A Knowledge Engineering Approach to Computational Creativity in Sound Design

Eugene Cherny[1,2], Johan Lilius[1],
Johannes Brusila[1], and Dmitry Mouromtsev[2]

[1] Åbo Akademi University, Turku, Finland
[2] ITMO Univesrity, St. Petersburg, Russia
{eugene.cherny, johan.lilius, jbrusila}@abo.fi,
mouromtsev@mail.ifmo.ru

Abstract. This paper proposes an approach to sound design based on knowledge engineering. The approach aims to include aspects of computational creativity into the process to ease the work of the sound designer. To form the basis of our approach we have interviewed a sound designer, and identified a number of problems that we plan to address. We believe that including a computational creativity engine into the tool will help the designer more efficiently explore the spectrum of design parameters for sounds faster, and lead to a more satisfying end result.

Keywords: Computational Creativity, Sound Design, Expert Systems, Knowledge Engineering, Ontology Modelling

1 Introduction

This paper describes an ongoing project exploring computational methods for aiding sound design tasks. The purpose of the project is to address a number of repetitive tasks in the sound design process by using computational creativity methods in sound synthesis and selection. The project's concrete goal is to develop a system for sound generation based on:

1. A domain ontology for structuring the design space of the sound designer
2. Sound descriptions based on the domain ontology that are given as input to the synthesis engine
3. A multi-agent based system that encompasses principles from computational creativity that proposes solutions to the sound descriptions
4. A synthesis engine that creates the sounds.

The scope of the project is limited to creativity support and it is not intended to replace the actual sound designer. This pragmatic limitation was introduced with the intention to make the project usable for real work. The limitation simplifies several technical and conceptual questions, such as we do not need to justify the algorithmic assessment of generated creative assets or the rigorous definition of a creativity concept for the project. Instead we view our tool as augmenting human creativity with intelligent computational methods in the sound design process. Thus in the end the success of the approach will be measured through the adoption by sound designers.

Computational Creativity is a multi-disciplinary research area exploring practical and theoretical issues of simulating creative behavior with a computer. One of the most influential works in this field is the research of Margaret Boden [1], proposing conceptual spaces as a way to structurally reason about creativity. Although she did not provide a formal description of her approach, the framework developed by Geraint Wiggins [2] provides a mathematical model of creativity based on Boden's ideas. We will reflect on the relationship between our knowledge engineering approach with computational creativity and the framework by Wiggins [2] in particular.

The aim of this paper is to give an overview of our approach, and focus on the general methodological issues in the knowledge elicitation process. We use interviews with professional sound designers, and already documented knowledge (books [3–5] blog posts [6–8] to build a sound design ontology. The ontology is expected to evolve in an interactive manner, as more information is added through the interviews, and from other sources. The interviews are also intended to critique the approach as the sound designers can easily comment what could be useful and what could not.

2 Related Work

There are a lot of examples of computational creativity projects, but for this paper we would like to mention some of those, which directly concern the sound itself, i.e. using computational/AI methods to generate sound.

In [9] (SeaWave) authors use a taxonomy to synthesise the sound. The taxonomy terms include informal descriptions of the timbral and temporal features of the sound. The additive synthesis technique was used, but the programmers predefined the set of timbres. Other work [10] (ARTIST) elaborates further the approach by moving a taxonomy of sound attributes into a knowledge base. A subtractive synthesier uses it to provide "some degree of automated reasoning which supports the laborious and tedious task of writing down number sequences for generating a single sound on a computer". A knowledge base approach was also used in a more recent work [11], where authors developed an expert system for additive sound synthesis controlled by fuzzy sound description attributes. Different machine learning methods were also used to create sound synthesisers, for example, in [12] two neural networks was used to map a verbal descriptor space onto an additive synthesier parameters and vice-versa; in [13] an evolutionary algorithm was used to find the optimal parameters for a FM synthesier for the sound matching purpose.

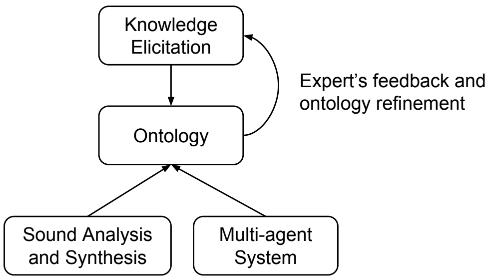

Fig. 1. Approach diagram.

3 Approach Overview

The project goal is to build a tool for sound designers and an approach was shaped this idea in mind. Fig. 1 demonstrates the main components of the approach: it starts with a knowledge elicitation to formalize sound designers' knowledge in the form of ontology, which then binds a synthesis and a multi-agent systems together. The approach is iterative and involves continuous improvement through interaction with professionals. We do not try to capture "an objective" domain knowledge and think about the approach in a pragmatic way: if the tool is useful for sound designers then the approach will be assessed as successful.

The knowledge elicitation is done through interviews and documents analysis. Subjects of interest here are sound design techniques and terminology used by experts. The former provides us with the knowledge how to manipulate sound, the latter—with an essential knowledge to describe sound in an informal manner. Many professionals often describe timbral, temporal or other aspects of the sound using mentions of real-world objects and phenomena ("brass", "rotating", etc.), but also ones which does not have strict definitions, for example ("dark", "dull", "sparkling", etc.) [14]. They are often metaphoric, representing how a subject perceives the sound, but there is evidence such verbalisations can be meaningfully used for timbre description [9, 15, 14, 16, 17] and even hold their meanings across people speaking different languages [18]. Also the creation of a controlled vocabulary of the most meaningful words to describe sound has its own value [19]. The outcome of the knowledge elicitation will be a domain ontology.

The ontology has three main parts: a sound design terminology, a taxonomy of sound objects and a model of sound design process. The first one provides the basic concepts used in the domain, like physical or psychoacoustic terms and also informal words used to communicate about sound, elicited from the interviews. The second one, the taxonomy, adds real-world object entities to the ontology to annotate sounds. These may include, for example, "a car", "a rotor", etc.—

the objects which have their own distinctive sound. Such terms complement the first part, as we can describe the sound itself not only by how we perceive it, but also by referring to the object that produce it ("flutey", "glassy", etc.). The third one holds the formal representation of the sound design techniques (like effect processing chains, etc.), actions (what can be done with the sound) and connections between actions and their outcomes (represented as concepts in the first part). The purpose of these connections is to annotate how actions affect the sound with already known terms and add semantic meaning to all manipulations with the sound. Thus, the ontology provides the means, from the one hand, to describe a sound in the formally encoded terms and, from the other, to add semantics to sound manipulations.

"The creativity" in the project will be implemented as a multi-agent system. Each agent is fully capable to produce creative assets by its own. To do this it composes knowledge from the ontology to make a sound from a combination of actions. Being connected with others, the agent is capable to look at the state of its neighbors and, depending on its behaviour traits, evolve own state in a specific way. For example in [20] authors assign a role to each agent, thus making one to mimic other agents, or the opposite, to oppose others in creative assets generation.

4 Knowledge Elicitation

4.1 Informal Interview with a Sound Designer

During the project work a semi-structured interview with a professional sound designer has been conducted. The purpose of this interview was to get the overview of the domain and to understand existing problems to be addressed. "Semi-structured" means that the interview was conducted with the concrete purpose, but without elaborated structure. This section summarizes the interview.

There are two approaches to sound design differing by the way the sound is generated and used: a sample-based and a procedural [3]. The former presumes processing of already recorded sounds to be triggered at the specific game event, and the latter—building the signal processing chain for use in a real-time context. Although the latter approach is promising, it is not commonly used due to implementation complexity, hence the sample-based approach is used most to do sound design. But it creates a burden of navigating through big sample libraries for the interviewee. Sample management software does not completely solve the problem, as it is based on a per-sample hand-added metadata, which often incomplete or even missing. This leads to the situations when a single sound query returns several dozens of sounds need to be listened in order to find a relevant one for a task.

Sound design is done based on the technical requirements or other concept document. As a sound is designed to support activities on the screen, the visual information usually defines the requirements for the sound. Hence, a sound designer has a visual reference for the sound to be created, which also can be described verbally. For example, the sentence "boxy watery human breathing", which is more elaborated in the next section, is a typical verbalization form that narrows down the possible solutions for the sound to be made.

And finally, in game sound design there is a need to create sets of similar sounds that will be triggered randomly during the game (e.g. gun shots).

During the interview the following list of repetitive tasks of the sound design process have been identified: 1) navigation through a big sample library, 2) construction of the effect chains based on the unstructured query, 3) creation of the similar sounds to create random sets for games.

4.2 Case: Heavy Armor Sound Design for the InSomnia Project

A sound designer working on the game InSomnia[3] has published a detailed story how he designed a sound for a heavy armored game character [6]. For the sound examples and visual references we suggest the reader to refer to the original post, and here we would like to give a brief sound design process summary for this case and discuss its relevance to the raised issues.

The author defined three stages in the design process: concept (defining how to make a sound), sound design in DAW, embedding sound into a game engine. The following paragraph is a short retelling of the art concept, provided by the author.

The game has a dieselpunk setting, hence the mechanical sounds should be used. No sci-fi elements or electronics. A human is inside the armor. The armor is heavy, hence the sound should convey the feeling of something massive. There is a helmet and a knapsack, the latter contains something, probably a water pipeline system. The armor has a lot of moving parts and small mechanical elements. The helmet is somehow connected to the water system in the knapsack, so the human inside breathes through some kind of a "water filter". The resulting armor sound consists of three different sounds: a voice, a knapsack, a metal/mechanics.

For the purpose of this paper to show the example of knowledge that can be elicited from the use case, we would like to elaborate only on the voice sound.

The sound effect chain is represented on the Fig. 2.

The authors of this paper describe the resulting sound as "boxy watery human breathing", that means that the purpose of the whole processing chain is to make the sound with this characteristic. This sentence has a clear structure: it contains of a noun ("human breathing") and two adjectives ("boxy" and "watery"). The noun corresponds to the sound source and the adjectives—to the effect blocks. This natural form of sound verbalization can be used to create a query for the sound specification.

To add 'watery" feeling to the breathing sound the designer decided to use a vocoder which applies spectral envelope of breathing to the distorted bubbles sound and hence makes the sound "watery". The combination of resonators and

[3] InSomnia project homepage, URL: `http://www.insomnia-project.com/`

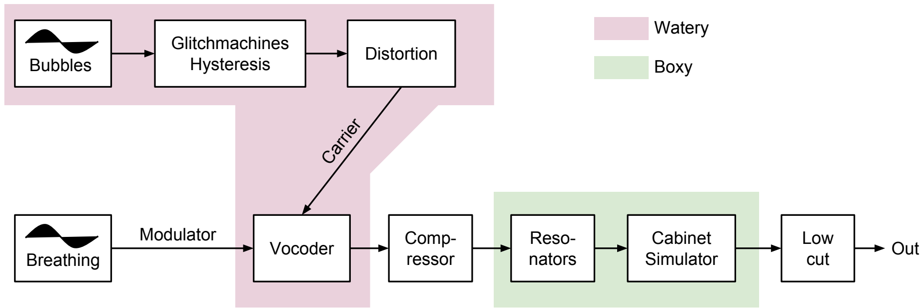

Fig. 2. The sound processing of the "boxy watery human breathing" sound.

cabinet simulator are, in our opinion, responsible for the "boxy" characteristic of the sound.

This example illustrated that some simple statements can be decomposed into small chunks that have correspondence to sound processing blocks. The interpretation we provided is one of many and in the tool a multi-agent system will be responsible for doing the search for ones.

5 Multi-agent System for Sound Generation

As discussed in the previous section, the sound design process starts with a concept requirement—an informal description of the sound to be created. Hence, we can see the creativity problem as a search for solutions satisfying this informal description. Such simplification has direct correspondence to the computational creativity model explored by Boden [1], and developed it into a more formal framework later by Wiggins[2]. The core of this approach is the existence of conceptual space—a knowledge gathered by a creative agent. This knowledge includes culture, history, skills and other concepts, which define agents thinking. "Boden conceives the process of creativity as the identification and/or location of new conceptual objects in a conceptual space" [2]. She also discriminates between exploratory and transformational creativity. The former is the search for a concept in the current space, and the latter involves changing of the agent's conceptual space in order to come up with a new concept. To provide more rigorous description of the model Wiggins [2] introduces some modifications. He defines the universe as all possible concepts, and conceptual spaces are the subsets of the universe. He also defines two distinct rule sets: to constrain the conceptual space and to traverse it—hence the artificial agents differ not only in the concepts they hold, but also in how they look for new concepts.

This project proposes a computational creativity tool based a multi-agent system. At the very heart of the system is the domain ontology, filled with the knowledge elicited from sound designers and domain-specific sources. It contains rules how the sound can be manipulated and a terminology to describe sound and the information how different sound processing blocks affect the sound. The multi-agent system, once initialized, sets up a number of agents, gives a slice of the ontology and assigns a behavioral trait for each of them and creates initial connections between them. Once system is executed, each of the agents looks for the solution for the requested sound query. Agents can evolve over time based on their behavioral traits and connections with other agents. In terms of Wiggins's framework the ontology plays a role of the universe, the ontology slices are the conceptual spaces, the sound query is the space constrain rule and the behavioral traits are the space traverse rules.

6 Conclusions and Future Work

In this paper we proposed an approach to sound design computational creativity based on the knowledge engineering and multi-agent system and discussed its connection to the existing computational creativity framework [2]. Although the project is still in its early stage, the results of the interviews and the case study have clearly demonstrated the importance of the knowledge elicitation for creativity process modeling.

At the current state a number of issues still present. Particularly, the multi-agent system and sound analysis and synthesis algorithms are not well elaborated yet. The multi-agent system concept lacks specification on agents communication and how the system should be embedded in the process of the actual sound designer. Sound analysis has a great importance for the project, as it should provide the means for the smart search in the sound library. The methods for automatic annotation of audio files should be developed in order to provide responses for an informal query. The tool will be evaluated by the professionals' feedback, assessing how well it integrates with their workflow and how they like to work with it.

Acknowledgements

We wish to thank Ivan Osipenko, a sound designer at Saber Interactive (St. Petersburg, Russia), for the interview and sound design discussions.

This work was partially financially supported by the Government of Russian Federation, Grant #074-U01.

References

1. Boden, M.A.: The creative mind: Myths and mechanisms. Psychology Press (2004)
2. Wiggins, G.A.: A preliminary framework for description, analysis and comparison of creative systems. Knowledge-Based Systems 19(7), 449–458 (2006)
3. Farnell, A.: Designing sound. Mit Press Cambridge (2010)
4. Cipriani, A., Giri, M.: Electronic Music and Sound Design - Theory and Practice with Max 7, vol. 1. Contemponet, 3 edn. (2016)

5. Viers, R.: The sound effects bible: how to create and record hollywood style sound effects. Michael Wiese Productions (2008)
6. Yakyamsev, P.: Heavy armor sound design for the game InSomnia (in russian) (2015), http://habrahabr.ru/post/272635/
7. Frank Bry: Frank Bry Special - Methods of a Metadata Madman, Part 2 (2010), http://designingsound.org/2010/12/frank-bry-special-methods-of-a-metadata-madman-part-2/
8. Tim Prebble: You and your Metadata.... (2010), http://www.musicofsound.co.nz/blog/you-and-your-metadata
9. Ethington, R., Punch, B.: Seawave: A system for musical timbre description. Computer Music Journal 18(1), 30–39 (1994)
10. Miranda, E.R., CSL-Paris, S.: Striking the right note with artist: an ai-based synthesiser. Recherches et applications en informatique musicale pp. 227–239 (1998)
11. Hamadicharef, B., Ifeachor, E.C.: Intelligent and perceptual-based approach to musical instruments sound design. Expert Systems with Applications 39(7), 6476–6484 (2012)
12. Gounaropoulos, A., Johnson, C.G.: A Neural Network Approach for Synthesising Timbres from Adjectives. In: Proceedings SMC'07, 4th Sound and Music Computing Conference. pp. 287–290 (2007)
13. Mitchell, T., Sullivan, C.: Frequency modulation tone matching using a fuzzy clustering evolution strategy. In: Audio Engineering Society Convention 118. Audio Engineering Society (2005)
14. Sundaram, S., Narayanan, S.: Classification of sound clips by two schemes: using onomatopoeia and semantic labels. In: Multimedia and Expo, 2008 IEEE International Conference on. pp. 1341–1344. IEEE (2008)
15. Samoylenko, E.: Systematic analysis of verbalizations produced in comparing musical timbres. International journal of psychology 31(6), 255–278 (1996)
16. Verbal expression of piano timbre: Multidimensional semantic space of adjectival descriptors. In: Williamon, A., Edwards, D., Bartel, L. (eds.) International Symposium on Performance Science. Ass. Europenne des Conservatoires, Acadmies de Musique et Musikhochschulen (AEC) (2011)
17. Asterios Zacharakis, Konstantinos Pastiadis, Joshua D. Reiss, George Papadelis: Analysis of musical timbre semantics through metric and non-metric data reduction techniques. In: Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC12) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM 08). pp. 1177–1182 (2012)
18. Zacharakis, A., Pastiadis, K., Reiss, J.D.: An Interlanguage Unification of Musical Timbre: Bridging Semantic, Perceptual, and Acoustic Dimensions 32(4), 394–412 (2015)
19. Font, F., Serra, X.: Analysis of the folksonomy of freesound. In: Serra X, Rao P, Murthy H, Bozkurt B, editors. Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey. Barcelona: Universitat Pompeu Fabra; 2012. Universitat Pompeu Fabra (2012)
20. Gabora, L.: An evolutionary framework for cultural change: selectionism versus communal exchange. Physics of Life Reviews 10(2), 117–145 (2013)