Discriminating Sound Textures

Jayson Lynch* Eric Mannes†

December 9, 2015

Abstract

We show sound textures are cool stuff.

Keywords: Machine Learning, Sound Textures, Audition

*MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, MA 02139, USA, {jaysonl}@mit.edu.

†MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, MA 02139, USA, {mannes}@mit.edu.

1 Introduction

Sound textures, such as rain or crackling fire, are the result of many similar acoustic events.

Recent work by McDermott and Simoncelli attempts to understand a special category of sounds called *sound textures* [MS11, MSS13]. They show that for many sound textures, the salient features that cause humans to classify these sounds are controlled by a number of time-averaged statistics of the sounds. This leads to an interesting question of why some sounds seem to be identified by time-averaged properties, when many other, such as speech or music, depend strongly on their temporal structure. McDermott et. al. conjecture that the 'sparsity' or 'compressibility' are related to the sounds that are perceived in this way. In addition, they mention that sound textures are 'temporally homogeneous' without giving formal definitions of these terms. This paper seeks to build upon these ideas both to investigate new features for audio classification as well as to lend evidence toward the previous conjectures about sound texture perception. We provide several formal definitions of audio compressibility, sparsity, and temporal homogeneity in Section 2. These are then extracted as a feature set and are used in a machine learning algorithm for sound textures, described in Section 3.

There has been other prior work on classifying sound textures. Saint-Arnaud's Master's Thesis attempts to extract the sound atoms that make up sound textures and use this as a basis for a classifier [SA95]. The thesis also discusses human perception of sound textures and the difficulties surrounding a good definition.

[Cite McDermott's papers [MS11, MSS13] and why we care about these features/classification] xxx

[Cite some other audio/sound texture classification work.] xxx

2 Definitions

We provide working definitions for examples of three main descriptors: sparsity, compressibility, and temporal homogeneity. Since these characteristics can be captured in different ways, which are not always equivalent, we choose to work with multiple formal definitions.

Compressibility was the simplest to work out. We chose a standard lossy and lossless compression algorithm and measured the compression ratio for the .wav flies. In this case we used **[what audio compression did we use?]** xxx

[Give formal and informal definitions of the features we used] xxx

3 Methods

[Describe how we set up our feature extraction. How we set up our learning. What our dataset is. Any other important process things.] xxx

4 Results

[What coorilated and what didn't? What feature (ensemble) lead to a good classifier?] xxx

5 Conclusion

[Speculate if these could be useful features in audio classification. Speculate about the implication to human audition. Give future directions.] xxx

Acknowledgments

We thank Prof Josh McDermott for his support in answering our questions about their research and providing us with their dataset. We also thank the course staff of 6.867 for their instruction and support this semester. In particular, we appreciate Marzyeh Ghassemi advice and guidance in shaping our project plan.

References

[MS11] Josh H. McDermott and Eero P. Simoncelli. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, 71(5):926 – 940, 2011.

[MSS13] Josh H McDermott, Michael Schemitsch, and Eero P Simoncelli. Summary statistics in auditory perception. *Nature neuroscience*, 16(4):493–498, 2013.

[SA95] Nicolas Saint-Arnaud. *Classification of sound textures*. PhD thesis, Massachusetts Institute of Technology, 1995.