StIM: Structural image method for augmentation of deep learning audio tasks.
M.Sc. seminar presentation by Erez Shalev. Supervised by Prof. Israel Cohen
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
Audio classification:
What is audio classification?
Audio classification

What is it good for?
Audio classification:
Motivation

- Smart cities.
- Smart homes.
- Surveillance systems.
- Alarm systems.
- Robots.
Audio classification:
Common audio classification problems

- Problem definition and hierarchical labeling.
- Scarcity of datasets.
- Polyphonic sounds.
- Environmental factors.
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StiM
- Speech emotion recognition
- Future work
Simulation methods for reverberations

- Wave-based: computationally complex.
- Statistical methods: Non-temporal.
- Ray-based.
Ray tracing

- Reverberated Sounds
- Echo
- Dry Sound
Simulation methods for reverberations:
Image method

\[ L_x \]

\[ L_y \]
Simulation methods for reverberations:
Image method
Simulation methods for reverberations:

Image method
Simulation methods for reverberations:

Image method

\[ O \]

\[ L_x \]

\[ L_y \]
Simulation methods for reverberations:
Image method
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
Audio classification from another room
Audio classification from another room:

Motivation

- Smart homes – economical benefits.
- Robots – Not static.
- Surveillance.
- Speech emotion recognition
Simulation methods for reverberations:
Image method with a source outside the room
Simulation methods for reverberations:
Image method with a source outside the room
Simulation methods for reverberations:
Image method with a source outside the room
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
StIM:
Receiver image method
StIM:
Receiver image method
StIM:
Receiver image method
StIM:
Receiver image method
StIM:
Receiver image method
StIM:
Structural image method
Audio classification from another room: Integrating RIR into the dataset

Example of Audio in K rooms
Audio classification from another room:

Models

- Baseline CNN model.
- AlexNet classifier.
- VGG16 classifier.
Audio classification from another room: Integrating RIR into the dataset
### Audio classification from another room:

#### Performances

<table>
<thead>
<tr>
<th>Model</th>
<th>Clean</th>
<th>Clean</th>
<th>StILM</th>
<th>StILM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base model</td>
<td>100 %</td>
<td>56.66 %</td>
<td>98.8 %</td>
<td>93.33 %</td>
</tr>
<tr>
<td>AlexNet</td>
<td>96.66 %</td>
<td>40 %</td>
<td>99.81 %</td>
<td>83.33 %</td>
</tr>
<tr>
<td>VGG16</td>
<td>100 %</td>
<td>53.33 %</td>
<td>95.85 %</td>
<td>90 %</td>
</tr>
</tbody>
</table>
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
Speech emotion recognition from beyond the wall:

Motivation

- Covid-19 isolation policy:
  - A rise in domestic violence
  - Higher suicide rate
  - Emotional responses - children, elderly community,
  - People coping with mental conditions.
- Aggression detection.
- Distress detection in isolation.
Speech emotion recognition from beyond the wall:

Models

- AlexNet.
- LSTM-RNN.
- GRU-RNN.
Speech emotion recognition from beyond the wall:

Results
Speech emotion recognition from beyond the wall:

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Clean</th>
<th>Clean</th>
<th>StIM</th>
<th>StIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>84.14 %</td>
<td>20 %</td>
<td>89.13 %</td>
<td>50.66 %</td>
</tr>
<tr>
<td>RNN-LSTM</td>
<td>83.74 %</td>
<td>21.33 %</td>
<td>89.34 %</td>
<td>61.33 %</td>
</tr>
<tr>
<td>RNN-GRU</td>
<td>85.94 %</td>
<td>26.66 %</td>
<td>87.82 %</td>
<td>53.33 %</td>
</tr>
</tbody>
</table>
Outline:

- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work
Future work:

Motivation

- Deep network to produce real RIR.
  - VAE vs. GAN
  - Time domain vs. frequency domain.
- CRNN models.
- Distillation of models.
- Interlacing augmentation for aggression detection.
Contributions:

- StIM: structural imaged method.
- Integrating RIR augmentation into an audio dataset.
- Evaluation of audio classification from beyond a wall.
- Baseline for speech emotion recognition from beyond a wall.
Thanks