Crowdsourcing a Reverberation Descriptor Map

Prem Seetharaman
Northwestern University
EECS Department
prem@u.northwestern.edu

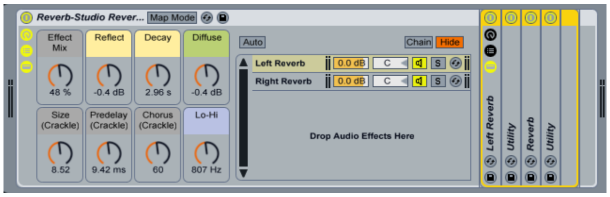
Bryan Pardo
Northwestern University
EECS Department
pardo@northwestern.edu

ABSTRACT

Audio production is central to every kind of media that involves sound, such as film, television, and music and involves transforming audio into a state ready for consumption by the public. One of the most commonly-used audio production tools is the reverberator. Current interfaces are often complex and hard-to-understand. We seek to simplify these interfaces by letting users communicate their audio production objective with descriptive language (e.g. "Make the drums sound bigger."). To achieve this goal, a system must be able to tell whether the stated goal is appropriate for the selected tool (e.g. making the violin warmer using a panning tool does not make sense). If the goal is appropriate for the tool, it must know what actions lead to the goal. Further, the tool should not impose a vocabulary on users, but rather understand the vocabulary users prefer. In this work, we describe SocialReverb, a project to crowdsource a vocabulary of audio descriptors that can be mapped onto concrete actions using a parametric reverberator. We deployed SocialReverb, on Mechanical Turk, where 513 unique users described 256 instances of reverberation using 2861 unique words. We used this data to build a concept map showing which words are popular descriptors, which ones map consistently to specific reverberation types, and which ones are synonyms. This promises to enable future interfaces that let the user communicate their production needs using natural language.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors; H.5.1 [**Multimedia Information Systems**]: Audio input/output; H.5.2 [**User Interfaces**]: User-centered design; H.5.5 [**Sound and Music Computing**]: Signal analysis, synthesis, and processing



Effect Mix 48 %
Reflect -0.4 dB
Decay 2.96 s
Diffuse -0.4 dB
Size (Crackle) 8.52
Predelay (Crackle) 9.42 ms
Chorus (Crackle) 60
Lo-Hi 807 Hz

Left Reverb 0.0 dB C S
Right Reverb 0.0 dB C S

Figure 1: A parametric reverberator from Ableton Live, a digital audio workstation.

General Terms

Human factors; crowd-sourced vocabulary; reverberation; audio

Keywords

Human computation; audio descriptors; audio synonyms; audio engineering; interfaces

1. INTRODUCTION

Audio production is central to every kind of media that involves sound, such as film, television, and music. It involves using tools such as parametric reverberators, equalizers, compressors, and limiters, to transform audio into a state ready for consumption by the public.

One of the most commonly-used audio production tools is the reverberator. In the physical world, reverberation is created by the reflections of a sound off of the solid surfaces (e.g. walls) of an enclosed space. These reflections result in a decaying series of echoes that modify the sound's loudness, timbre, and perceived spatial characteristics. In the digital realm, reverberation (reverb) can be simulated using networks of delays and gains to create a decaying series of echoes.

Reverberators allow the creation of echo effects. They can make the audio sound as if it were recorded in a different acoustic environment (e.g. change a recording made in a sound booth to sound like one recorded in a cathedral), and are used to increase the pleasantness of the sound. Nearly all commercially-recorded singing has reverberation added. People often describe reverb as making a sound "deep", "spacious" and "warm", although the exact relationships between these words and the changes in the control parameters for reverberators are not widely known.

Figure 1 shows the interface of a professional quality parametric reverberator. While the "effect mix" and the "decay"

MM'14, November 3–7, 2014, Orlando, Florida, USA.

ACM 978-1-4503-3063-3/14/11 ...$15.00.
http://dx.doi.org/10.1145/2647868.2654908.

dials may make intuitive sense, the other dials like "predelay" and "Lo-Hi" have no intuitively obvious meaning to the average person or to many musicians (e.g. acoustic and orchestral musicians). This is because the controls are conceptualized in terms of the underlying processes used to create the reverberation effect, as opposed to perceptually relevant terms that people may commonly use to describe reverberation, like "boomy." The result is that musicians without technical expertise can spend a great deal of time stumbling through numerous parameter settings, disrupting the creative process. Musicians should not have to reconceptualize their ideas in terms of a fixed interface with esoteric parameters.

Musicians, both amateur and professional, often conceptualize creative goals for production in terms of natural language that may not have obvious, clear mappings onto the controls available on the audio production tools. Much of the work of audio production involves bridging the gap between artistic goals that are expressed in natural language ("Make the guitar sound 'bigger"') and the tools available to manipulate the sound (e.g. the controls of a parametric reverberator).

We seek a vocabulary that makes audio production interfaces accessible to laypeople, rather than experts. We do this by finding mappings between commonly used descriptive words ("boomy") and the hard-to-understand controls of production tools ("predelay"). The point of this is to develop interfaces that give novices an easy point of entry into audio production. This will support the creativity of acoustic musicians without forcing them to learn interfaces with opaque and esoteric controls. What is more, mappings between the vocabulary of non-technical people and the parameters of production tools may give expert audio engineers an easier way to communicate with their clients. The following quote from Jon Burton, a respected audio engineer, illustrates the communication problem many audio engineers face.

"The idea had sprung from a problem that has arisen in studios ever since the beginning of the recording age: how can you best describe a sound when you have no technical vocabulary to do so? It's a situation all engineers have been in, where a musician is frustratedly trying to explain to you the sound he or she is after, but lacking your ability to describe it in terms that relate to technology, can only abstract. I have been asked to make things more 'pinky blue', 'Castrol GTX'y' and... 'buttery'." [5]

In this work, we describe SocialReverb, a project to crowdsource a vocabulary of audio descriptors that can be mapped onto concrete actions using a parametric reverberator. Using the data collected by SocialReverb, we can create a map that places audio descriptors in relation to one another. This allows us to answer questions about the relationships between descriptors in terms of how they map onto reverberation. We can find descriptive terms whose definitions vary across users, and words that have high agreement between users.

A crowdsourced concept map of terms relating to reverberation would also allow the creation of a reverberator that responds to simple commands in plain English ("make the sound boomy"). To achieve this goal, the tool must be able to determine whether or not the stated goal can be achieved using the selected tool (e.g. making a violin sound "warmer" using a panning tool does not make sense). It should also know what actions must be taken, given the correct tool ("Use a parametric reverberator with an RT60 of 4 seconds and a cutoff frequency of 100 Hz to make the drums rumble warmly"). Further, it should be aware of possible variations in the mapping between word and audio among users (Bob's "warm" $\neq$ Alice's "warm". The concept map learned by SocialReverb is key to learning these mappings and building a plain English interface for reverberation tools.

2. BACKGROUND AND RELATED WORK

There has been much prior work on learning descriptive terms for audio. One common approach to creating a dictionary of descriptors is that of using using text co-occurrence, lexical similarity and dictionary definitions (e.g. WordNet [12]). Approaches based strictly on text are not applicable to our task, because they do not provide mappings between words and measurable sound features or control settings for audio manipulation tools.

Psychologists have explored the mappings between descriptive terms and measurable signal characteristics for sound. Some terms, specifically terms that relate to pitch (high, low) and loudness (soft, loud) have relatively well understood [8, 19] mappings onto measurable sound features. Many other terms, however (e.g. "muffled" sound) do not have simple correlations that have been identified by psychologists.

There have been numerous studies performed in the past half century that hope to find universal sound descriptors that relate to a set of canonical perceptual dimensions [7, 11, 21, 23]. In the past decade, researchers from many different backgrounds, such as recording engineering [9], music composition [20], and computer science [17], have sought a universal set of English terms that describe sound. [22] extracted features from onomatopoeia recordings and computed distances between words, similar to this work. They embedded these distances into a 2D space, similar to the one in Figure 7. This work, however, addresses the distances between onomatopoeia, rather than reverberation. It also has no crowdsourcing element, as the words and sounds were generated by 4 lab members. Our work involves hundreds of users across thousands of sessions, contributing thousands of words.

These studies have varied from finding a vocabulary of descriptors for sound in general to finding a specific set of descriptors for particular instruments. The typical approach is to start by determining a set of natural descriptors by performing a survey. The descriptors provided by participants of the survey are then mapped onto sound measures such as spectral tilt, sound pressure level, or spectral centroid. The research community, however, has not focused on learning vocabularies of words that map to actionable sound manipulations using audio production tools. They have also focused on words to describe timbre, rather than the effect of reverberation. Our work is distinct in both these regards.

Recording engineers [9] use a few specific terms that are used to describe effects produced by recording and production equipment which are straightforward to map onto measurable sound properties. In the case of reverberation, "wet" is, perhaps, the best-known term and refers to the "wet/dry" mix control found on many reverberators. This gives the amplitude ratio between the direct audio signal and the reverberated signal. The "wetter" the sound is, the more reverberation there is. Unfortunately, many descriptive terms applied to reverberation (e.g. "boomy" and "warm") do not map clearly onto single reverberation parameters and the general population of acoustic musicians do not share the vocabulary of recording engineers. Our goal is to discover the vocabulary of this more general population.

The SocialEQ project is the most similar work to the current project. SocialEQ was a project to crowdsource sound adjectives relevant to parametric equalization [6]. It creates a dictionary of words defined both in terms of subjective experiential qualities and measurable properties of a sound. The current work is distinct in that it crowdsources a vocabulary of audio descriptors relevant to reverberation, rather than equalization. Further, the collection technique described in [6] is a time-consuming task that involves rating 40 equalization settings to learn a single word. Their task takes around 15 minutes per word. This limited the size of the resulting vocabulary learned. The simple task in the current work takes just 2 to 3 minutes for multiple words, letting us learn many more word associations.

[14] describes a reverberator that can be controlled through measures of reverberation (RT60, echo density, clarity, central time, spectral centroid). This reverberator is the one used in this work. We also map audio descriptors to these same 5 measures of reverberation. [15] expands on this reverberator to develop a system for learning the settings of a parametric reverberator by users teaching words to the system and rating examples, like in [6]. However, [15] is limited to only a few words that were given to the users to teach to the system (bright, clear, boomy, church-like, bathroom-like). Our system expands and greatly extends this approach by collecting far more audio descriptors (thousands, instead of five) from far more people (hundreds, instead of dozens), priming the pump for a future system where relevant words can be taught to the system using fewer audio examples.

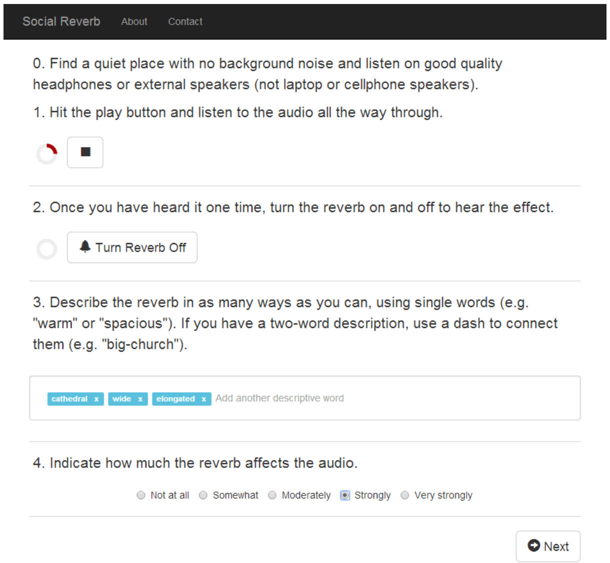
3. SOCIALREVERB

Simplifying interfaces, such as the controller shown in Figure 1, to a simple dial that makes the sound more "church-like" or less "church-like," or more "rumbly" or less "rumbly" requires the collection of a vocabulary that we can map onto actionable changes made by a parametric reverberator. However, it is not immediately obvious what words actually describe reverb and how those descriptions map on to changes in signal statistics or the parameters controlling a reverberator.

We address this problem by asking a large number of users to describe the difference between pairs of audio recordings: an original recording, and that same recording after reverberation is applied. This creates a folksonomy of words that relate to reverberation. It also gives us the data needed to map these words into both a feature space describing the audio and the space of parameter settings for a reverberator.

3.1 The Software

To deploy the software to a large audience for data collection, we have implemented it as a web application using Web Audio API, an experimental low-level audio processing API implemented in Chrome, Opera, and Safari. We implement the reverberation unit with Web Audio API using the built in Gain nodes, Delay nodes, and Lowpass filter, so all of the audio processing required for the reverberation is done within the user's browser. The software is deployed on Google App Engine.



0. Find a quiet place with no background noise and listen on good quality headphones or external speakers (not laptop or cellphone speakers).

1. Hit the play button and listen to the audio all the way through.

2. Once you have heard it one time, turn the reverb on and off to hear the effect.

3. Describe the reverb in as many ways as you can, using single words (e.g. "warm" or "spacious"). If you have a two-word description, use a dash to connect them (e.g. "big-church").

cathedral x wide x elongated x Add another descriptive word

4. Indicate how much the reverb affects the audio.

○ Not at all ○ Somewhat ○ Moderately ● Strongly ○ Very strongly

Figure 2: First part of survey for participants. Here they contribute words freely, and rate how much the reverb affects the audio.

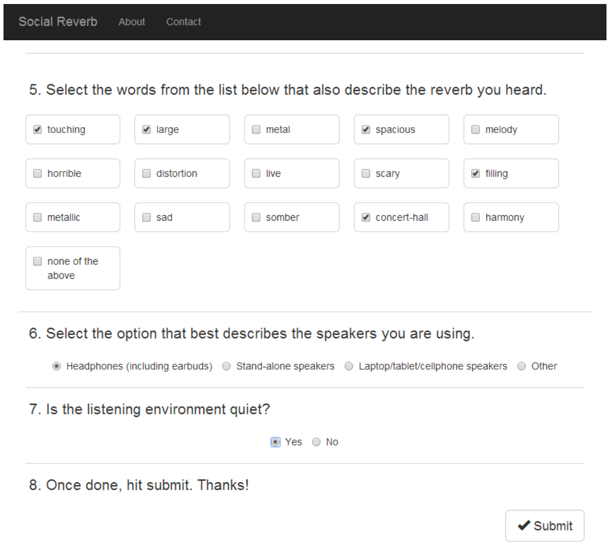
3.2 The Interaction

Participants are recruited through Amazon's Mechanical Turk [2]. Once someone agrees to participate, they are presented with a brief (30 second) listening test to ensure they have a listening environment conducive to hearing reverberation (see Section 4.2 for details).

If they pass the listening test, they begin the task and are presented with a looped 10-15 second audio recording of a musical instrument (drums, guitar, or piano) recorded without reverberation. When the audio example has played through once, the 'Turn Reverb On' button is enabled, letting the user turn the reverb on and off to hear the effect of the reverberator. The user can only add or remove reverb. No manipulation of the reverberator parameters is allowed. Once they have toggled the reverb on and off and listened for at least another iteration of the audio example, they are asked to provide a list of words that describe the effect of the reverberator. We encourage the user to use single words, like "warm" and "spacious", but also allow them to use multi-word descriptions by connecting them with a dash.

We ask each participant to contribute words before being presented words other participants contributed to avoid any premature convergence of user vocabularies. This avoids a narrowly focused resulting vocabulary. Since we collected 2861 unique words, we achieved this goal. The prompt in the task mentions three words (warm, spacious, and big-church). While one might expect these words would end up over represented in the final data set, these are just three words out of 2861 provided by participants and none of these three words ended up in the top ten words, ranked by consistency. We then ask them to rate how much the reverb affects the audio, using a Likert scale with values 'not at all', 'somewhat', 'moderately', 'strongly', 'very strongly'.

If there are prior user responses for the current reverberation settings, we then present 15 random words from the set of words previously used to describe the reverb. We ask



5. Select the words from the list below that also describe the reverb you heard.

☑ touching ☑ large □ metal ☑ spacious □ melody

□ horrible □ distortion □ live □ scary ☑ filling

□ metallic □ sad □ somber ☑ concert-hall □ harmony

□ none of the above

6. Select the option that best describes the speakers you are using.

● Headphones (including earbuds) ○ Stand-alone speakers ○ Laptop/tablet/cellphone speakers ○ Other

7. Is the listening environment quiet?

● Yes ○ No

8. Once done, hit submit. Thanks!

Figure 3: Second part of survey for participants. Here they are presented with a list of words and asked whether or not they agree that they describe the reverberation effect. They are asked two questions about their listening environment for inclusion criteria purposes.

them to select words from this list that are good descriptors of the reverberation, or indicate that none are good descriptors. An important note is that this set of words is selected randomly to avoid any bias in the resultant data. If a participant contributes a new word to the data in the first part of the survey, it now has a chance of being presented to other users.

Finally, we ask two questions about the listening environment. We ask what sort of speakers the user has (headphones, stand-alone speakers, laptop/tablet/phone speakers, other) and whether the listening environment is quiet. These are used as inclusion criteria (see Section 4.2). One session consists of repeating this process 5 times, labeling 5 reverberation settings. Each reverberation setting in a single session has a random audio file associated with it. The audio changes between the piano, drums, and guitar sample. In the Amazon Mechanical Turk task used to collect the data, we restrict users to at most 3 sessions.

3.3 The Reverberator

To implement the data collection, we needed a reverberator able to generate a wide variety of impulse response functions on the fly. Rather than use a convolution reverberator, which selects from a fixed library of precomputed impulse responses, we used a digital stereo reverberation unit inspired by Moorer's work ([13]), described in [15] and [14] and seen in Figure 4. The developed reverberator is a version of a well known state-of-the-art algorithm, comparable to reverberators found in professional digital audio work stations such as Ableton, Cubase, etc.

The reverberator is controlled with 5 parameters: the delay and gain of the first comb filter, the delay between the channels of the all-pass filter, the cutoff frequency of the low pass filter and the gain of the overall reverberation. The reverberation is added back to the dry signal to produce the modified signal. The reverberator uses six comb filters in parallel to simulate the complex modal response of a room by adding the echoes together. The delays and gains of the other five comb filters are derived from the delay (d_1) and gain (g_1) of the first comb filter. After this is computed, the signal is doubled into a left and right channel. The left channel is delayed by $d_a + m$ seconds, while the right channel is delayed by $d_a - m$ seconds, where d_a is .01 sec, and m ranges between 0 and 12 msec. This creates a slight stereo effect. The signal is then put through a low-pass filter with a cutoff frequency f_c to simulate air and walls absorption. Finally, a gain parameter G, controls the wet/dry effect.

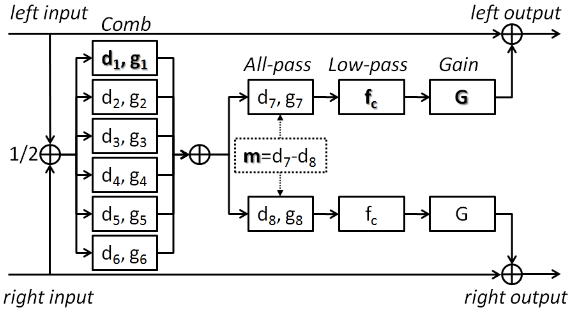


Figure 4: The digital stereo reverberation unit.

These five parameters (d_1, g_1, m, f_c, G) provide us direct control over the sound of the reverberator. To make data collected using this reverberator useful for other reverberators with different control parameters, we use the methods described in [14] to map these five control parameters to five signal measures of the resulting impulse response function. The signal measures are:

1. *Reverberation time (RT60)*, the time it takes for the reflections of a direct sound to decay by 60 dB below the level of the direct sound [1].
2. *Echo density*, the number of echoes per second at a time t.
3. *Clarity*, the ratio in dB of the energies of the impulse response before and after a given time t, indicating how "clear" the sound is.
4. *Central time*, the time of the center of the energy in the impulse response.
5. *Spectral centroid*, the frequency of the center of energy in the magnitude spectrum of the impulse response.

By characterizing the reverberation in this way, results from this study can be used for any reverberation unit where a mapping has been made between the control parameters and the resulting impulse response characteristics, regardless of the construction of the reverberator. This makes the learned data highly generalizable.

In selecting reverberation settings to present, we chose a set of 1024 impulse response functions that evenly cover a wide range of reverberations. The reverberation time ranged from .5 to 8 seconds, the echo density from 500 to 10000 echoes/sec, the clarity from -20 to 10 dB, the central time from .01 to .5 sec, and the spectral centroid from 200 to 11025 Hz.

From this set S of 1024 settings, we estimate a maximally varying subset P of 256 settings as described in [16]. First, we select a random setting $s \in S$ and initialize P to include s. Then we search through S to find the next setting that maximizes the variance of P, if included. That is,

$$s_{next} = \arg\max_{s \in S} v(s)$$

, where $v(s) = \frac{1}{D} \sum_d std_d(P \cup f)$. Here, d is a dimension (one of the five signal measures) and D is the number of dimensions (5). We normalize each dimension so one does not dominate the sum. This process walks through the space of reverberation settings, finding ones that are maximally different from ones that we have already chosen. This set of 256 widely-varying reverberation settings was used for all sessions in the study.

For each session with a participant we select a setting from P as described in 4.3, and present it to the user. Our selection method ensures even coverage across P, and the nature of P is such that a single user is unlikely to get two similar reverberation settings in a single session.

3.4 The Audio

We selected three monophonic (as opposed to stereo) dry (no reverberation) signals for source audio that are representative of sounds that would be used in real musical projects. The first is a dry electric guitar sound with no effects processing, playing a 10 second chord progression. The second is a 16 second passage from Bach's Chaconne in D minor recorded with a dry piano sound, created using a dry sampled piano from the East West Quantum Leap Pianos Gold library. The last is a 14 second recording of a drum kit performing in a rock style, taken from a studio recording used in a commercially released album. Each of these signals is representative of real-world music. We chose three very different signals that each broadly cover the frequency range. In the "agreement" part of the survey, we present words from other sessions with the same reverberation, but applied to all 3 sounds. High-agreement words are likely in the intersection of descriptions of three audio signal/reverberation combinations. As the three audio signals are very different, we see the words in the intersection as descriptions of the effect, rather than the audio signal.

The audio was recorded with a sample rate of 48000 Hz. We then compressed each audio file as a high-quality .mp3 file using the LAME [10] .mp3 encoder at a bitrate of 320kbps. This bitrate has been shown to produce audio that listeners over the internet find indistinguishable from uncompressed compact disc quality audio [3]. Compression was done to limit the bandwidth cost of sending the audio to the client's computer. Once the audio is transferred to the client, the audio is expanded again to a 16-bit pulse-code-modulated (PCM) audio buffer. Reverberation is then applied locally to the audio in the buffer.

4. DATA COLLECTION

4.1 Participant Recruiting

We recruited participants through Amazon's Mechanical Turk. We paid participants $.50 (USD) for every 5 audio examples described, if they pass certain inclusion criteria. Participants could elect to perform the task from 1 to a maximum of 3 times. A single Amazon Mechanical Turk user thus describes 5, 10, or 15 reverberation settings.

4.2 Inclusion Criteria

We ensure cooperative and attentive participants in a variety of ways. Amazon Mechanical Turk provides measures of worker reliability that may be used to pre-screen participants. For this study, we only allowed workers wtih a 97% positive review rating and who have performed at least 1000 tasks on Mechanical Turk.

Much of the effect of reverberation occurs in low frequencies (below 100 Hz). Many laptop speakers are not able to reproduce sounds below 100 Hz. To ensure quality contributions, participants were asked to take a listening test prior to performing the task. The test randomly selects 2 audio files from a set of 8. These audio files consists of high, midrange, and low tones in some random sequence. The high tones were selected to be audible on any speaker. The midrange tones were selected to be slightly audible on laptop, phone, or tablet speakers, but clear on any quality speakers or headphones. The low tones were selected to be completely inaudible on laptop, phone, or tablet speakers. The user is asked how many tones they heard clearly in each audio file. The correct answers vary between 1 and 8 depending on the audio file selected. The user is given three chances to pass the speaker test. If they fail, they cannot participate.

Remaining participants are filtered further. We measure the total amount of time spent listening to the audio example, with the effect on and off. If the user listens to the clean audio for less than the length of the audio file, we exclude them from the data. If the user listens to the audio with the effect on for less than the length of the audio file, we exclude them from the data. We also removed sessions where the participant answered no to the question: "Was the listening environment quiet?" Finally, we only include participants who self-report listening on "headphones (including earbuds") and "stand-alone speakers", excluding those listening on "laptop/tablet/phone speakers" or "other."

4.3 Experimental Design

There are 3 possible audio recordings and 256 possible reverberation settings, making for a total of 768 combinations. The audio examples are selected randomly for each session. The reverberation settings are selected carefully for each session to ensure even coverage of the 256 settings. Each time the system selects a reverberation setting to present, a count for that reverberation is incremented by one. When we select a reverberation setting, we select randomly amongst the settings with the minimum count in the database. This is equivalent to random selection without replacement until all reverberation settings have been used an equal number of times.

Each time a participant performs the task, they are asked to describe 5 reverberation settings (see Section 3.2 for a description of the interaction). A single user may perform the task at most 3 times, describing at most 15 reverberation settings. Recall there are 768 combinations of reverberation and audio file to be labeled. This requires a minimum of 52 (if each describes 15 reverberations) and a maximum of 154 (if each describes only 5 reverberations) participants to ensure each combination was labeled at least once.

5. RESULTS

As of this writing, 513 unique users have described 256 instances of reverberation using 2861 unique words. We have made the data available for use by the research community at `interactiveaudiolab.org/data/socialreverb`. We have also included audio examples, where a descriptor is applied to one of the dry sounds described in 3.4.

5.1 Definitions

1. **example**: A reverberation applied to an audio file for a participant to describe.
2. **descriptor**: An adjective used to describe the effect of reverberation on one or more examples.
3. **session**: A participant takes the SocialReverb survey, providing descriptors for 5 examples. Participants may perform up to 3 sessions.
4. **descriptor instance**: Each time a descriptor is entered in the free-response question (Question 3 in Figure 3.2) or a user agreed that it described a reverberation from the list of words (Question 5, Figure 3.2).
5. **reverberation measure vector**: a vector of measures of reverberation described in Section 3.3. Each descriptor instance is associated with a corresponding reverberation measure vector.
6. **reverberation parameter vector**: a vector of control parameter settings for the reverberator described in Section 3.3. Each descriptor instance is associated with a corresponding reverberation parameter vector.
7. **descriptor definition**: The set of reverberation measure vectors that share a common descriptor, and built as described in 5.3. The preciseness of the definition depends on the normalized variance in the set of measure vectors.

5.2 The descriptors

There were a total of 1074 sessions, in which we collected 14628 descriptor instances from 513 users. These 14628 descriptor instances represented 2861 unique descriptors. Of the 2861 descriptors, 1791 had at least 2 instances. The most popular descriptor by far was *echo*. Other popular terms include *warm* and *spacious*. While these words were expected, we now have correlations between specific reverberation effects and the words used to describe them. This is a unique contribution as no prior study has done this for a vocabulary of more than 5 descriptors. The top ten most common descriptors are shown in Table 1. These words are those one would expect to crop up in a discussion of reverberation. Interestingly, they are not those with the most specific definitions, in terms of the measured qualities of the reverberation, as we shall see in Section 5.4.

5.3 Representing reverberation concepts

Recall that these words were elicited from users in a format that asks them to describe the change to a sound when the reverb is applied, as opposed to an absolute measure of what is a "warm" sound. This lets us correlate the actual signal changes caused by the reverberator to the descriptive term elicited by that reverberator.

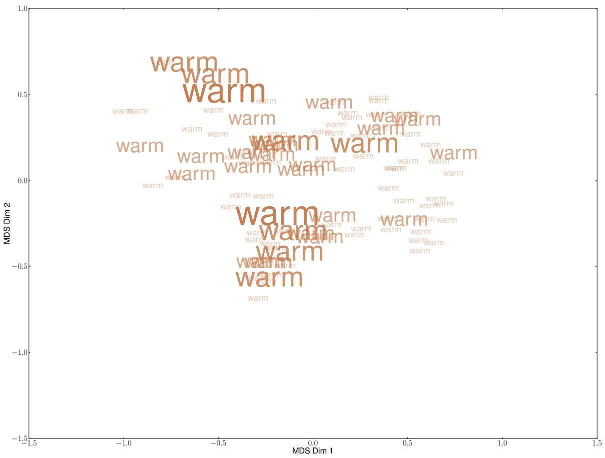


Figure 5: A map of reverb measure vectors for "warm". Font size encodes how much agreement there was on the word for that specific reverb measure vector. "Warm" is a less consistent descriptor, than "underwater," as can be seen by its spread across the map above.

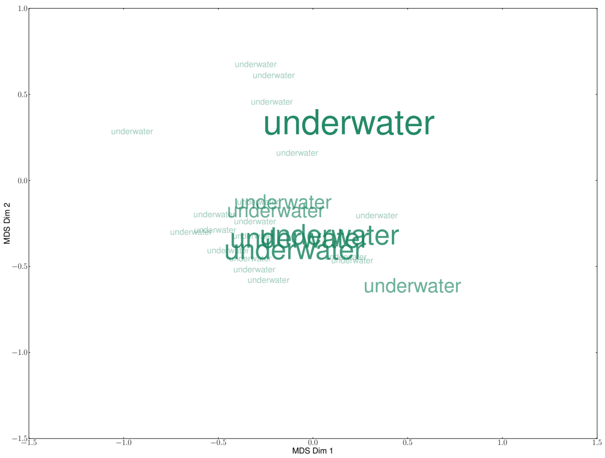


Figure 6: The same visualization technique as in Figure 5 for "underwater", rather than "warm". This map is more consistent than Figure 5, as it has much less spread across the map.

Rank	Word	Instances
1	echo	1220
2	distant	662
3	warm	596
4	spacious	596
5	loud	538
6	muffled	451
7	deep	423
8	church	327
9	echoing	312
10	big	255

Table 1: Top ten words in database by number of descriptor instances.

Using the parameters and measures associated with each descriptor (see Section 3.3), we can calculate an actionable definition for the descriptor. Each instance of a descriptor is associated with an example that had a particular reverberation applied. To define the actionable meaning of a descriptor, we add the associated reverberation measurement vector (consisting of reverberation time, echo density, clarity, central time, and spectral centroid) to the set of associated measurement values.

Once we find all reverberation measures associated with a word (e.g. "warm"), we take the average of each measure and use the resultant vector as a first approximation of the descriptor definition. Here, by definition, we mean the changes in the measurable signal qualities that a typical listener would label as making the sound "warm." This gives us a definition that can be used to change a sound to make it more "boomy" (or "warm", or some other word). This can be done using the same technique applied in [14].

5.4 Consistent reverberation descriptors

Some words are more specific than others. For example, "echo" or "echoing" is a very broad term that we know from the data is applied to many varied examples of reverberation. While it is true that taking the average, as described in the previous section will result in something many would call "echoing," we seek words that would let a user have more nuance than merely having or not having echo effect. Therefore, we would like to establish how specific reverberation descriptors are. For example, does the word "boomy" describe a specific range of reverberation settings, or is it used more generally? To answer this, we calculate the normalized variance within each descriptor's definition.

A descriptor definition is the set of reverberation measure vectors (one five-element vector per instance) associated with the descriptor. Each of the measures is normalized to the range from 0 to 1, so that one measurement (e.g. RT60) does not dominate the calculations. For each descriptor, we then calculate the within-measurement variance (e.g. variance of normalized RT60 for all instances of "boomy"). The variances for all five measures (e.g. the variance of RT60, echo density, clarity, central time and spectral centroid) are then averaged for each descriptor. The top 10 descriptors that had at least 15 instances in the database each are shown in Table 2 , ranked by variance of descriptor definition. Note that none of the top 10 highly-specific terms is in the top 10 most-used descriptors. This indicates that the most widely-used words may contain less specifically applicable information than many less-used words.

Figures 5 and 6 are a visual representation of consistency. They were generated by gathering every unique reverberation measure vector associated with the descriptor, and performing multidimensional scaling [4] on the resultant data. Here, the size of a word indicates the number of instances where that descriptor shared the same specific reverberation measure vector. The larger the text, the more times that specific reverberation setting was given this label. Less consistent maps have the descriptor spread across the map. This means many different reverberation settings elicited the same description. More consistent maps have descriptors more focused in specific areas of the map. This indicates the word has a specific meaning, particular to a certain kind of reverb.

Rank	Word	Normalized Variance	# of Instances
1	chaotic	.17452	26
2	underwater	.19031	70
3	watery	.20074	29
4	boomy	.2014	31
5	distorted	.20172	235
6	messy	.20595	35
7	haunting	.21077	157
8	broad	.21518	26
9	overdone	.21743	18
10	ominous	.21839	84

Table 2: Top ten descriptors with at least 15 instances, sorted by variance of the definition. Lower variance means higher cross-participant agreement in the definition of the word.

Figure 5 shows that some words have multimodal definitions not captured by the averaging scheme we use for the descriptor definition. We can see that "warm" appears to group into 3 clusters. This would indicate that there are 3 different "warm" reverberations in the data. This is in contrast to Figure 6, which has a single definition, since the distribution of reverberation settings is clearly centered on one location.

5.5 Mapping the reverb word space

A reverberation setting is described by the five-parameter signal descriptor of the reverberation described in Section 3.3. We can thus place any word-label applied by a participant to a reverberation setting in a five dimensional-space, defined by the signal descriptor values.

We can visualize the reverberation measure space by using descriptor definitions. For each descriptor with 15 or more instances, we calculate a definition and measure the variance of the definition as described in Section 5.4. From the resultant descriptor definitions, we calculate pairwise Euclidean distances. We use these distances to project the original five dimensional space onto two dimensions, using multidimensional scaling [4]. Each descriptor is scaled using its variance. The resultant reverberation descriptor map is shown in Figure 7. To rephrase, word position is the center of the distribution of reverberation settings associated with the word. Word size is associated with the variance of the distribution. Larger words indicate greater consistency (less variance) among participants.

5.6 Audio descriptor synonyms

Distance between descriptor definitions is measured using Euclidean distance in the space of reverberation measures. We consider two descriptors synonyms if the pairwise distance between their descriptor definitions is within the first percentile of all pairwise distances between descriptor definitions. We restrict the set of possible words to those whose descriptor definition variance score falls in the lower 50th percentile of all descriptor definition variance scores. We only create a descriptor definition for a word if it was mentioned in the database (either agreed with or contributed freely) at least 15 times.

Table 3 shows synonyms of some high consistency descriptors found through comparing descriptor definitions.

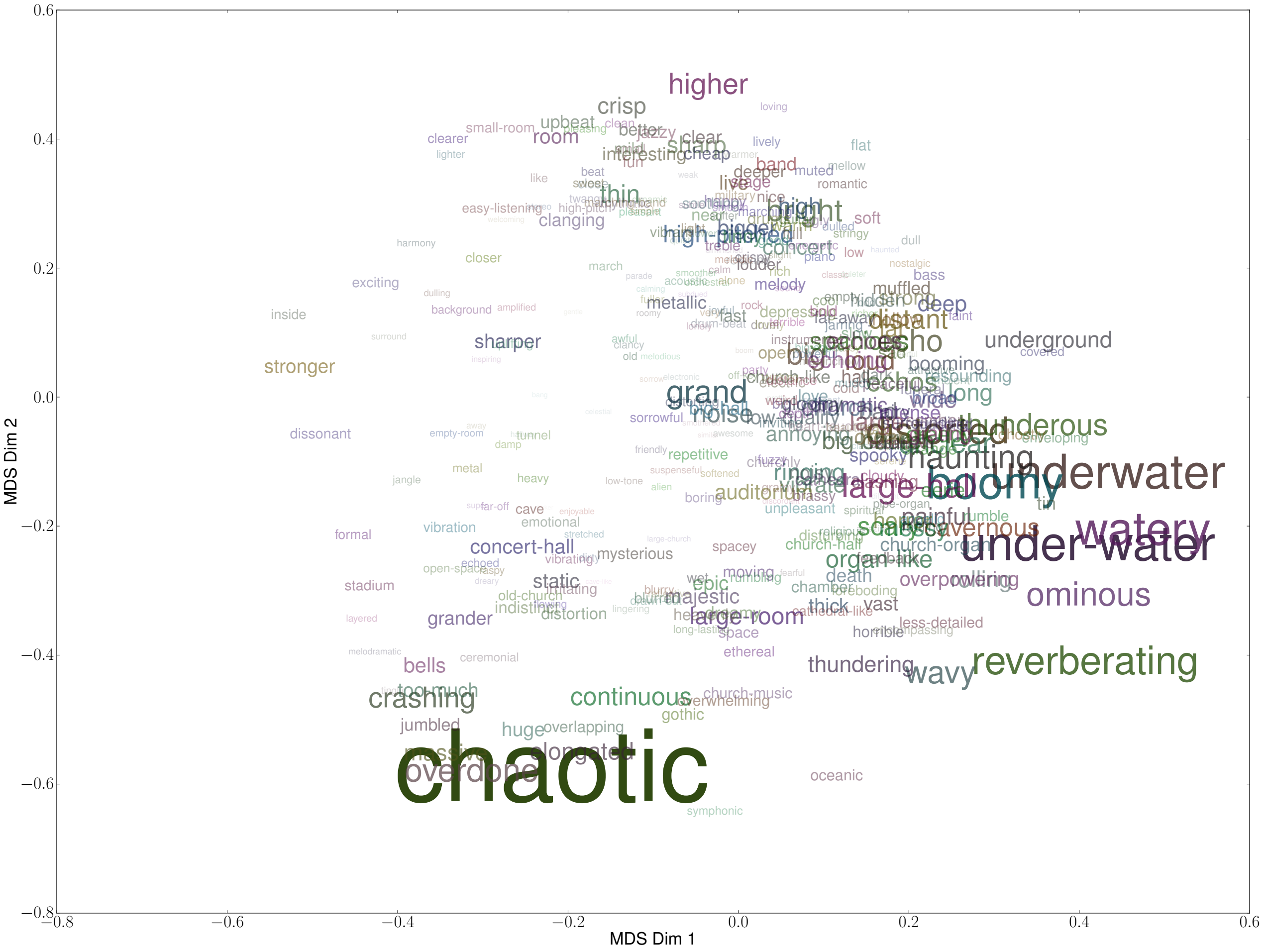


Figure 7: A visualization of the reverberation descriptor space. The map was created using multidimensional scaling to project the data onto the two dimensions shown. The font size of a word inversely correlates to the variance in its definition among participants. Big words that are close together can be interpreted as reliable audio synonyms for reverberation.

Descriptor	Synonyms
distorted	big-church, strange
harsh	big-church, bad, strange
echo	loud, spacious, echos
spacious	echo, distant, echoing, far, slow, cool, jumbled
deep	distant, hollow, bass
church	dramatic, dark
far	distant, spacious
wide	airy, peaceful
ringing	annoying, spooky, intense
hollow	distant, deep, far-away, muffled, bass, thick
big-church	distorted, harsh, surrounding, strange

Table 3: Some descriptors with low normalized variance and their synonyms.

5.7 Parameter space versus measure space

To this point, all figures and tables have shown words in the space of measured signal statistics for reverberated sounds. We did this so that the data would be usable across multiple kinds of reverberators that do not share the same set of control parameters. A question may arise, however, about whether new insight can be gained by building a map of descriptors using the distance in parameter space rather than measure space. The resultant map is shown in Figure 8.

The parameters control the reverberation directly, manipulating the coefficients and delay times of the reverberator shown in Figure 4. The measures of the impulse response generated by the parameters are what we use to measure distance between reverberation effects. In Figure 7, the words are more evenly spaced around the environment, and the space can be divided cleanly into a few different types of reverberation by looking at it. Around “chaotic” falls reverberations in large halls, to the point of distortion, like in a parking garage. Around “underwater” falls calmer reverberations in slightly smaller halls. Above “underwater” are stranger reverberations (“clashing”, “distorted”) in small

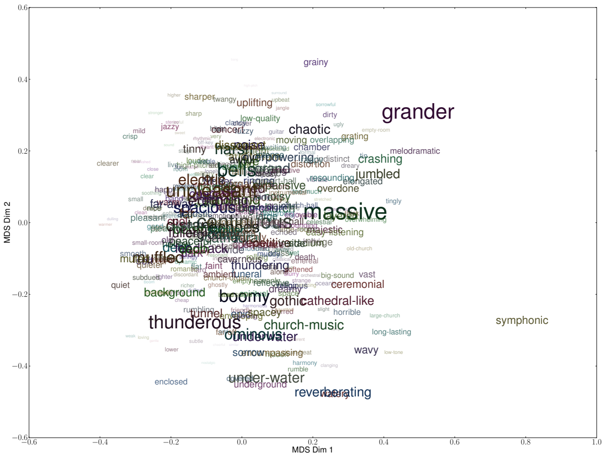


Figure 8: The same visualization technique as in Figure 7, except now done in parameter space rather than measure space

halls (bathrooms, stairwells, perhaps) and above that are the more realistic reverberations, that sound like churches and auditoriums. Around "upbeat" exist reverberations that sound like studio recordings, or very small rooms.

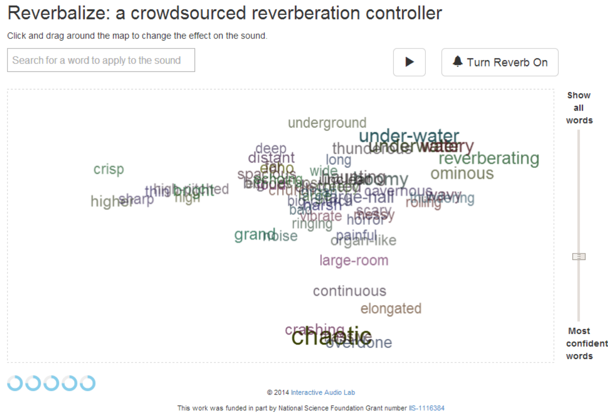
In parameter space, shown in Figure 8, we cannot draw these divisions so easily. The words are clustered together much more closely and evenly. While some of the relationships in measure space are somewhat preserved ("massive", "jumbled", "crashing" are still near each other), many are not. "Upbeat", for example, is now close to "low quality" and much closer to "chaotic" than it was before. This suggests that there are larger perceptual discontinuities in parameter space than in measure space. From this map, we see that operating a parametric reverberator using the measures of the reverberation, rather than direct control of the system that produces the reverberation is more predictable, making it easier to use. Indeed, many parametric reverberators behave this way, such as the one in Figure 1. This is, in fact, what makes them difficult to use and is, in part, a motivation for this work.

6. FUTURE WORK

We have used the data collected to implement a novel reverberation controller [18], shown in Figure 6. Words are projected onto a map, using a mapping from their 5 dimensional descriptor definitions to the 2 dimensional space of the map. Users traverse the map to explore effects, using the words as a guide. They can also search for a word, and if it's in the map, it's applied to the audio.

This interface lays the groundwork for a future validation study on the effectiveness of an interface built using the data from the current work. In this validation study, we will focus on a population of acoustic musicians who have musical ideas and goals, but are not the typical population of tech-savvy production engineers and electronic musicians that are the typical user base of existing audio production tools. We will compare our vocabulary-map interface to a traditional parametric interface, where both interfaces control the same underlying reverberation tool.

Participants will be given two kinds of production tasks. In the first task, an audio file with reverberation already


Reverbalize: a crowdsourced reverberation controller

Click and drag around the map to change the effect on the sound.

© 2014 Interactive Audio Lab

This work was funded in part by National Science Foundation Grant number IIS-1116384

Figure 9: Reverbalize: a novel reverberation controller implemented using the data collected via SocialReverb.

applied to it will be presented. Then the participant will be given the original, un-reverberated audio and asked to match the reverberation effect using either the new interface or a traditional parametric reverberation interface. We will measure two parameters: how quickly does the participant complete the task and how closely does the participant-applied reverberation match the original reverberation. Participants will also be asked to complete a survey about their satisfaction with the interface along various dimensions (e.g. ease of use, clarity of affordances, etc.).

In the second task, users will be given a production goal (e.g. "make the music sound as if it is being played underwater.") and asked to achieve that goal using either the traditional or the new interface. We will measure how quickly the user reports having achieved the goal, as well as the users' level of satisfaction with how well the goal was achieved. Further, the outcome will be presented to a second set of users who will be asked to measure how well the stated goal was achieved. As with the first task, participants will be asked to complete a survey about their satisfaction with the interface along various dimensions (e.g. ease of use, clarity of affordances, etc.).

Other future work includes a tool to automatically apply appropriate descriptive labels to the output of any existing reverberator. This can be done by measuring the impulse response, placing it on the descriptor map, and using the words nearby as a description of the reverberation effect. This will allow the creation of two-way language-based control of reverberation tools. The user can ask the question "What is the effect of changing knob A?" and the tool could predict "It will make the sound 'boomy'." Similarly, one could ask "How do I make the sound 'boomy'?" and be told by the tool. This opens up the possibility of a new kind of interactive instruction for such tools, informed by a real understanding of the mappings between words and audio effects.

7. CONCLUSIONS

In this work, we described an approach to learning actionable words to describe reverberation that could be used to produce a crowd-sourced folksonomy of descriptive terms for reveberation. This was embodied in a web-based data collection tool, SocialReverb.

We also developed methods to determine which descriptors are commonly used to describe reverberation, which are specific to a range of reverberation settings, and which are synonymous. We also created a map of words visualizing their relationship to each other and examined the map in both measure space (abstract control of the reverberator) and parameter space (direct control of reverberator). The data underlying this map has been made available to the broader community.

This map of descriptors, allows the creation of reverberation tools whose controls are conceptualized in terms of perceptually relevant terms ("make the drums sound chaotic") rather than in terms of the underlying processes used to create the reverberation (reverberation time, echo density, and so on). The data we gathered also tells us which words are actionable by a reverberator, like "echo", "distant", or "underwater". By collecting this vocabulary, we can create audio production tools that are accessible to novices, as well as understand the way people interact with reverberation tools and audio effects in general.

8. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation, Grant No. IIS-1116384. Additional thanks to Zach Puller for providing source audio for the study.

References

[1] Fons Adriaensen. "Acoustical Impulse Response Measurement with ALIKI". In: *4th International Linux Audio Conference*. 2006.

[2] Amazon. *Amazon Mechanical Turk*. 2005-2014. URL: http://mturk.com.

[3] Jeff Atwood. *Concluding the Great MP3 Bitrate Experiment*. 2012. URL: http://blog.codinghorror.com/concluding-the-great-mp3-bitrate-experiment/.

[4] Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer, 2005.

[5] Jon Burton. *Ear Machine iQ: Intelligent Equaliser Plugin*. June 2011. URL: http://www.soundonsound.com/sos/jun11/articles/em-iq.htm.

[6] Mark Cartwright and Bryan Pardo. "Social-eq: Crowdsourcing an equalization descriptor map". In: *14th International Society for Music Information Retrieval*. 2013.

[7] J. Grey. *Multidimensional perceptual scaling of musical timbres*. The Journal of the ASA, 61(5):1270-1277, 1977.

[8] H. Helmholtz and A. Ellis. *On the sensations of tone as a physiological basis for the theory of music*. Dover, New York, 2nd english edition, 1954.

[9] D. Huber and R. Runstein. *Modern recording techniques*. Focal Press/Elsevier, Amsterdam ; Boston, 7th edition, 2010.

[10] LAME. *LAME MP3 Encoder*. 1998-2014. URL: http://lame.sourceforge.net/.

[11] S. McAdams et al. *Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes*. Psychological Research, 58(3):177-192, 1995.

[12] George A. Miller. *WordNet: a lexical database for English*. 1995. DOI: 10.1145/219717.219748.

[13] James A. Moorer. "About this Reverberation Business". In: *Computer Music Journal* (1979).

[14] Zafar Rafii and Bryan Pardo. "A Digital Reverberator Controlled through Measures of the Reverberation". In: *Northwestern Electrical Engineering and Computer Science Department* (2009).

[15] Zafar Rafii and Bryan Pardo. "Learning to Control a Reverberator using Subjective Perceptual Descriptors". In: *10th International Society on Music Information Retrieval* (2009).

[16] AT Sabin, Zafar Rafii, and Bryan Pardo. "Weighted-Function-Based Rapid Mapping of Descriptors to Audio Processing Parameters". In: *Journal of the Audio Engineering Society* (2011), pp. 419–430.

[17] M. Sarkar, B. Vercoe, and Y. Yang. "Words that describe timbre: a study of auditory perception through lan- guage". In: *Proc. of Language and Music as Cognitive Systems Conference*. 2007.

[18] Prem Seetharaman and Bryan Pardo. "Reverbalize: a crowdsourced reverberation controller". In: *ACM Multimedia, Technical Demo* (2014).

[19] R. Shepard. *Geometrical approximations to the structure of musical pitch*. Psychological Review, 89(4):305-333, 1982.

[20] D. Smalley. *Spectromorphology: explaining sound-shapes*. Organised Sound, 2(02):107-126, 1997.

[21] L. Solomon. *Search for physical correlates to psychological dimensions of sounds*. The Journal of the ASA, 31(4):492-497, 1959.

[22] S Sundaram and S Narayanan. "Analysis of audio clustering using word descriptions". In: *ICASSP: Acoustics, Speech and Signal Processing* (2007).

[23] A Zacharakis, K Pastiadis, and G Papadelis. "An Investigation Of Musical Timbre: Uncovering Salient Semantic Descriptors And Perceptual Dimensions". In: *12th International Society for Music Information Retrieval Conference*. 2011.