Predicting reflection patterns from binaural activity maps using deep neural networks

By: Jeramey Tyler
Advised by: Mei Si (Cognitive Science), Jonas Braasch (Architecture)

Abstract

A new model architecture is presented to predict room acoustical parameters from a running binaural signal. For this purpose, a deep neural network architecture is combined with a precedence effect model to extract the spatial and temporal locations of the direct signal and early reflections. The precedence effect model builds on the modified BICAM algorithm (Braasch, 2016), for which the 1st layer auto-/cross correlation functions are replaced with a Cepstrum method. The latter allows a better separation of features relating to the source signal's early reflections and harmonic structure. The precedence effect model is used to create binaural activity maps that are analyzed by the neural network for pattern recognition. Anechoic orchestral recordings were reverberated by adding four early reflections and late reverberation to test the model. Head-related transfer functions were used to spatialize direct sound and early reflections. The model can identify the main reflection characteristics of a room, offering applications in numerous fields, including room acoustical assessment, acoustical analysis for virtual-reality applications, and modeling of human perception.

Roadmap

You are here

1. The problem
2. The current approach
3. The new approach
4. Experiment design
5. Methodology
6. Results
7. Discussion and Questions
8. References

The problem

Sound source localization

Sound source localization is the ability to detect the spatial location of a sound's origin.

The precedence effect

The psychoacoustical effect in which two sounds separated by a sufficiently short delay are perceived as occurring at the same time and spatial location is called the *precedence effect (Blauert, 97)*.

The current approach

Interaural time difference

The *interaural time difference* (ITD) is a contributing factor to a person's ability to localize sounds and represents the difference in time between when a sound arrives at the first ear and when it arrives at the second ear.

**Honorable mentions

Other factors that contribute to sound source localization include the *interaural phase difference* (IPD) , *interaural intensity difference* (IID), and *interaural level difference* (ILD) which represent the difference respectively in phase, intensity, and level between the first and second ears.

The current approach

Autocorrelation

Autocorrelation is the correlation of a signal with a time-delayed copy of itself and it is thought to play an important part in human hearing. In this instance the signals from each ear are correlated with each other to extract . This reveals harmonic properties and allows for detection of early reflections.

2nd-layer autocorrelation

A second layer autocorrelation is used to determine the ITD and time align the left and right ear autocorrelated signals. The autocorrelated and aligned signals are then used to generate binaural activity maps.

The new approach

Cepstral analysis

Whereas BICAM analyzes signals with respect to the time domain CEPBIMO analyzes signals with respect to the cepstral domain. The *cepstral domain* can be most simply described as the spectrum of a spectrum, hence *spec*-tral becomes *ceps*-tral and *fre-que*-ncy becomes *que-fre*-ncy.

The cepstrum is defined as the inverse fourier transform of the logarithm of the spectrum, or more formally:

$$\mathscr{F}^{-1}\{\log(|\mathscr{F}\{f(t)\}|)\}$$

Experiment design

Neural networks

The efficacy of CEPBIMO in comparison to BICAM is determined by analysing their predicted accuracy in traditional deep neural networks and convolutional neural networks.

Data representations

The type of neural network being trained determines the way in which CEPBIMO and BICAM data are represented. Both BICAM and CEPBIMO produce images as output which is used to train the convolutional neural networks. Additionally deep neural networks are trained on the raw tensor representations of BICAM and CEPBIMO.

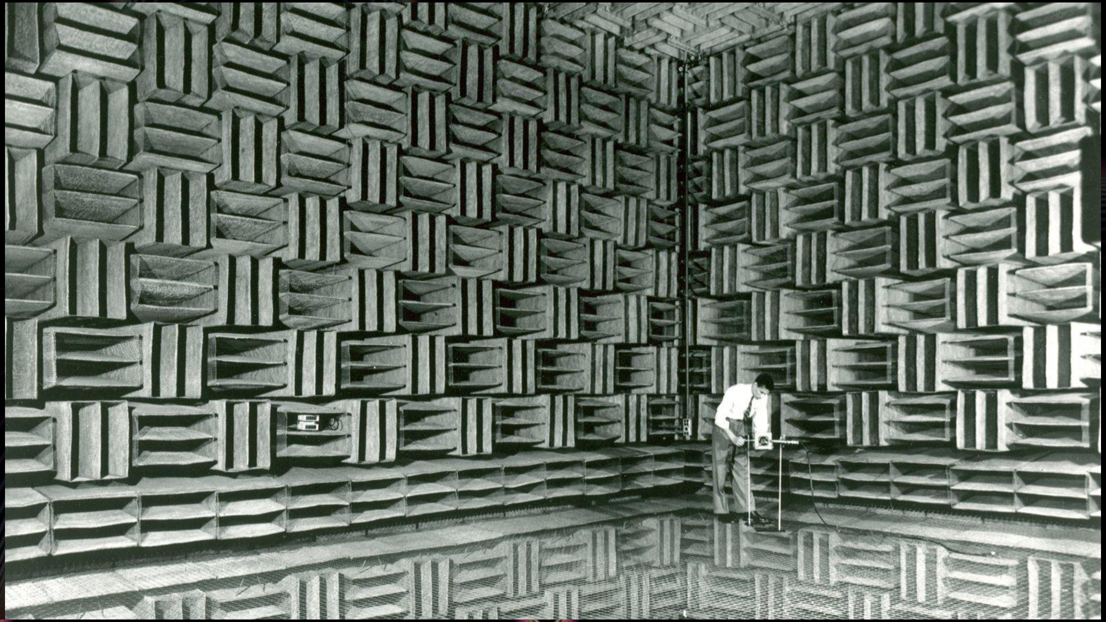
Methodology

Anechoic recordings

To create a dataset of simulated reflections we need to start with recordings produced in an *anechoic* environment, or an environment that does not produce echos. A collection of orchestral pieces recorded in an anechoic chamber is available courtesy of (Pätynen, et. al, 08) at https://https://research.cs.aalto.fi//acoustics/virtual-acoustics/research/acoustic-measurement-and-analysis/85-anechoic-recordings.html

Methodology

Anechoic Chamber



Methodology

Head related transfer functions

Head related transfer functions (HRTF) are used to change the perceived spatial location of the simulated reflections. An HRTF describes the rotation of the head in azimuth degrees on the horizontal axis and zenith degrees on the vertical axis. A set of HRTFs is made available courtesy of (Martin & Gardener, 94) and MIT media lab at https://sound.media.mit.edu/resources/KEMAR.html.

Methodology

1. The problem
2. The current approach
3. The new approach
4. Experiment design
5. Methodology
   a. Dataset
   b. Neural networks
   c. Transformations
6. Results
7. Discussion and questions
8. References

Kemar dummy head

Methodology

Dataset features

- 10,000 samples
- 10s each
- Randomized direct signal spatial origins
- 4-8 randomized reflections per sample
- Randomized reverberation
- Train, test, validation split
- Hosted online
    - https://reflections.speakeasy.services/train.zip
    - https://reflections.speakeasy.services/test.zip
    - https://reflections.speakeasy.services/validate.zip
    - https://reflections.speakeasy.services/full.zip

Methodology

1. The problem
2. The current approach
3. The new approach
4. Experiment design
5. Methodology
    a. Dataset
    b. Neural networks
    c. Transformationa
6. Results
7. Discussion and questions
8. References

Neural networks

Deep neural network

Convolutional neural network

Methodology

1. The problem
2. The current approach
3. The new approach
4. Experiment design
5. Methodology
    a. Dataset
    b. Neural networks
    c. Transformations
6. Results
7. Discussion and questions
8. References

Transformations

- 1st layer autocorrelation
- 2nd layer autocorrelation

Results

Results

Discussion and questions

Discussion and questions

References

1. Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.
2. Braasch, J. (2016). Binaurally integrated cross-correlation/auto-correlation mechanism (BICAM). The Journal of the Acoustical Society of America, 139(4), 2211-2211.
3. Kaernbach, C., & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. The Journal of the Acoustical Society of America, 104(4), 2298-2306.
4. Martin, K., & Gardener, B. (1994). HRTF Measurements of a KEMAR Dummy-Head Microphone. Media Lab Perceptual Computing.
5. Pätynen, J., Pulkki, V., & Lokki, T. (2008). Anechoic recording system for symphony orchestra. Acta Acustica united with Acustica, 94(6), 856-865.