
Audio Production with Intelligent Machines

Mark Cartwright

Interactive Audio Lab
Department of EECS
Northwestern University
Evanston, IL
mcartwright@u.northwestern.edu

Bryan Pardo

Interactive Audio Lab
Department of EECS
Northwestern University
Evanston, IL
pardo@northwestern.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CHI'15 Workshop 05 "Collaborating with Intelligent Machines: Interfaces for Creative Sound", April 18, 2015, Seoul, Republic of Korea.

Abstract

To effectively utilize and collaborate with audio production tools, users (especially novices) must be able to communicate their ideas in the ways that work well for them. However, many artists' preferred means of interaction are not well aligned with the interfaces of most audio production software. The resulting mismatch can disrupt the flow of ideas and inhibit creativity. A new approach to interface design is required.

Author Keywords

audio production, HCI, intelligent machines, collaboration

ACM Classification Keywords

H.5.2. [User Interfaces]: Interaction styles; H.5.5. [Sound and Music Computing]: Systems

Introduction

Audio production is an integral process of music composition in the 21st century. However, most audio production tools are still controlled in terms of low-level technical parameters that map directly onto signal processing algorithms. These do not have clear mappings into the conceptual spaces of many artists and therefore require significant experience to use effectively. Without such experience, novices have a particularly difficult time communicating their ideas to audio production tools and

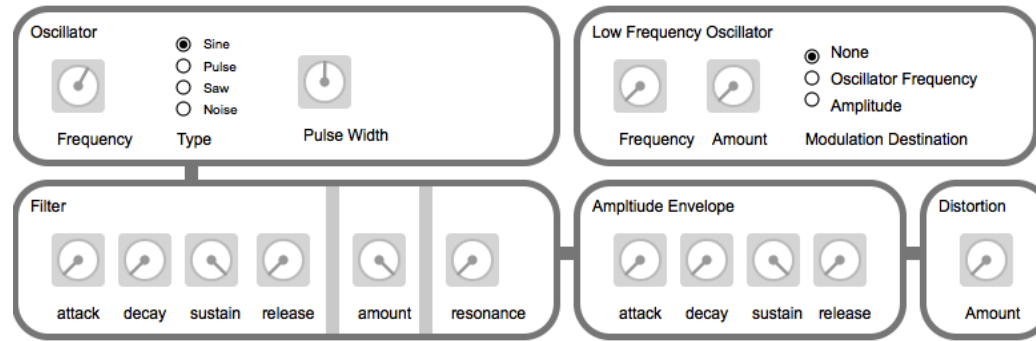


Figure 1: A traditional synthesizer interface.

achieving their creative goals.

For example, when a novice participant in one of our studies was asked to program a given synthesizer sound with a traditional synthesizer interface (see Figure 1), they stated:

I give up. I can't find any improvement. It's as if you put me in the control room of an airplane, and I can't even take off.

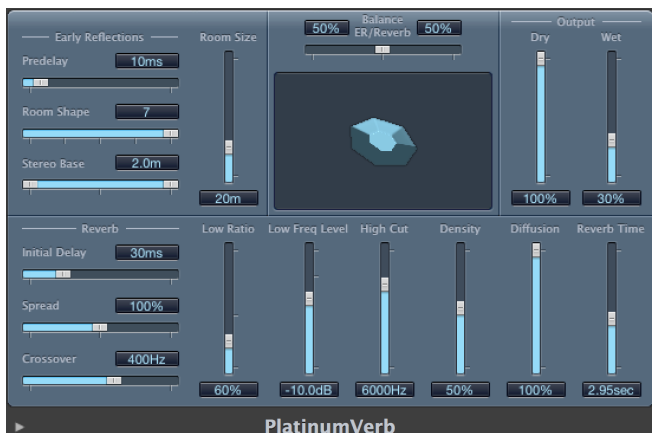
In another example, a novice user on Reddit stated:

I have been playing guitar 30 years. I bought the recording interface, software, etc. 6 months ago. As I am 48 and work as a carpenter, I am just too damn tired all the time to learn this stuff. There is so much to learn at the same time, I don't know all the terminology. I have given up for now. Sad, because I have lots of ideas.

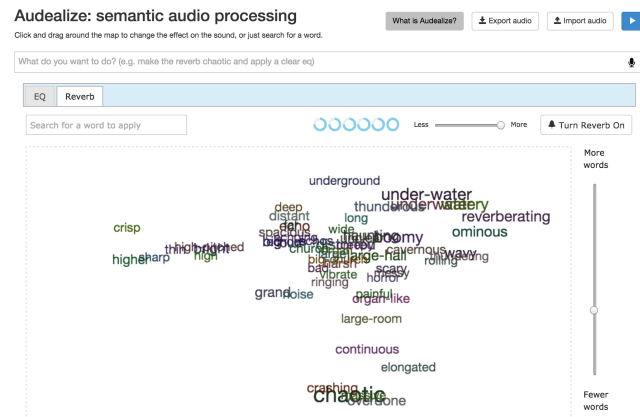
The problem is that these users do not know how to communicate their ideas using low-level technical parameters. They get frustrated and give up. By combining machine learning, signal processing, and human computer interaction, intelligent audio production tools may be able to bridge this gap.

Communication

When dealing with audio production tools, there are several conceptual spaces the user must be able to negotiate to achieve their goals: the *parameter space*, the *perceptual space*, and the *semantic space*. The *parameter space* is the space of low-level technical parameters of audio production tools. On a reverbation tool, such a control might be a knob labeled "reverb time" (see Figure 2a). This is the space that users typically have to navigate in order to achieve their goals. If the production tool is an audio processing tool rather than a synthesis tool, then the relation of this space to the audio output of the system is also dependent on the input audio signal and its features. When we perceive the audio output of an audio production tool, it is projected to a *perceptual*



(a) A traditional reverb interface (Apple Inc’s PlatinumVerb).



(b) Audealize, a semantic reverb and EQ interface based on the descriptor maps learned in [1, 9].

Figure 2: Reverb interfaces

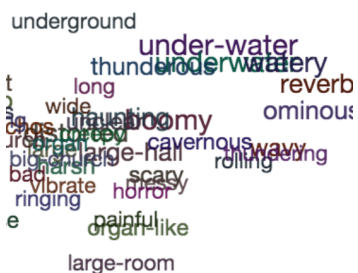


Figure 3: Reverb descriptions in Audealize (see Figure 2b).

space, which describes how and what aspects of the air pressure waves actually form our aural perception (we notice longer echos). Lastly, what we perceive also relates to an even higher level *semantic space* which is how we describe the world and is therefore more aligned with our goals (it sounds like a “big cave”).

To communicate with current audio production tools, one must understand the parameter space and how it relates to the higher level perceptual and semantic spaces. This mapping is often highly nonlinear and dependent on multiple settings. This makes it difficult to learn mappings between desired effects (“sound like you’re in a cave”) and the low level parameters (a knob on a reverb unit labeled “diffusion”). We envision future audio production tools that let users communicate their ideas through natural language, perceptual dimensions, vocal imitation,

gesture / performance, brain-computer interfaces, and other ways they find effective.

Exploration

In addition, while being able to communicate an isolated idea (“make it sound ‘cave’ like”) is a good first step, exploration is also very important. Therefore, future audio production tools should not only enable jumping to a desired location in the parameter space but also enable a user to fluidly traverse the space in a semantically meaningful way (“Go from ‘cave’ reverb smoothly to ‘small closet’ reverb”). For an example, see Figures 2b and 3).

By learning mappings between the spaces and meaningful subspaces and manifolds in high dimensional parameter spaces, we can help users communicate and explore. We

believe these spaces are dynamic as well, dependent on the user and the context. Therefore, achieving this goal requires learning more about these spaces, incorporating context, and personalizing for users. Some work has already been done [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], but more work is needed in order to control complex production tools and to combine multiple tools to achieve users goals.

Collaboration

Users do not always communicate audio concepts using the same words, language, or even modality. Therefore, by enabling such communication and exploration, users can also then begin to collaborate meaningfully with different agents, either artificial or other humans, opening entirely new means of effective interaction.

Conclusion

We imagine future audio production tools in which users can communicate their ideas with ease, enabling both novices and experienced users to achieve their goals and more easily collaborate with others.

Acknowledgements

This work was supported by NSF Grant Nos. IIS-1116384 and DGE-0824162.

References

- [1] Cartwright, M., and Pardo, B. Social-eq: Crowdsourcing an equalization descriptor map. In *Proc. of International Society for Music Information Retrieval* (Curitiba, Brazil, 2013).
- [2] Cartwright, M., and Pardo, B. Synthassist: Querying an audio synthesizer by vocal imitation. In *Conference on New Interfaces for Musical Expression* (2014).
- [3] De Man, B., and Reiss, J. D. A

- knowledge-engineered autonomous mixing system. In *Proceedings of the Audio Engineering Society Convention 135*, Audio Engineering Society (2013).
- [4] Heise, S., Hlatky, M., and Loviscach, J. Aurally and visually enhanced audio search with soundtorch. In *Proc. of International Conference Extended Abstracts on Human factors in Computing Systems* (2009).
- [5] Huang, C.-Z. A., Duvenaud, D., Arnold, K. C., Partridge, B., Oberholtzer, J. W., and Gajos, K. Z. Active learning of intuitive control knobs for synthesizers using gaussian processes. In *Proc. of Int'l Conference on Intelligent User Interfaces* (Haifa, Israel, 2014).
- [6] Mecklenburg, S., and Loviscach, J. subject: controlling an equalizer through subjective terms. In *Proc. of CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, Canada, 2006).
- [7] Reed, D. A perceptual assistant to do sound equalization. In *Proc. of the 5th international conference on Intelligent user interfaces*, ACM (2000), 212–218.
- [8] Sabin, A., Rafii, Z., and Pardo, B. Weighting-function-based rapid mapping of descriptors to audio processing parameters. *Journal of the Audio Engineering Society* 59, 6 (2011), 419–430.
- [9] Seetharaman, P., and Pardo, B. Crowdsourcing a reverberation descriptor map. In *Proceedings of the ACM International Conference on Multimedia*, ACM (2014), 587–596.
- [10] Stowell, D. *Making music through real-time voice timbre analysis: machine learning and timbral control*. PhD thesis, Queen Mary University of London, 2010.