Interacting With Algorithmic Music Through Semantic Descriptors

Fabio Morreale, Raul Masu, Antonella De Angeli

University of Trento, Italy

Abstract. The last decades have witnessed a significant growth of interest in interactive systems that allow the audience to alter the auditory content of an artwork. Generally, the artwork itself is the primary goal of the artist, who arbitrarily maps audience behaviours into musical output following their personal aesthetic. However, given this subjective interpretation, the musical output might not meet audience's expectations. In this paper we discuss a new model for automatic music expression that mediates audience's intentionality. The audience interacts with the system through semantic descriptors (e.g. emotions, movements, narratives), while an algorithm processes this information and returns it to the listeners in the form of semantically coherent music.

1. Introduction

The advent of modern technologies has been creating a new demand for an access to music based on the experience of the listener [2]. In interactive art, this advent dramatically opened new ground for novel musical experiences. General public is now empowered to actively participate in the execution of the artwork by interactively operating on the music content while listening [1]. To foster audience interaction, mapping strategies have to be defined, which can transform audience behaviours into musical output. So far, this mapping has been often set-up arbitrarily: artists themselves determine these strategies in conformity with their personal aesthetics. In these cases, the input is directly translated into musical parameters or acoustic events [4, 12]. This approach is *artist-centric,* as it focuses on the actual artwork and gives the artist complete freedom of expression. By contrast, an *audience-centric* approach would shift the focus to the audience's experience [11]. To this end, more intuitive mapping strategies are needed, as to foster audience understanding. An intermediate layer should be introduced, that "enables users to deal with perceptual and semantic descriptions rather than low-level features that are not accessible to everybody" [8].

This paper reflects on the conceptualization and the operationalization of this intermediate layer. The product of this research is Robin, a rule-based algorithmic composer that loads semantic information and outputs musical patterns consistent with this information. In Section 2, the theoretical issues and the practical implications of using an in-

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 561-564

termediate layer based on semantic descriptors are discussed. Instead of directly mapping audience behaviours into low-level musical parameters, this layer mediated the intentionality of the users through semantic descriptors and translates it into a number of musical features that are used by Robin to compose consistent music. Besides interactive art, this study is of interest for a variety of applications domains where an adaptive music generation system with semantic or affective meaning is needed: film directors and computer games developers, for instance, could generate adaptive soundtracks. In Section 3, the algorithm is detailed: and, finally, in Section 4 two practical application of Robin are presented: The Music Room and TwitterRadio, two interactive artworks that use semantic information supplied by from the visitors themselves direct the composition.

2. Semantic mediators

Traditionally, algorithmic composers adopted in interactive musical expressions directly maps the input sensed from the user into musical and sonic features [5]. Iamascope, for instance, processes visual information describing the current status of the installation and maps it into specific pitches [4]. Similarly, Metaphone detects bio-data from the visitors and uses this information to modulate the frequency and the amplitude of predefined tones [12]. In both cases, scenario-specific information is gathered from the audience. This information is detached from specific semantic features, and the artist freely transforms it into musical parameters and acoustic events. While keeping the mapping between input and musical output unintuitive and ambiguous can be a specific choice of the artist, the audience's experience might not be taken into proper consideration, too. Indeed, listeners might fail to give a meaning to music, as semantic and emotional characterizations are virtually missing. In particular, the musical intentionality of the listeners does not find correspondence in the resulting music.

To tackle this challenge we propose an intermediate layer that mediates the musical intentionality of the audience. This layer implies a symbolic communication of musical intentionality, as described by Cope: "The sender encodes the experience in a symbol, a formal entity chosen by convention which the receiver decodes and understand" [2]. In our model the sender-receiver communication is circular. First, the listener conveys his musical intentionality by using semantic descriptors that are naturally associated with music, such as emotion, narratives, activity, genre, kinaesthetic and movements [7]. Then, the intermediate layer (i.e. the *receiver*) decodes this information using predefined rules and encodes it into patterns of sounds. These encoding rules manage to evoke in the listeners (i.e. the new *receiver*) the associated feelings or images. As an example, the intensity of the movements of the audience can be mapped into musical parameters that describe an increase in the intensity of music such as loud volume and fast tempo. An image or a video depicting storm clouds over a prairie can be mapped into the sounds of mournful saxophone [3].

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 561-564

3. Robin

The conceptual model proposed in the previous Section was operationalized in Robin, an algorithmic composer that generates an original tonal music in real time (details of the algorithm can be found in [9]). The present investigation describes the issues related to designing an algorithmic composer coping with semantic descriptors.

First, the connection between the sensed input and the musical output has to be understandable for the listener. This led us to adopt tonal music, as experimental, complex or atonal compositions usually target niche listeners. Furthermore, a mediator that intuitively describes listeners' intentionality must be exploited. Emotions seems to be the mediator which best meets this requirement, as music is one of the arts that most effectively elicit emotions [6]. The mapping between emotions and musical parameters adopted in Robin is reviewed in [9].

Second, the compositional process has to be semi-deterministic, in the sense that we need to control the evolution of the composition while allowing unexpected behaviours. A rule-based approach guarantees an accurate control of the compositional process as well as the definition of uncommon decisions. The algorithm is taught a series of basic compositional rules of tonal music, which are used to run a number of stochastic processes that mould original Western monodic with accompaniment compositions.

Lastly, the composition needs to be influenced in real time. This choice precludes the definition of predefined musical structures, such as verses and sections as the evolution of user interaction with the system cannot be predicted. The only structural elements composed by Robin are theme repetitions, which simulate refrain, and cadences, that define phrases.

4. Case studies: The Music Room and TwitterRadio

In this section we present two case studies that describe our experience with Robin: The Music Room and TwitterRadio. These two interactive installations show the potential of the algorithm to work with semantic contexts. The Music Room is an interactive installation designed with the aim of allowing everybody to experience musical creativity[1] [8]. The installation takes places in a closed space and it is intended to be experienced by visitors in couple. Couples entering The Music Room can influence the music that is composed by Robin by moving throughout the space of the installation. The metaphor of emotion mediates the interaction between people's behaviour and an algorithmic composer. Specifically, distance between people influences the emotional valence (positive vs. negative) of the mu-

[1] Some excerpts of the generated music can be found at: goo.gl/YgGzux and goo.gl/jm2P4L

sic, while their average speed influences its arousal (level of activation). When they stand close, the music sounds positive and romantic, while it will be sadder and more tragic while they move apart. If they start moving fast, running or dancing, the intensity of the music increases. Body movements are used as semantic descriptors as they provide alternative non-linguistic description of music [2]. The Music Room exploits this association by defining an intuitive mapping between movements and musical structures by means of evoked feelings.

TwitterRadio is an interactive installation designed to explore the social world of Twitter through music [10]. The idea is to access the musical domain to display information about the latest trends and news. Robin automatically generates compositions that match a number of semantic features connected to tweets. The visitors of TwitterRadio can explore themes they are interested in by typing in theme-related hashtags. The system collects all recent associated tweets and retrieves information about emotional valence (positive vs. negative), popularity and retweet frequency. These features are then mapped into music to create compositions semantically mirroring the character of the tweets: rare hashtags will result in slow melodies while trending topics will sound animated or agitated depending on the valence of the tweets. In this case the main semantic descriptor is linguistic: a sentiment analysis algorithm detects the affective character of the tweets. The two remaining descriptors, popularity and retweet percentage, also provide information on the emotionality and intensity of the input.

References

1. Camurri, A., Canepa, C. and Volpe G. Active Listening to a Virtual Orchestra through an Expressive Gestural Interface: The Orchestra Explorer. In Proc. of NIME (2007).
2. Cope, D. Computer Models Of Musical Creativity. Cambridge: MIT Press (2005).
3. Csikszentmihalyi, M. Finding flow: The psychology of engagement with everyday life. Basic Books (1997).
4. Fels, S. and Mase, K. Iamascope: A Graphical Musical Instrument. Computers & Graphics 23(2): 277–86 (1999).
5. Franinovic, K. and Salter, C. The Experience of Sonic Interaction. *Sonic interaction design*: 39 (2013).
6. Juslin P. and Sloboda J. Handbook of Music and Emotion: theory, research, applications: theory, research, applications. Oxford University Press, Oxford (2009).
7. Leman, M. Embodied Music Cognition and Mediation Technologies. MIT Press (2008).
8. Morreale, F., De Angeli, A., Masu, R. Rota, P., and Conci, N. Collaborative Creativity: The Music Room. Personal and Ubiquitous Computing, (2013a).
9. Morreale, F., Masu, R. and De Angeli. A. Robin: An Algorithmic Composer For Interactive Scenarios. In Sound and Music Computing Conference 2013, 207–12, (2013b).
10. Morreale, F., Miniukovich, A. and De Angeli, A. TwitterRadio: Translating Tweets into Music. CHI '14 Extended Abstracts on Human Factors in Computing Systems: 0–3 (2014a).
11. Morreale, F., De Angeli, A., O'Modhrain S. Musical Interface Design: An Experience-oriented Framework. Proc. of NIME '14, accepted for publication (2014b).
12. Šimbelis, V., Lundström, A. and Höök, K. Metaphone: Machine Aesthetics Meets Interaction Design. Proc. of CHI 2014: 1–10 (2014).

Workshop Proceedings of IAS-13
13th Intl. Conf. on Intelligent Autonomous Systems
Padova (Italy) July 15-19, 2014
ISBN 978-88-95872-06-3 pp. 561-564