Matching Artificial Reverb Settings to Unknown Room Recordings:
A Recommendation System for Reverb Plugins

Nils Peters [1,2], Jaeyoung Choi [1], Howard Lei [1]

[1] ICSI - International Computer Science Institute, Berkeley, CA, USA

[2] CNMAT - Center for New Music and Audio Technologies, UC Berkeley, CA, USA

[nils|jaeyoung|hlei]@icsi.berkeley.edu

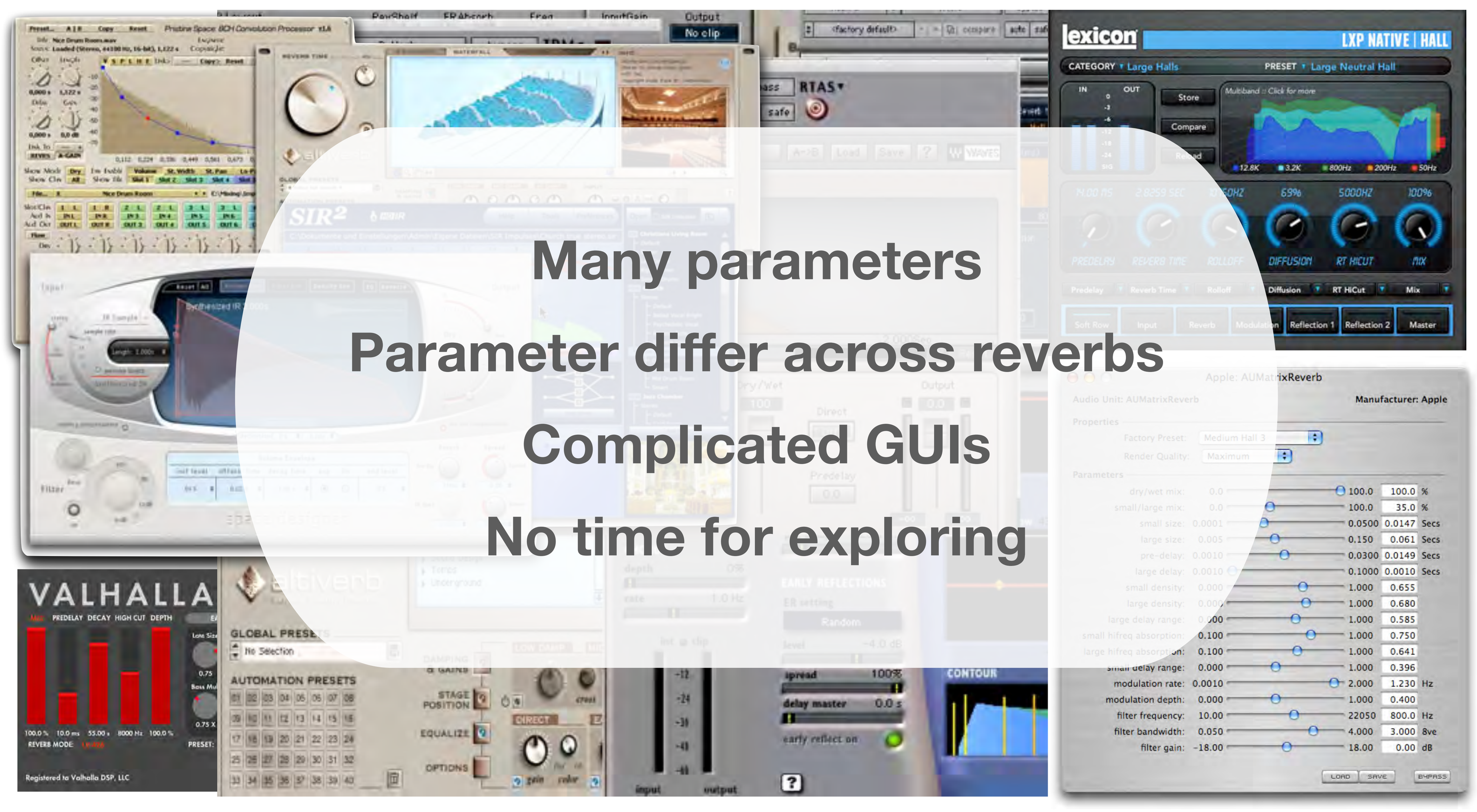
Abstract

For creating artificial room impressions, numerous reverb plugins exist and are often controllable by many parameters. To efficiently create a desired room impression, the sound engineer must be familiar with all the available reverb setting possibilities.

Although plugins are usually equipped with many factory presets for exploring available reverb options, it is a time-consuming learning process to find the ideal reverb settings to create the desired room impression, especially if various reverberation plugins are available.

For creating a desired room impression based on a reference audio sample, we present a method to automatically determine the best matching reverb preset across different reverb plugins. Our method uses a supervised machine-learning approach and can dramatically reduce the time spent on the reverb selection process.

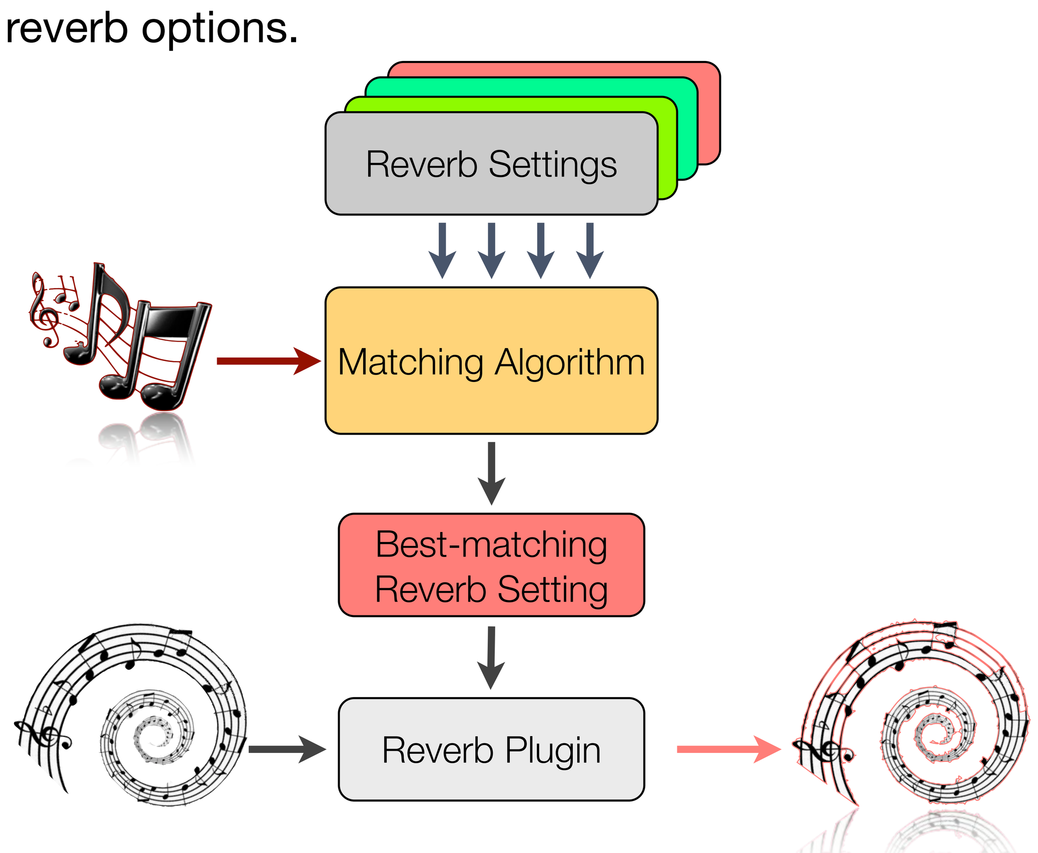
Problem: Reverbs Are Complicated



Our Solution

Often a user has an ideal reverb in mind that he is trying to create by means of an artificial reverb, e.g., a reverb heard in another recording.

Based on a provided reverberant audio file, we propose a system that automatically selects the best matching reverb setting from all of their available reverb options.



Prior Art

To simplify the workflow with artificial reverbs, a few control strategies have been proposed in the past:

Perceptual Parameter Layer

Rather than technical parameter, the user controls perceptual parameter (e.g., *room presence, intimacy*) that are linked to underlying technical parameter. *Ircam's Spatialisateur* pioneered this approach.

A Personalized Reverb UI

On an individual level, [1] described a method to train a personal interface on top of an artificial reverb. A linear regression model is used to map the parameters of the underlying reverb to the personal interface. This mapping can be learned by rating 35 reverberated audio samples in less than 3 minutes.

A Unified Reverb Interface

The authors of [2] propose a unified API based on Open Sound Control (OSC) to control different artificial reverbs from one single user interface.

References

[1] Z. Rafii & B. Pardo. *Learning to control a reverberator using subjective perceptual descriptors*, ISMIR 2009

[2] M. F. Zbyszynski & A. Freed. *Control of VST plug-ins using OSC.* ICMC 2005

[3] D. Reynolds, T. Quatieri, & R. Dunn. *Speaker verification using adapted gaussian mixture models.* Digital signal processing, 10(1- 3), 2000.

[4] J. Bonastre, F. Wils, & S. Meignier. *ALIZE, a free toolkit for speaker recognition.* ICASSP 2005.

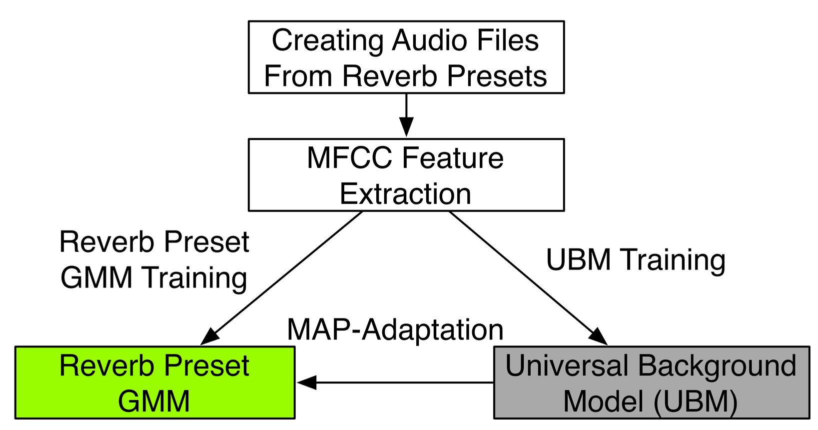
[5] P. Kenny & P. Dumouchel. *Experiments in speaker verification using factor analysis likelihood ratios.* In Proc. of Odyssey, 2004.

[6] http://www.1-1-1-1.net

The Matching Algorithm

Our recommendation system is derived from a GMM-based system historically used in speaker recognition [3] and uses *Mel-Frequency Cepstral Coefficient* (MFCC) audio features. From each audio file, MFCC C0-C19 along with deltas and double-deltas are extracted, 60 dimensions in total. The window length is 25 ms and frame intervals are 10 ms.

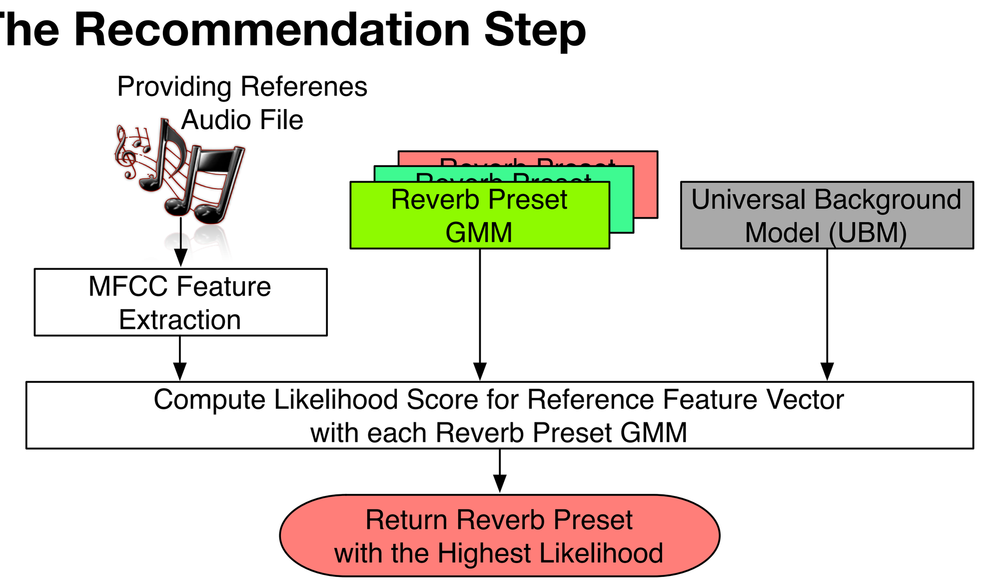
The Training Step



One *Gaussian Mixture Model* (GMM) is trained for each reverb preset, using a bag of randomly selected MFCC vectors from all the audio files processed with that reverb preset. The GMM training is done via MAP-Adaptation [3] from a reverb-independent GMM, which is trained using MFCC features from all audio tracks of all reverb-presets in the development set. A total of 128 mixtures are used for each GMM.

The open-source ALIZE toolkit [4] is used for GMM training, factor analysis implementation, and likelihood ratio computation for the reference audio. Factor analysis seeks to obtain a low-dimensional subspace representing the undesired variations of RIR-processed audio [5], with the RIRs coming from the same reverb preset. The undesired variations would be subsequently removed from each preset-related GMM and reference audio. Factor analysis is computationally intensive and can be done offline. The resulting model is stored and will be recalled when a reverb recommendation is requested.

The Recommendation Step



The same 60 MFCC features are extracted from the provided reference audio file and used to compute the likelihood ratio between each of the preset-related GMMs and the preset-independent GMM. The recommended reverb preset is the one with the largest likelihood ratio.

A Prototype

We created a prototype of the recommendation system for 97 monaural reverb presets. All of those reverb presets are available as IRs in the web [6]. The system was trained with 3880 reverberant audio files created with those reverb units from 40 anechoic speech files (20 seconds).

Apple Matrix Reverb

13 presets

TC Electronics M3000

53 presets

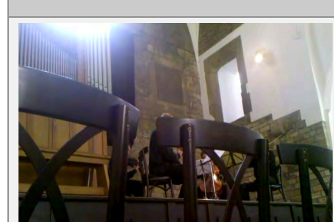
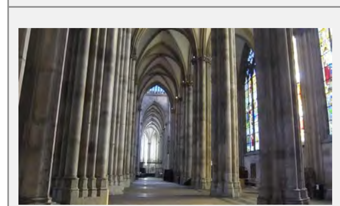
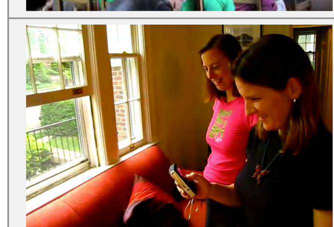
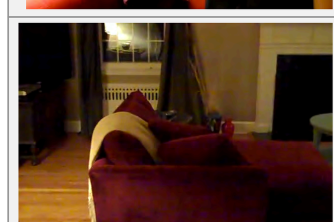
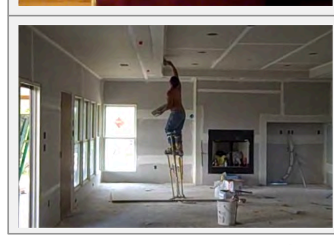
Lexicon MPX1

31 presets

Testing Results

Audio tracks of 18 videos from the Flickr video database were extracted. Half of these videos were tagged with *Living room*, the other half were tagged with *Church acoustics.*

The videos have a maximum length of 30 sec. and all were captured with consumer video cameras or mobile phones. In some of the footage, either the audio is clipping or an active gain compression notably affects the dynamic.

Flickr Video	5 best matching reverbs	5 least matching reverbs
	1. Plate Space 2. M SQ Garden 3. Large Hall 4. Vocal Plate 5. Tajma Hall	93. Long Swimming Pool 94. Recording booth 95. Room Large 1 96. Subway tunnel 97. Drum Booth
	1. Plate Space 2. Hall 4 PA 3. Large Hall 4. Vocal Plate 5. M SQ Garden	93. Walk in closet 94. Ambience 4 PA 95. Dialog Booth 96. Drum Booth 97. Subway tunnel
	1. Overhead Mics 2. Plate Space 3. Stage and Hall 4. Studio 40*40 ft 5. Large Hall	93. Phonebooth 94. Ambience 4 PA 95. Drum Booth 96. Dialog Booth 97. Subway tunnel
	1. Live VO booth 2. Small foley blue 3. Bright Room 4. Walk in closet 5. Wide garage	93. Hall Large 2 94. Warm Cathedral 95. Space Hall 96. Plate Space 97. Subway tunnel
	1. Bright Room 2. Small foley blue 3. Live VO booth 4. Right side garage 5. Claustrophobia	93. Frankfurt Main Station 94. Empty indoor pool 95. Plate Space 96. Warm Cathedral 97. Subway tunnel
	1. Small Hall 2. Subway platform 2 Small 3. Church 4. Wide garage 5. Band Rehearsal Room	93. Hall Large 2 94. Dialog Booth 95. Long Swimming Pool 96. Drum Booth 97. Subway tunnel

Example results of the recommendation system

133RD AES CONVENTION
October 26–29, 2012, San Francisco