Dynamic Music Generation, Audio Analysis-Synthesis Methods

Gilberto Bernardes and Diogo Cocharro
INESC TEC and University of Porto, Faculty of Engineering, Porto, Portugal

Synonyms
Adaptive music systems; Audio collage; Audio mosaicing; Concatenative sound synthesis; Interactive music systems

Definitions
Dynamic music generation systems create ever-different and changing musical structures based on formalized computational methods. Under scope is a subset of these methods which adopt musical audio as a strategy to formalize musical structure which then guides higher-level transformations to be synthesized as new musical audio streams.

Introduction
Technologies which adhere to nonlinear, as opposed to fixed, storytelling are becoming pervasive in the digital media landscape (Lister et al. 2003). In this context, methods for dynamically generating music have been prominently and increasingly adopted in games, virtual and augmented reality, interactive installations, and 360 video. Their adoption is motivated by a wide range of factors from computational constraints (e.g., limited memory) to enhanced interaction with external actuators and artistic endeavor.

Dynamic music generation systems are typically driven by formalized or algorithmic methods whose history is interleaved with modern computing (Nierhaus 2009). This article reviews a subset of these systems which adopt musical audio as a source to formalize musical structure algorithmically, which is then used to guide the generation of new musical streams by synthesizing sub-audio clips from the musical audio source – an approach addressed as musical audio analysis-synthesis.

The remainder of this entry details a typical architecture of a generative musical audio analysis-synthesis system (section “From Sound to Musical Audio Analysis-Synthesis Systems”) and presents an overview of its applications scenarios (section “Applications”).

From Sound to Musical Audio Analysis-Synthesis Systems

Audio analysis-synthesis methods break down a sound into some essential, measurable attributes (e.g., amplitude or pitch) to guide sound transformations during (re-)synthesis (Jehan 2005). Historically, these transformations exist at the
sample level and follow an adaptive audio effect architecture (e.g., compression) (Verfaille and Arfib 2001). Recent advances in the hierarchical analysis and generation of musical audio structure have expanded the transformations beyond the sample level towards music processing (e.g., automatic remixing).

Analysis-synthesis systems for dynamic generation of musical audio mimic fundamental perceptual and cognitive human functions in a threefold component architecture of machine listening, learning, and composing (see Fig. 1). Machine listening and learning are two intertwined components which primarily adopt bottom-up (or content-driven) processing methods to infer hierarchical structure from audio samples. It comprises two main tasks: multilevel (beat, downbeat, phrase, and section) segmentation and the description of its temporal structure. Inferred information tends to be represented as graphs, whose nodes represent segmented musical structures, and directed pairwise links their temporal relations (see Fig. 2). The resulting representation provides a robust means for computational structure discovery, notably by finding redundant information across the temporal structure which can be clustered according to some (perceptually guided) similarity metric (see Fig. 2).

Typically, each node or segmented musical audio structure is represented in the system by a feature vector, i.e., a set of numerical features that result from measurable musical attributes, such as pitch, loudness, and percussiveness. The choice of such attributes, the metrics used to compare them, and the perceptual threshold used to define the maximum degree of dissimilarity among segments are crucial to the graph design. All these variables are commonly set manually by the users due to their subjective nature and implications in the resulting musical output of the system (i.e., synthesis). To a certain extent, the choice of such variables is akin to the role of sketches and raw material in the compositional process. For a comprehensive comparison of these variables and their implications in the musical results, please refer to Norowi et al. (2017).

If the resulting structural model has sufficient redundancy, machine composing can then traverse the directed links of the graph to create ever-different and changing musical streams which retain the (higher-level) structure of the audio source captured by the model. The output of an analysis-synthesis systems is musical collages made of rearranged segments or audio snippets from the original or source audio. The structure of the output musical audio can be understood as a variation of particular features temporal evolution of the source, thus retaining some level of similarity while allowing its indeterminate temporal expansion.

**Applications**

We present a user-centered perspective of applications for dynamic generation of musical audio
Dynamic Music Generation, Audio Analysis-Synthesis Methods, Fig. 2 Illustration of a graph model of musical audio structure. The space is defined by (perceptual) attributes (reduced in this illustration to two dimensions). Musical audio segments are understood as data points, represented as circles, whose locations are defined by their attributes in the descriptor space. Small segment distances denote higher (perceptual) similarity in the descriptor space. The circumferences define a perceptual threshold used to cluster similar-sounding segments through analysis-synthesis approaches. Our twofold fuzzy categorization is rooted in the nature of processed signals and distinguishes (natural or synthetic) sound textures and soundscapes from music. A distinction is made concerning the structure of the audio signal content (e.g., noisy vs. pitched or nonmetric vs. metric) rather than any discrimination based on artistic merit, since both categories (music and soundscapes) can be understood as the product of creative-oriented practices.

The generation of sound textures and soundscapes using analysis-synthesis methods has been applied in audio postproduction for television and film as well as in sound design for games and interactive installations. It mostly tackles the pervasive problem of extending a given audio clip in postproduction whenever the prerecorded audio does not cover the entire duration of a scene (Frojd and Horner 2007; Bernardes et al. 2016). Moreover, some analysis-synthesis soundscapes also enable to procedurally generate highly controlled nuances that match external actuators (Bernardes et al. 2013; Schwarz and Schnell 2010). For example, in a game engine, the player behavior can be mapped to soundscape parameters, such as density of events and spectral richness, to enhance the playability through symbiotic relations across modalities.

The generation of music using analysis-synthesis methods has been mainly applied as (online) performance or (offline) computer-assisted composition tools. In the first scenario, it has been highly explored in interactive music systems for human-machine improvisation, where co-improvising machines aim to capture and emulate the musician’s style (Schwarz et al. 2006; Assayag et al. 2006; Surges and Dubnov 2013; Pachet et al. 2013; Einbond et al. 2016). The second scenario has been highly explored to guide the search for variations of a given musical audio excerpt in the realm of entertainment technologies or composition of a given user-defined musical audio (Jehan 2005; Bernardes et al. 2013; Lamere 2012; Davies et al. 2014).

Cross-References

▷ Audiogame
▷ Procedural Audio in Video Games
▷ Sonic Interactions in Virtual Environments
References


