Learning to Control a Reverberator using Subjective Perceptual Descriptors

Zafar Rafii & Bryan Pardo

EECS Department, Northwestern University, *Evanston, IL, USA.*

ZafarRafii2011@u.northwestern.edu • pardo@northwestern.edu • http://music.cs.northwestern.edu/lab/

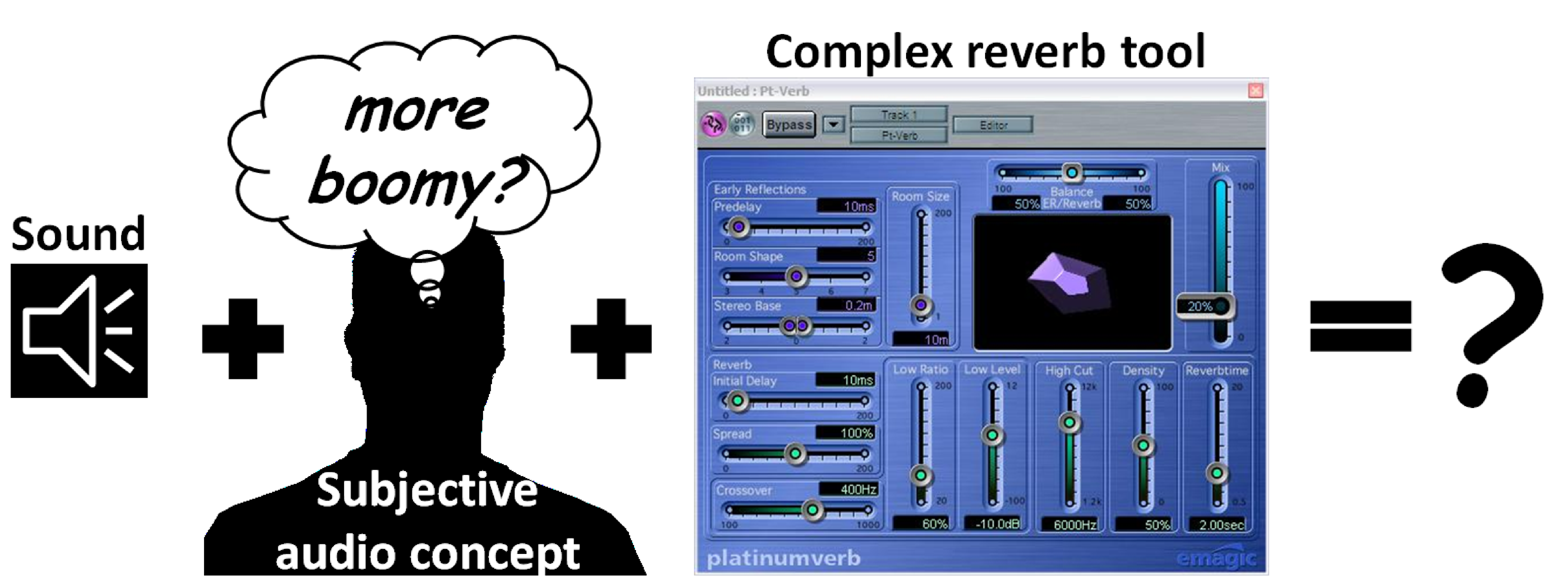
NORTHWESTERN UNIVERSITY

interactive audio lab

Introduction

We have built a system where a user can easily and rapidly teach the machine a subjective audio concept (such as "boomy" or "church-like") in order to build a simple controller that can manipulate sound in terms of that audio concept, bypassing the bottleneck of complex interfaces and individual differences in descriptive terms. Here, we propose to simplify and personalize the interface of one of the most widely used audio effect: *Reverberation*.

Motivations



▶ **Problems**:

- *What do you mean?* - Audio descriptors are subjective...
- *How do you proceed?* - Audio tools are too complex...

▶ **Solution**:

- Mapping audio descriptors onto the controls for audio tools (here, *Reverberation*) on a case-by-case basis.

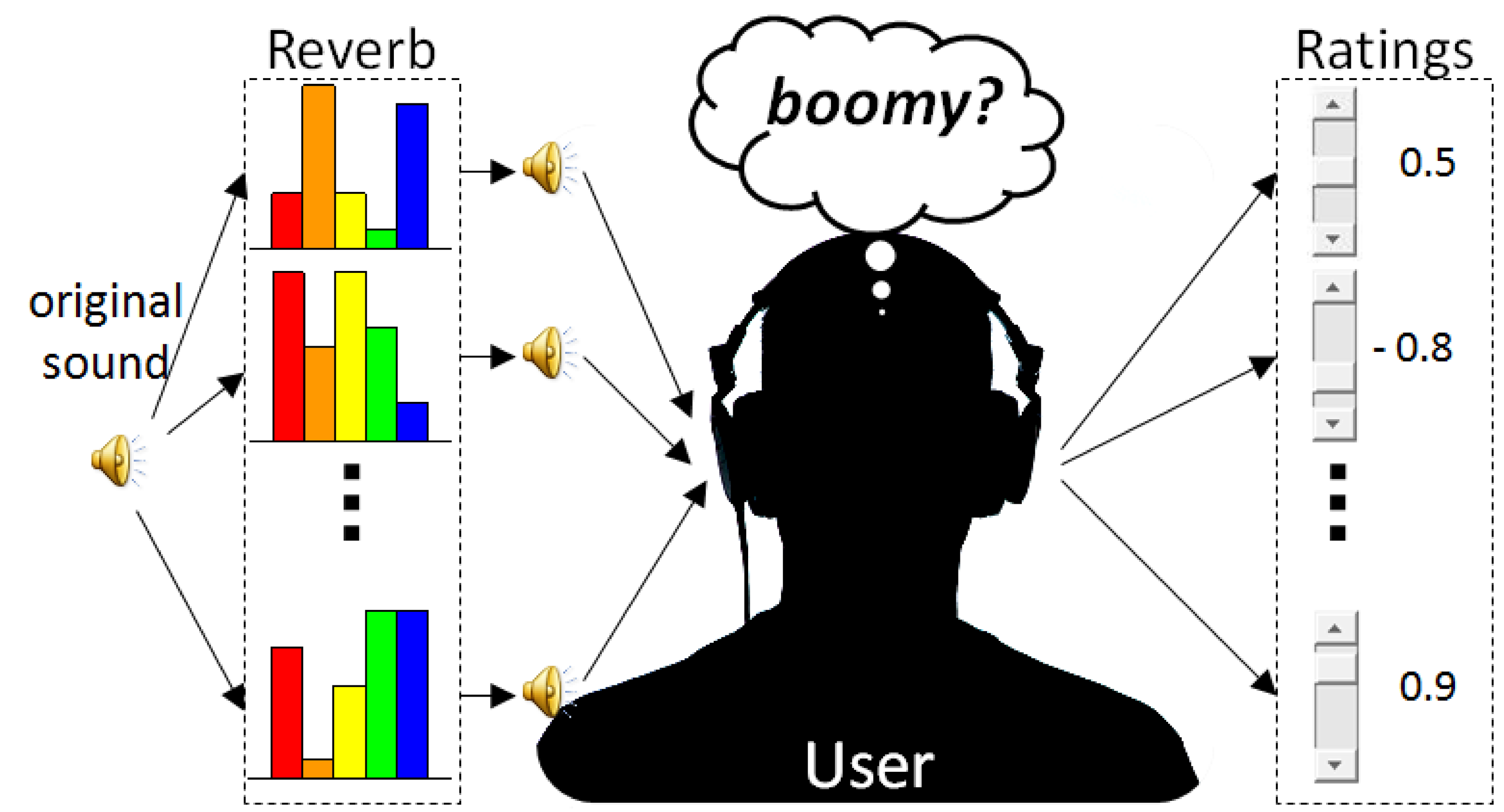
Method

- We have developed a **stereo reverberator** composed of different digital filters (comb, all-pass and low-pass).
- This reverberator is controlled through **5 measures** commonly used to characterize reverberation:

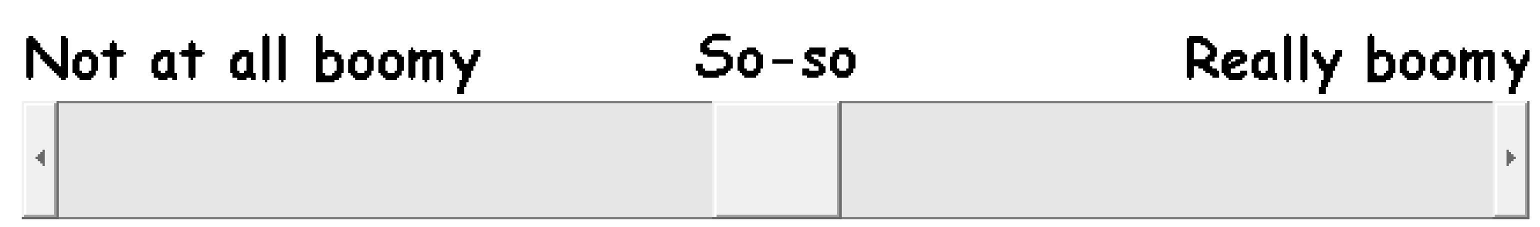
▶ *Reverberation Time*, *Echo Density*, *Clarity*, *Central Time*, & *Spectral Centroid*.

❶ Audio examples are generated from an original sound using different reverberation measures settings.

❷ The listener uses a slider to rate how well each example fits the audio concept she/he has in mind.



❸ The system maps user ratings to the 5 measures and builds a simple subjective audio controller.



Evaluation

- 22 participants (14 males and 8 females).
- 5 perceptual descriptors: *Bright*, *Clear*, *Boomy*, *Bathroom-like*, & *Church-like*.
- 3 tasks to perform, for each descriptor:
  - ❶ Rate a first set of audio examples (-1 to 1)
  - ❷ Rate a new set generated from the learned mapping
  - ❸ Give a rating to the final controller (0 to 10)

▶ Average *learning time* = **2 min 20 sec**

▶ Average *user consistency* correlation = **0.65**

▶ Average *system predictiveness* correlation = **0.75**

▶ Average *human rating* = **7.4** (out of 10)

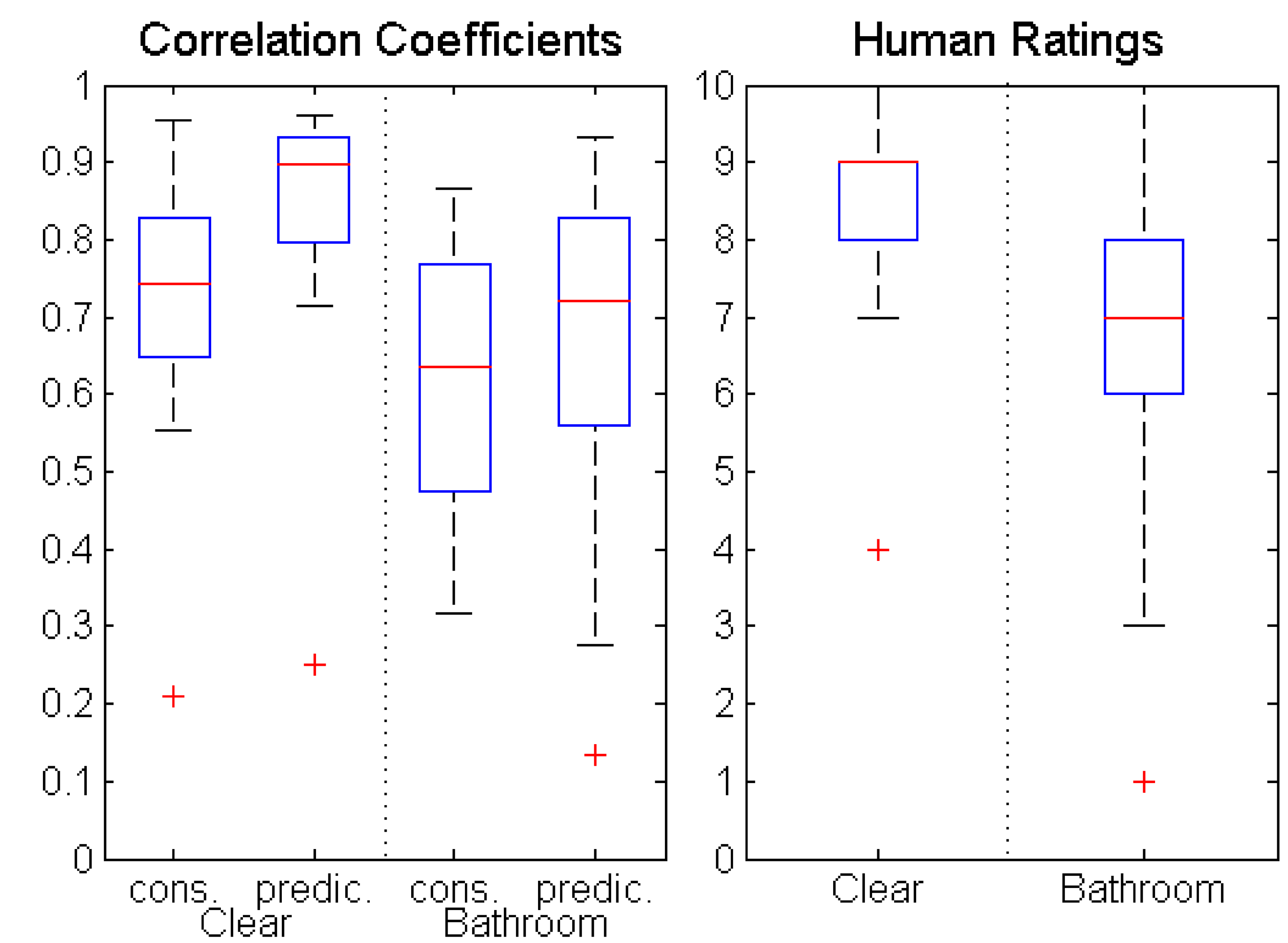


Figure: Left: distributions of *consistency* and *predictiveness* correlations for *Clear* and *Bathroom-like* ; Right: distributions of *human ratings* for *Clear* and *Bathroom-like*.

Boomy	*user 11*	*user 12*	*user 13*	*user 22*
Reverb. Time	0.01	-0.04	-0.10	-0.18
Echo Density	0.26	-0.08	0.24	0.01
Clarity	-0.43	0.10	0.14	**0.36**
Central Time	-0.33	-0.17	**0.69**	-0.32
Spec. Centroid	**-0.74**	**-0.58**	-0.15	0.17
predictiveness	0.90	0.77	0.86	0.79
human rating	7.0	10.0	8.0	8.0

Table: Correlations of the measures, *predictiveness* and *human ratings* for four different users for *Boomy* (highest absolute correlation value in bold, for each user).

Conclusion

Our study showed that people have different definitions of the same descriptor, and yet our system succeeds in learning rapidly and effectively the concept so that listeners are satisfied with the final controller. This supports our idea that individualizing controllers is a useful approach.

This work was supported by NSF grant number IIS-0757544.