SynthAssist: An Audio Synthesizer Programmed with Vocal Imitation

Mark Cartwright
Northwestern University
EECS Department
mcartwright@u.northwestern.edu

Bryan Pardo
Northwestern University
EECS Department
pardo@northwestern.edu

ABSTRACT

While programming an audio synthesizer can be difficult, if a user has a general idea of the sound they are trying to program, they may be able to imitate it with their voice. In this technical demonstration, we demonstrate SynthAssist, a system that allows the user to program an audio synthesizer using vocal imitation and interactive feedback. This system treats synthesizer programming as an audio information retrieval task. To account for the limitations of the human voice, it compares vocal imitations to synthesizer sounds by using both absolute and relative temporal shapes of relevant audio features, and it refines the query and feature weights using relevance feedback.

Categories and Subject Descriptors

H.5.5. [**Sound and Music Computing**]: Systems; H.5.2. [**User Interfaces**]: Interaction styles; H.5.1. [**Multimedia Information Systems**]: Audio input/output

General Terms

Human Factors

Keywords

Audio; music; synthesis; vocal imitation

1. INTRODUCTION

As software-based audio synthesizers have become more advanced, their interfaces have become more complex and therefore harder to use. Many synthesizer interfaces have one hundred or more controls (e.g. Apple's ES2 synthesizer has 125 controls), many of which are difficult to understand without significant synthesis experience. The search space of even simple synthesizers (15 controls) is complex enough that it is difficult for novices to actualize their ideas. Even for experienced users, the tedium of these interfaces takes them out of their creative flow state, hampering productivity. Some manufacturers address this problem by having many, many presets (e.g. Native Instruments Kore Browser). However, searching through a vast number of presets can be a task as daunting as using a complex synthesizer.

In this technical demonstration, we demonstrate SynthAssist [1], a system that interactively helps the user find their desired sound (synthesizer patch) in the space of sounds generated by an audio synthesizer. The goal is to make synthesizers more accessible, letting users focus on high-level goals (e.g. 'sound brassy') instead of low-level controls (e.g. 'What does the LFO1Asym knob do?'). SynthAssist accomplishes this by allowing the user to communicate their desired sound with a vocal imitation of the sound, a method of communication shown to be effective in a recent study [2]. The inspiration for SynthAssist was how one might interact with a professional music producer or sound designer: imitate the desired outcome vocally (e.g. "make a sound that goes like <sound effect made vocally>"), have the producer design a few options based on the example, give them evaluative feedback (e.g "that doesn't sound as good as the previous example.")

2. SYNTHASSSIST SYSTEM

With SynthAssist, a user can quickly and easily search through thousands of synthesizer sounds to find the desired option. The interaction for a session in SynthAssist is as follows:

1. To communicate their desired audio concept, the user provides one or more input examples by either recording a new sound (e.g. vocalize an example) or choosing a prerecorded sound.
2. The user rates how close each example is to the target sound.
3. Based on the ratings, SynthAssist estimates what the target sound must be like.
4. SynthAssist generates suggestions from the synthesizer that are similar to the estimated target sound.
5. The user rates how close the suggestions are to the desired target sound.
6. If a suggestion is good enough, return the synthesizer parameters. Else, repeat Steps 3 through 6.

Figure 1 shows a screenshot of SynthAssist. Each suggestion is represented by one of the colored circles. When a user clicks or moves a suggestion, the synthesizer sound

MM'14, November 3–7, 2014, Orlando, Florida, USA.
ACM 978-1-4503-3063-3/14/11.
http://dx.doi.org/10.1145/2647868.2654880.

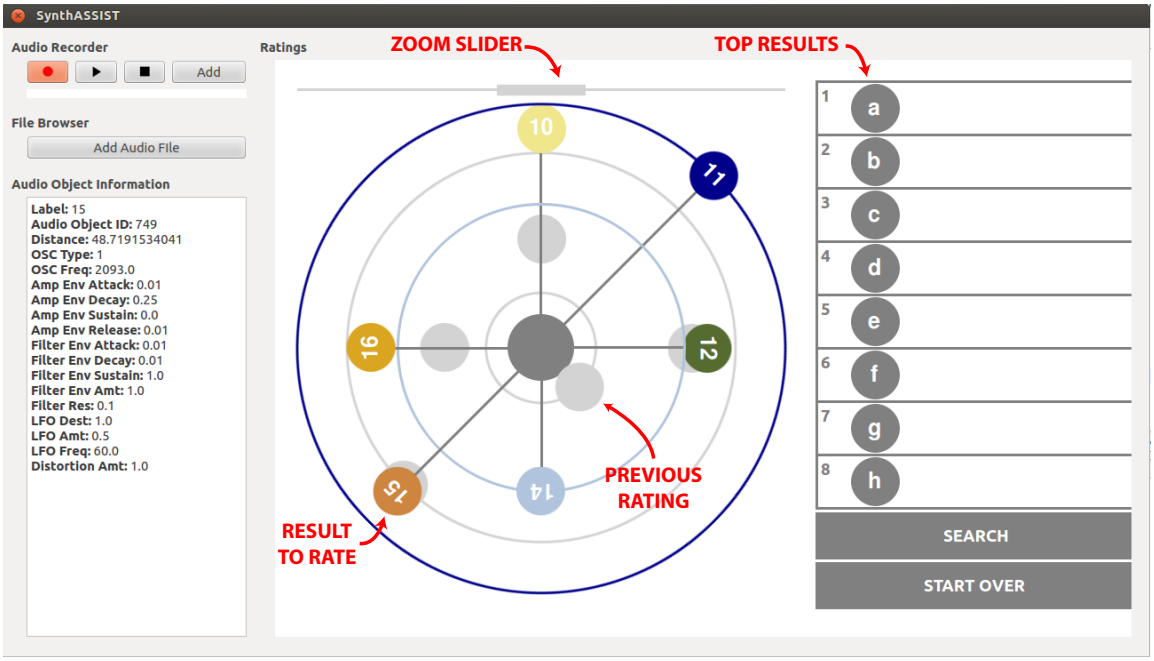

Figure 1: Screenshot of SynthAssist

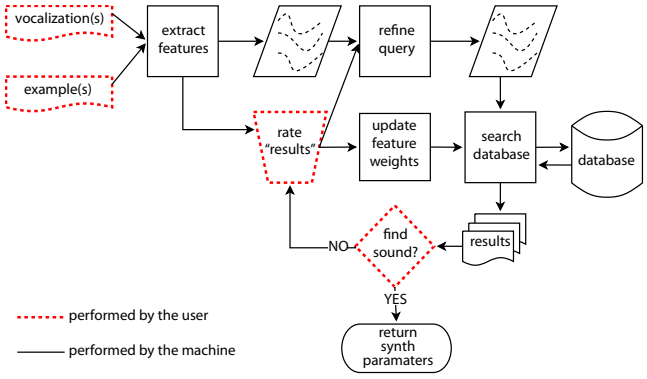

Figure 2: System flow of SynthAssist.

changes. Users rate how similar suggestions are to their target by moving them closer to or farther from the center, "hub", circle. If a suggestion is irrelevant, the user can inform SynthAssist and remove it from the screen by double clicking on it. Dragging a suggestion to the center of the circle indicates indicates this is the desired sound and terminates the interaction. The overall system architecture is shown in Figure 2.

In the SynthAssist system, we want to support vocal imitation queries. Our approach is described in [1]. While the human voice has a limited timbral range, it can very expressive with respect to how it changes in time. For example, your voice may not be able to exactly imitate your favorite Moog bass sound, but you may be able to imitate how the sound changes over the course of a note (e.g. pitch, loudness, brightness, noisiness, etc.). When comparing sounds in SynthAssist, we focus on these changes over the course of a note.

In this demonstration, 10,000 parameter combinations and their resulting sounds have been sampled from an audio synthesizer. Absolute and relative time series of several audio features are extracted from both the query and sampled sounds. The time series of the query are compared to those of the sampled sounds using dynamic time warping. The distance measure and query are refined using a relevance feedback mechanism for time series.

3. THE DEMONSTRATION

In the demonstration of SynthAssist, a computer will be set up running the SynthAssist software. Users will be able to record their vocal imitations and use them as input to the system to interactively program a synthesizer sound. Users can interact with the system either by itself or in a musical context, programming a sound for an existing keyboard line in a song. In both modes, the user can use a piano keyboard to play the synthesizer.

A video of an example search session and an explanation of the system can be found at `http://music.eecs.northwestern.edu/videos/SynthAssist.mov`.

4. ACKNOWLEDGMENTS

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

5. REFERENCES

[1] M. Cartwright and B. Pardo. Synthassist: Querying an audio synthesizer by vocal imitation. In *Conference on New Interfaces for Musical Expression*, London, 2014.

[2] G. Lemaitre and D. Rocchesso. On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America*, 135(2):862–873, 2014.