic emotion, is added as the sixth element of this list, as Ekman considers it perfectly describes a sudden change in feelings:

- Anger: Wrath, ire. Generates displeasure and inclination to belligerence.
- Disgust: Repugnance, aversion. Implies distaste for something.
- Fear: Urgency to avoid a certain situation which can inflict pain, damage, anxiety or simple aversion.
- Joy: Being delighted, glad, fortunate or pleased by something or someone.
- Sadness: Sorrow, unhappiness and grief. The feeling associated to dark, negative or bad things.

– Surprise: A sudden and unexpected feeling. Associated to discovery or learning without warning, whether the discovered piece of information is negative or positive.

Although Paul Ekman himself and other authors have identified many more subtle and complex emotions, the simplicity and universality of this list makes it suitable for our needs, as it would make the process of sorting sounds according to emotions and desirable player reactions a task achievable in most game scenarios.

Yanagisawa and Murakami [4] even propose a system to quantify emotional quality ("*kansei* quality") of a sound, by using certain parameters like "strong", "silent", "dull" or "hard" in a survey taken by neutral subjects. Though being applied to physical product design in the mentioned article, this method could be of use in video games if utilised properly. For example, we can increase the level of fear induced by an atonal melody by raising its intensity.

Besides, Jørgensen [5] points out that game audio has a dual role: on one hand, it serves the purpose of supporting the general feeling of an environment ("ambient music"); on the other, it gives players pieces of information about the game. A *layered* audio design system may then be appropriate, because it would make easy to select a set of sound tracks that fulfill both the requirements of the environment and the characters.

As will be explained in the next section, numerous music design systems have tried to adjust to events happening in-game and to player emotions. However, most initiatives are merely environmental — they depend on a change of scenery or a cinematic sequence to work properly. This usually happens due to two main reasons: lack of human resources or insufficient computational capacity. Our proposal aims to slightly decrease human intervention while greatly decreasing computational needs when composing layered, fully adaptive music for video games.

3 Problems of current dynamic music systems

The interactive entertainment industry offers many cases of *ad hoc* systems for generating dynamic music. Modern examples include the vertical re-orchestration of Red Dead Redemption, which uses a large pool of "music stems", all recorded in A minor at 130 beats per minute in order to be easily combined. Tomb Raider Legend developed the concept of micro-scoring for adapting its more than 180 minutes of music to different game characters and environments; Halo[3] series combines vertical and horizontal techniques with some randomization in its proprietary audio engine, etc.

Though dynamic audio design is a valued asset in this industry, as all previous examples show, most of the time it is achieved in a traditional, non-adaptive manner. Dynamic audio design techniques are more expensive, and have a practical limitation: because it is not possible to predict every potential in-game situation, an undetermined amount of "sound holes" or inconsistencies may appear

[3] https://www.halowaypoint.com

during the experience unless a completely adaptive approach is taken. Classic dynamic music is usually associated with seamlessly changing between different audio environments or scripted situations [6]. What we propose, however, is to fully adapt to player actions, even when there is not a way to predict what could happen during gameplay.

On the other hand, Collins [7] states that the major handicap of procedural music in current video games is available technology. As most of these methods produce a lower quality result in terms of audio fidelity, producers are worried about reducing the level of player immersion. Besides, system performance lowers considerably when using procedurally generated sound and music, which becomes a problem when playing on low-end equipment, like game consoles or mobile devices. Even popular game engines like Unreal Engine [4] understand procedural audio as an experimental, uncommon feature, which offers low quality results.

Jewell, Nixon & Prügel-Bennett [8] try to improve the complexity and general quality of procedurally-generated music by adding semantic information to genetic algorithms, achieving better results. Nevertheless, this technique cannot be applied to real-time gaming, due to their processing cost and the level of adjustment needed to achieve a sufficiently good piece of music.

Luhtala et. al. [9] find another interesting solution to the problem of audio quality in procedural music by adding a variety of synth modules to their system, in a way that makes it possible to process "raw" generated data by using Virtual Studio Technology (VST) instruments, among other techniques. Again, the same problem emerges from this approach: the processing power required for utilising high quality VSTs in multiple tracks while generating dynamic music is such that a commercial game with state-of-the-art graphics could not run, even on a current high-end system.

Another interesting issue was noted by Böttcher et. al. [10], who found a clear lack of attention from most users to sound quality and music composition when they are playing. We also observed this effect in a previous experiment with users of a Virtual Reality demo [11]. Emotions arise as a side effect of music and, as stated by Eladhari [12], sound does not have a protagonist role in a video game unless it conveys important, gameplay-related information.

Lastly, Livingstone & Brown [13] stand up for the need of a sound system that allows to dynamically adjust music in a video game depending on player emotions. This would bring a "cinematic" feel to video game soundtracks, thus improving the utility perception of music in video game production. We followed this suggestion when proposing our adaptive system for dynamic music.

4 Proposal of an emotion-driven adaptive system

Our take on the problems mentioned above is materialized on a system which allows to create audio atmospheres which adapt to the emotions arisen in a

[4] https://www.unrealengine.com/what-is-unreal-engine-4

narrative game. This system does not generate music from scratch, due to the limitations found with procedural techniques; instead, it combines pre-designed fragments of music (in the form of short sound tracks) to achieve the desired effect, avoiding the problem of low-fidelity sound. In the current version, the system is meant to work with in-game dialogues that allow the player to confront different characters and emotional situations, choosing from a variety of possible responses.

The proposed architecture (see figure 1) works as follows:

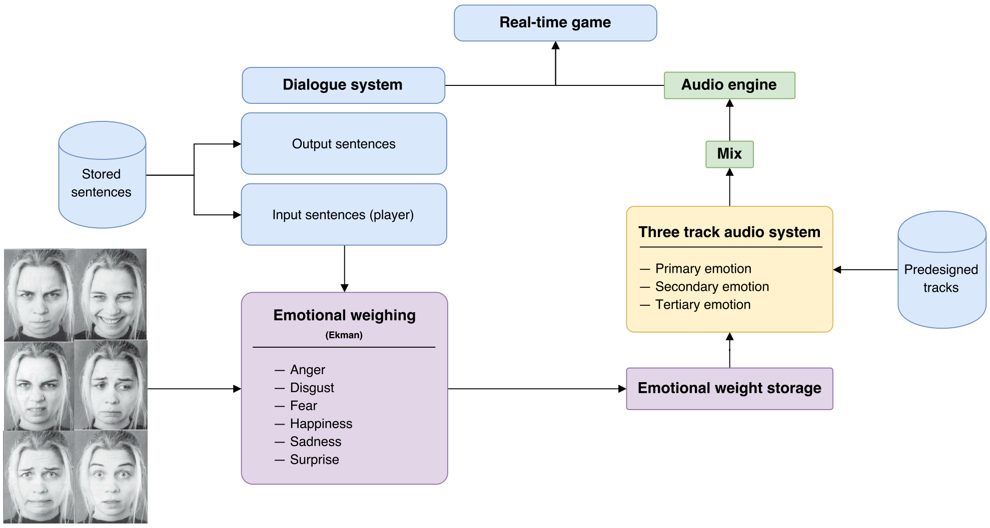

Figure 1. Architecture of our emotion-driven adaptive music system. Photos from Bennett & Šabanović [14].

– The game engine includes a dialogue system which the user can interact with in two ways: receiving messages from a non-player character (NPC) (*text output*) and choosing answers to those messages (*input choice*) from a series of options.
– The dialogue system takes information from a database of tree-form short dialogues. These dialogues are designed by a human, and try to induce a variety of emotional responses in the player. In our initial configuration, the player can choose from 3 different answers each time a text output is shown on screen, representing the expression of a feeling.
– Every possible answer has an "emotional weighing", containing a real value in a range from 0 to 1 for one of the 6 basic emotions of Ekman [3]. For example, receiving a declaration of war will have an anger value of 1, and watching a butcher work will have a disgust value of 0.5.
– The "emotional weights" of characters' messages and player choices are stored, so that the system can remember the emotional context of a playing

session. With each addition, the stored basic emotion variables with a higher value in the last seconds of gameplay are selected. Our intention is to select a maximum of 3 concurrent emotions (that correspond to 3 different music tracks) depending on their accumulated weight.

- We establish a maximum of 3 tracks for two reasons: Sometimes, emotions can overlap (e. g. anger and disgust generate loathing), which means at least 2 simultaneous tracks are needed to create such a complex atmosphere. Besides, we add a third track so as to be able to include remembrance in the experience. If, for example, our chosen tracks are prioritized in the following order: anger, disgust and happiness, we can produce a sense of loathing (anger and disgust) while hinting at a lost happiness.
- Due to this decision, currently we work with a 3-track audio system managing fragments of music each delta second. Once in that period of time, the emotion variables are read and a new piece of music related to those feelings starts playing if necessary. The resulting music should represent the selected basic emotions. If no emotional answers are given, the previous loop will keep playing indefinitely.
- All tracks are chosen from a database of predesigned fragments of music, created by a human composer, sharing a tempo of 110 beats per minute. All of them will be loops, with lengths that range from 5 seconds to 15. The system is not designed to allow a sequence of quick changes in ambient music (less than 5 seconds each), due to the minimum duration of the loops used.
- Every time a new track is selected and played, it adjusts to a fixed tempo *grid*, skipping up to a beat if necessary. Once a combination of two or more tracks is playing, they are mixed in the audio engine. Intensity (volume) is established depending on the values of emotional weight, for a normalized intensity of 1. Primary emotions have a normalized (non-logarithmic) intensity of 0.5, followed by secondary (0.3) and terciary (0.2) emotions.
- The resulting ambient music plays through the game engine's audio system. This makes it possible to spatialise the mix or to add an extra layer of effects.

5 Conclusions

Considering the disparity of approaches given to dynamic music generation in commercial video games, the proposed system architecture aims to constitute a versatile option for having emotion-driven ambient music without the low quality sound and high processing cost of a more complex, procedural system. Due to its architecture, it can work in every commercial game engine in the market, and can potentially adapt to every type of game, as long as there are emotional interactions in it.

Besides, the proposed audio managing technique acts as an "assistant" instead of as a "proxy" for musicians. This means the final result can be shaped by both human creativity and computational adaptability, resulting in sound tracks which are natural to compose and flexible to recombine at the same time. Although this means our system requires expert knowledge, it gives musicians the capacity to create realistic audio tracks for video games, while maintaining instantaneous adaptability.

For further research and testing of the system, we are currently building a prototype and a testbed game in Unity[5]. This prototype will be played by real users while taking both quantitative and qualitative measures, such as: live recording of play sessions, Self-Assessment Manikin Test (SAM) for measuring emotions [15], etc.

Once the system works as intended, and after adjusting it according to the results achieved in the experimental phase, a plugin for popular game engines as Unity or Unreal can be developed, so as to allow musicians and sound designers to use it with ease.

Acknowledgements

This article was funded by the Complutense University of Madrid (grant CT27/16-CT28/16 for predoctoral research), in collaboration with Santander Bank and NIL research group.

References

1. E. Asutay, D. Västfjäll, A. Tajadura-Jiménez, A. Genell, P. Bergman, and M. Kleiner, "Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design," *AES: Journal of the Audio Engineering Society*, vol. 60, no. 1-2, pp. 21–28, 2012.
2. I. Ekman, "Psychologically motivated techniques for emotional sound in computer games," in *AudioMostly 2008*, pp. 20–26, 2008.
3. P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3, pp. 169–200, 1992.
4. H. Yanagisawa, T. Murakami, S. Noguchi, K. Ohtomi, and R. Hosaka, "Quantification method of diverse kansei quality for emotional design: Application of product sound design," in *ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 461–470, 2007.
5. K. Jørgensen, *"What are these grunts and growls over there?" Computer game audio and player action.* PhD thesis, 2007.
6. W. Strank, "The Legacy of iMuse: Interactive Video Game Music in the 1990s," in *Music and Game: Perspectives on a popular alliance*, pp. 81–91, Wiesbaden: Springer Fachmedien Wiesbaden, 2013.
7. K. Collins, "An introduction to procedural music in video games," *Contemporary Music Review*, vol. 28, pp. 5–15, feb 2009.
8. M. O. Jewell, M. S. Nixon, and A. Prugel-Bennett, "CBS: A concept-based sequencer for soundtrack composition," in *Proceedings - 3rd International Conference on WEB Delivering of Music, WEDELMUSIC 2003*, pp. 105–108, 2003.
9. M. Luhtala, M. Turunen, J. Hakulinen, and T. Keskinen, "'AIE-studio' - A pragmatist aesthetic approach for procedural sound design," in *Proceedings of the 8th Audio Mostly Conference on - AM '13*, pp. 1–6, 2013.

[5] https://unity3d.com/es

10. N. Böttcher, H. P. Martínez, and S. Serafin, "Procedural audio in computer games using motion controllers: An evaluation on the effect and perception," *International Journal of Computer Games Technology*, no. 6, 2013.
11. M. López Ibáñez, N. Álvarez, and F. Peinado, "A study on an efficient spatialisation technique for near-field sound in video games," in *Proceedings of the 4th Congreso de la Sociedad Española para las Ciencias del Videojuego*, (Barcelona), 2017.
12. M. Eladhari, R. Nieuwdorp, and M. Fridenfalk, "The soundtrack of your mind: Mind music - Adaptive audio for game characters," in *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, (New York), p. 54, 2006.
13. S. Livingstone and A. Brown, "Dynamic response: real-time adaptation for music emotion," in *Second Australasian Conference on Interactive Entertainment*, (Sydney), pp. 105–111, Creativity & Cognition Studios Press, 2005.
14. C. C. Bennett and S. Šabanović, "Deriving minimal features for human-like facial expressions in robotic faces," *International Journal of Social Robotics*, vol. 6, pp. 367–381, aug 2014.
15. M. M. Bradley and P. J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential.," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.