TECHNION
Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER ENGINEERING

SIPL
Signal and Image Processing Lab.

# 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


BRAIN GAMES
ANOTHER NEW
EPISODE
NEXT
EXPLORER
natgeotv.com

# Audio classification:

What is audio classification?

# Audio classification

Audio classification

~~Why~~

What is it good for?

# Audio classification:

## Motivation

- Smart cities.
- Smart homes.
- Surveillance c
- Alarm systems
- Robots.

# Audio classification:

Common audio classification problems

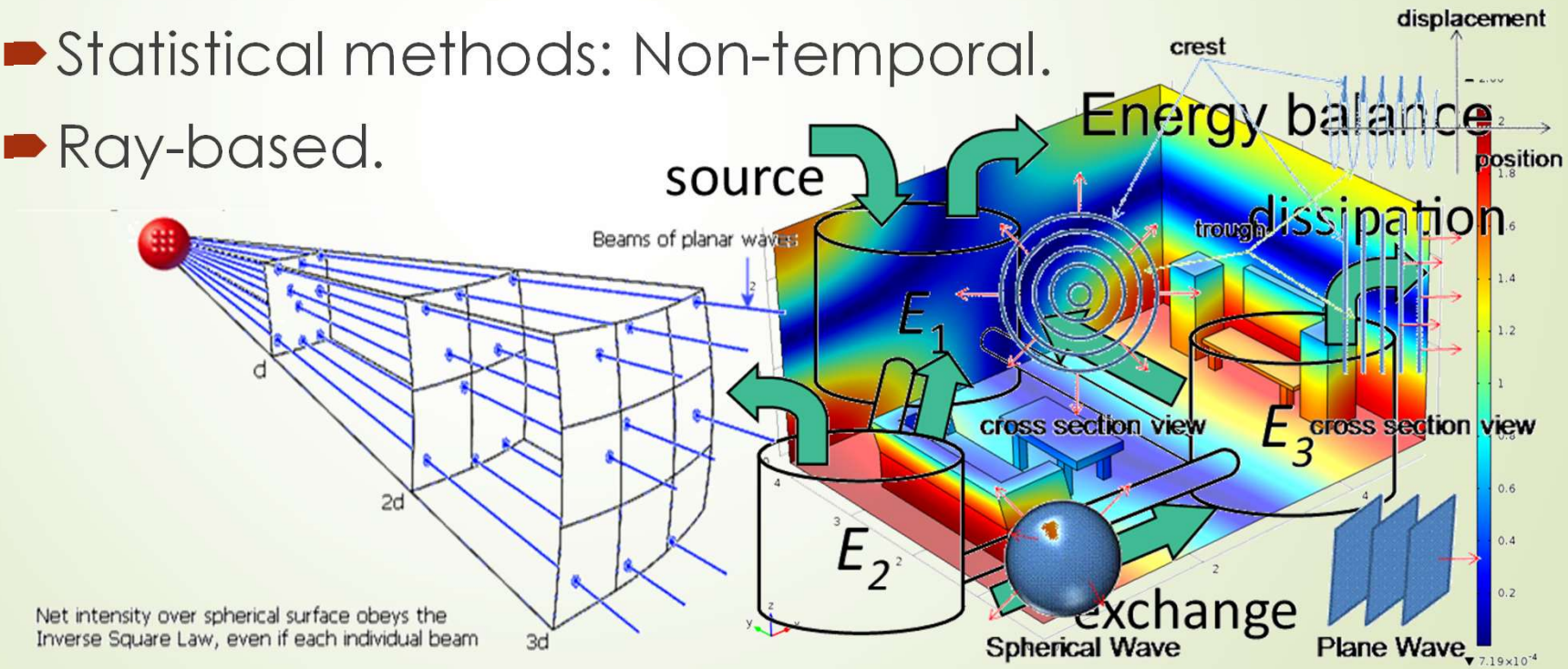
- Problem definition and hierarchical labeling.
- Scarcity of datasets.
- Polyphonic sounds.
- Environmental factors.

# Outline:

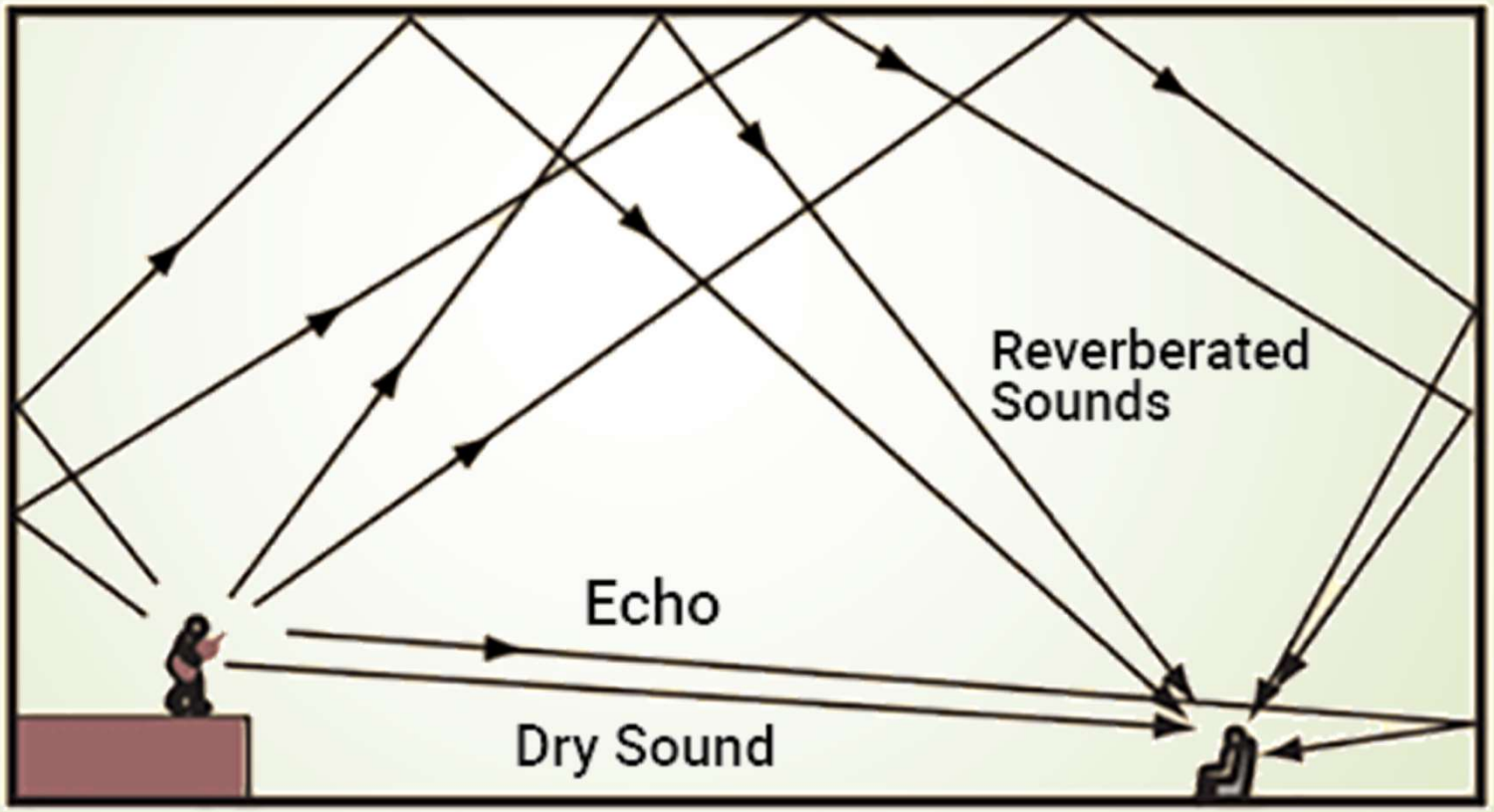
- Audio classification and common problems
- Simulation methods for reverberations and IM
- Audio classification from another room
- StIM
- Speech emotion recognition
- Future work

# Simulation methods for reverberations

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

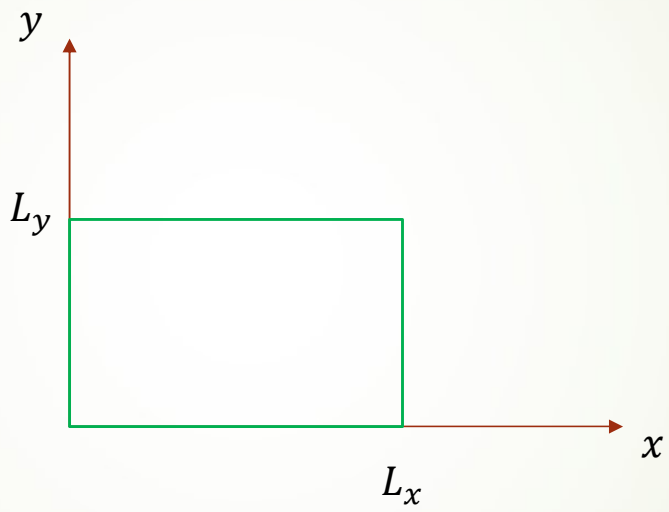


# Ray tracing



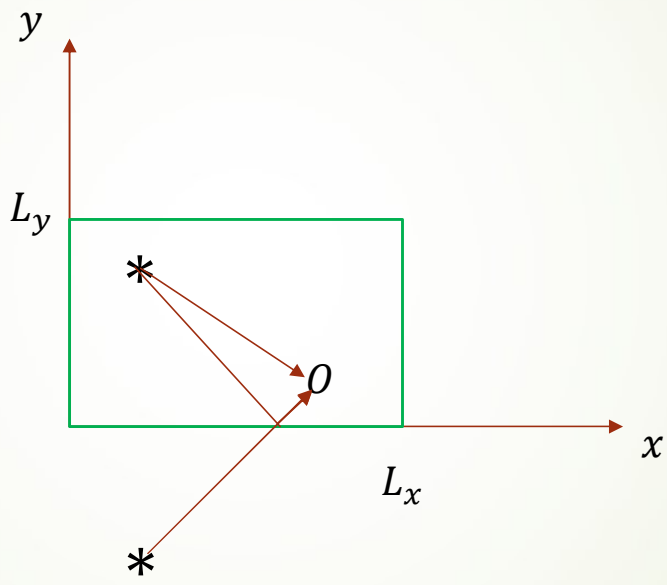
# Simulation methods for reverberations:

## Image method



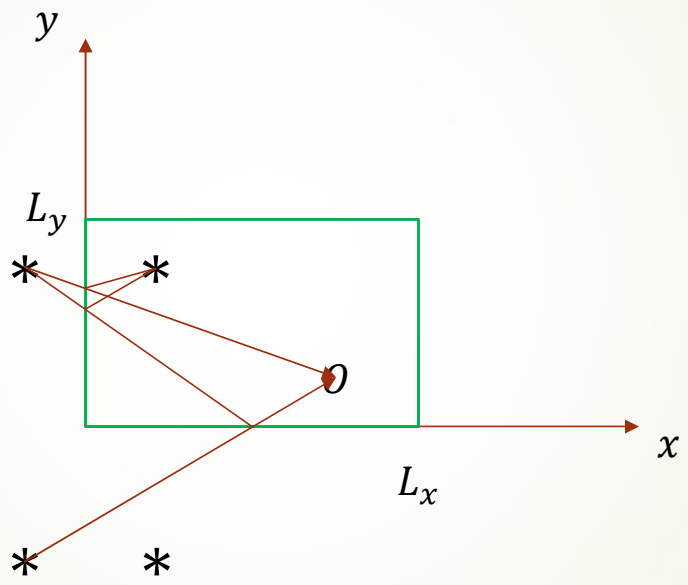
# Simulation methods for reverberations:

## Image method



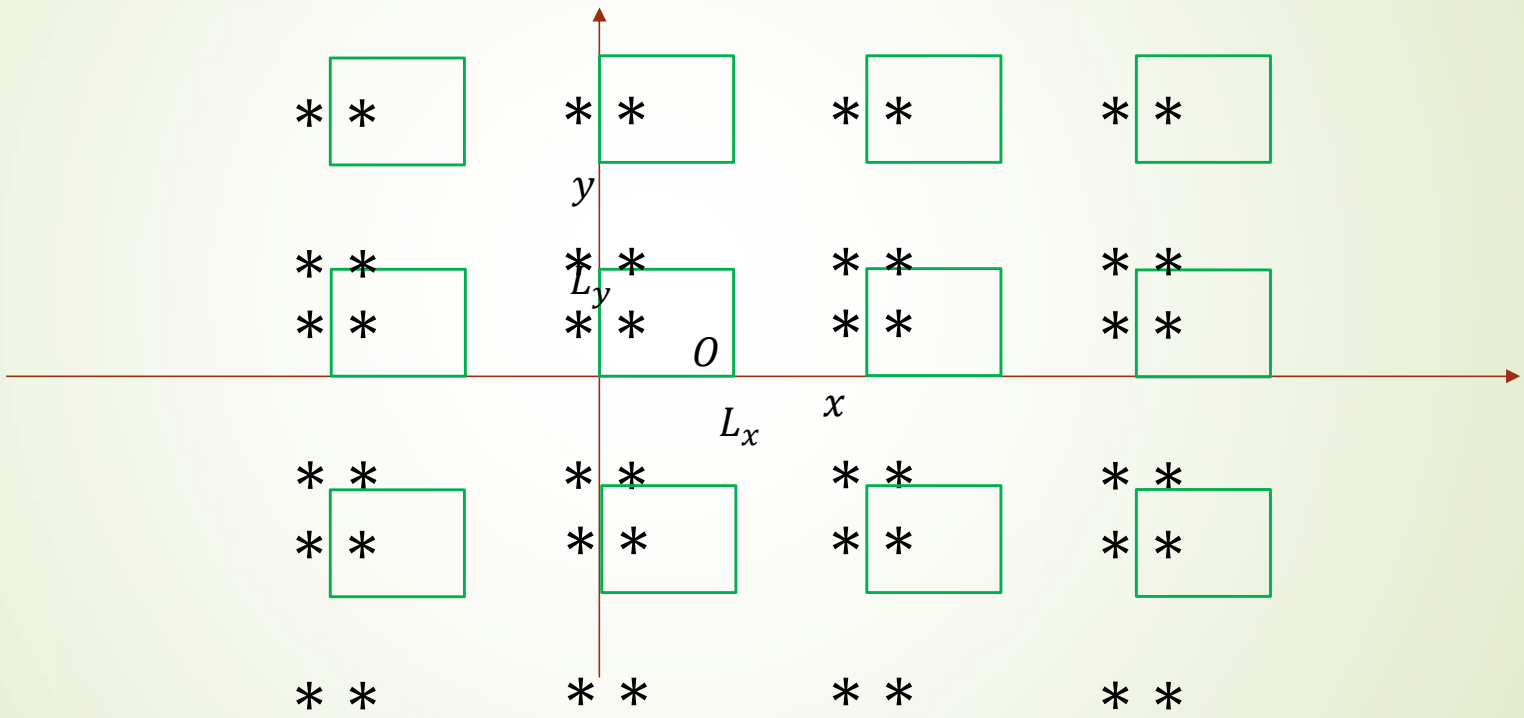
# Simulation methods for reverberations:

## Image method



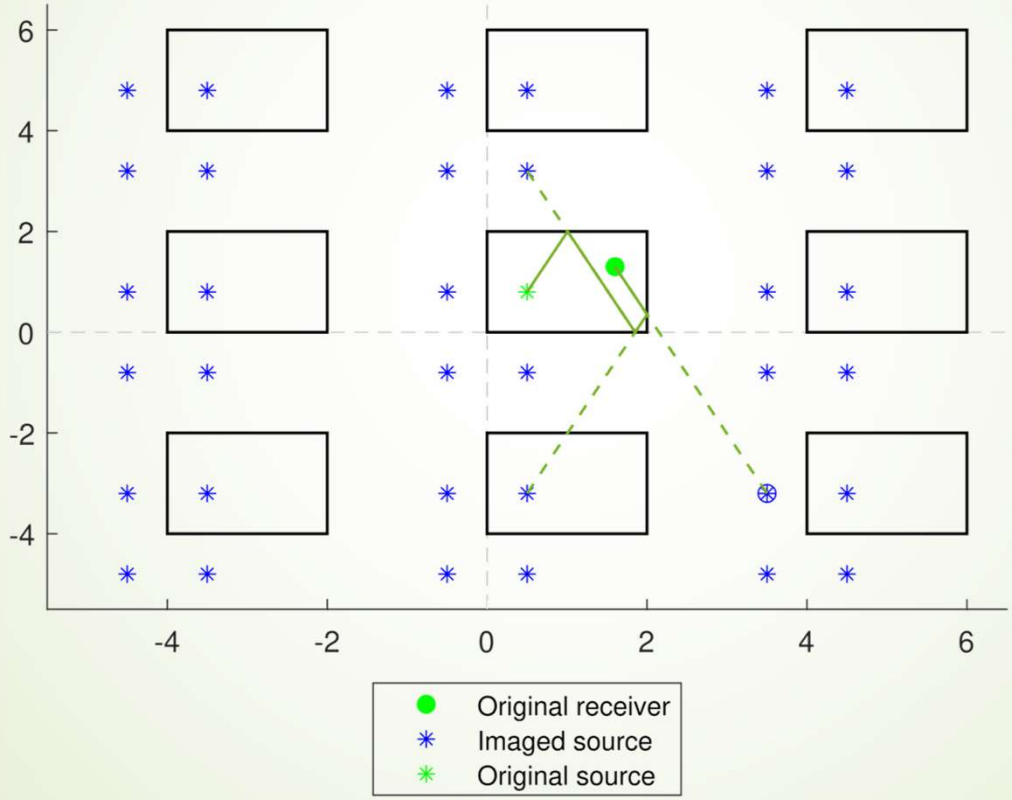
# Simulation methods for reverberations:

## Image method



# 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

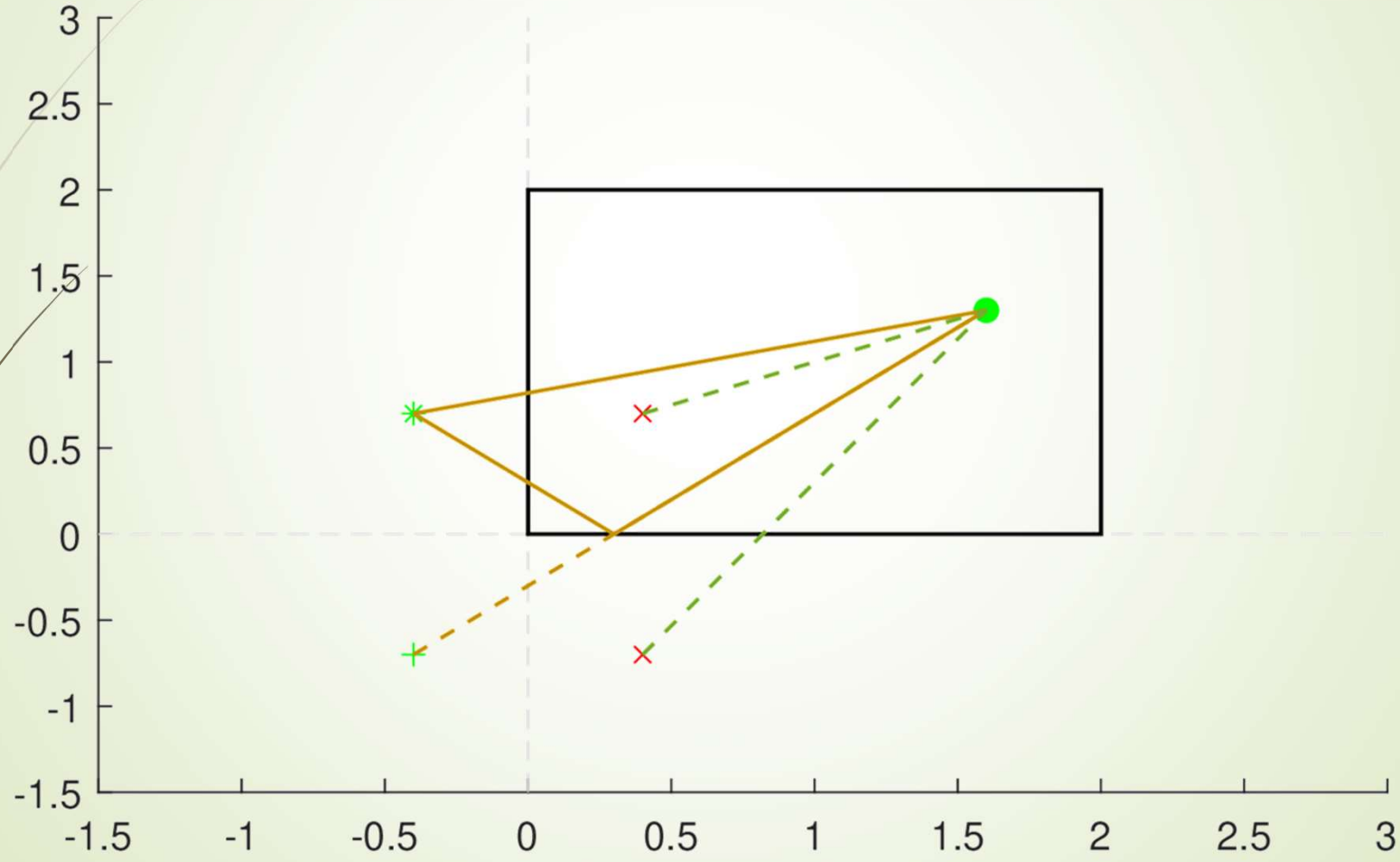
# Audio classification from another room

# Audio classification from another room:
## Motivation

- Smart homes – economical benefits.
- Robots – Not static.
- Surveillance.
- Speech emotion recognition

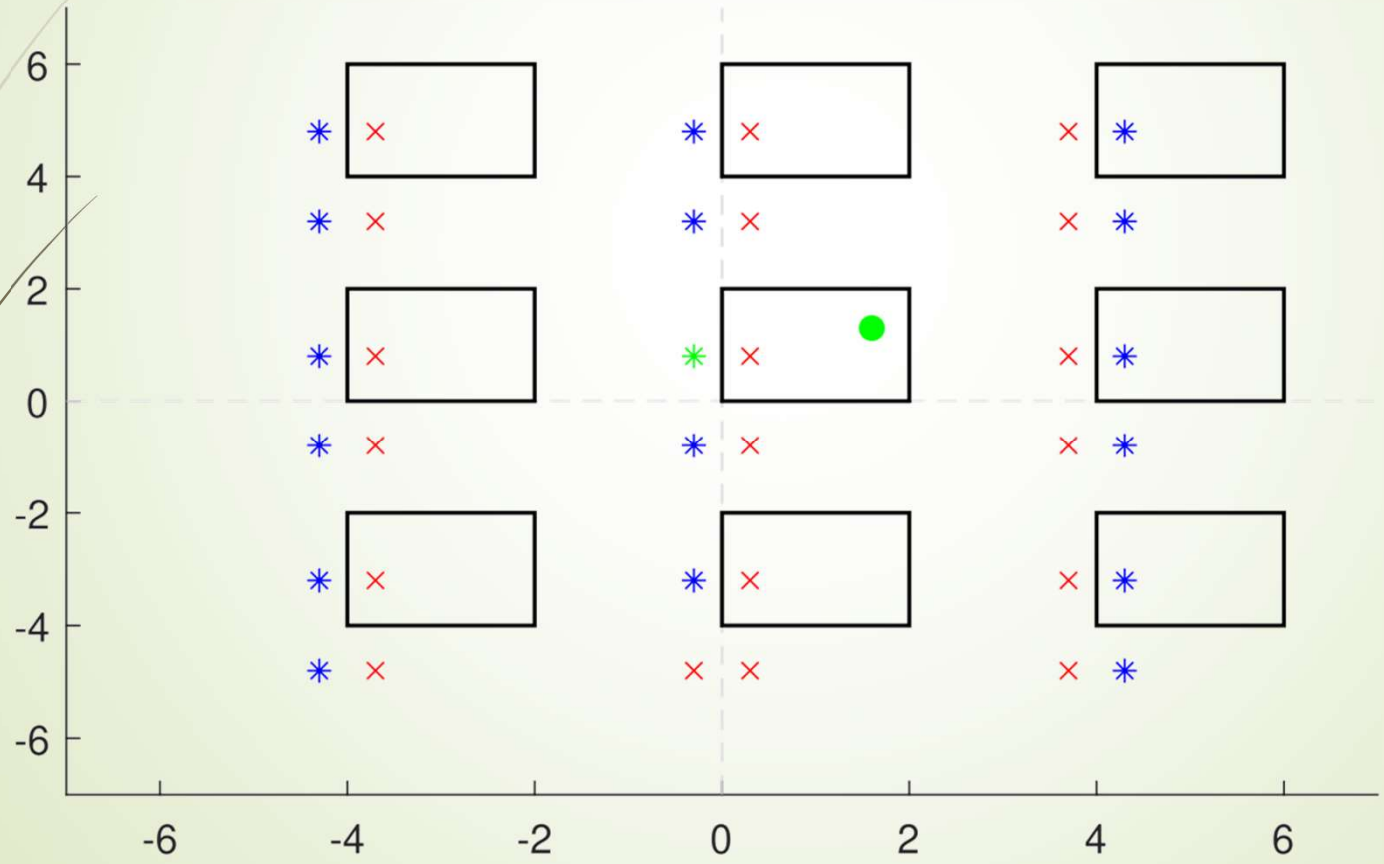
# Simulation methods for reverberations:

Image method with a source outside the room



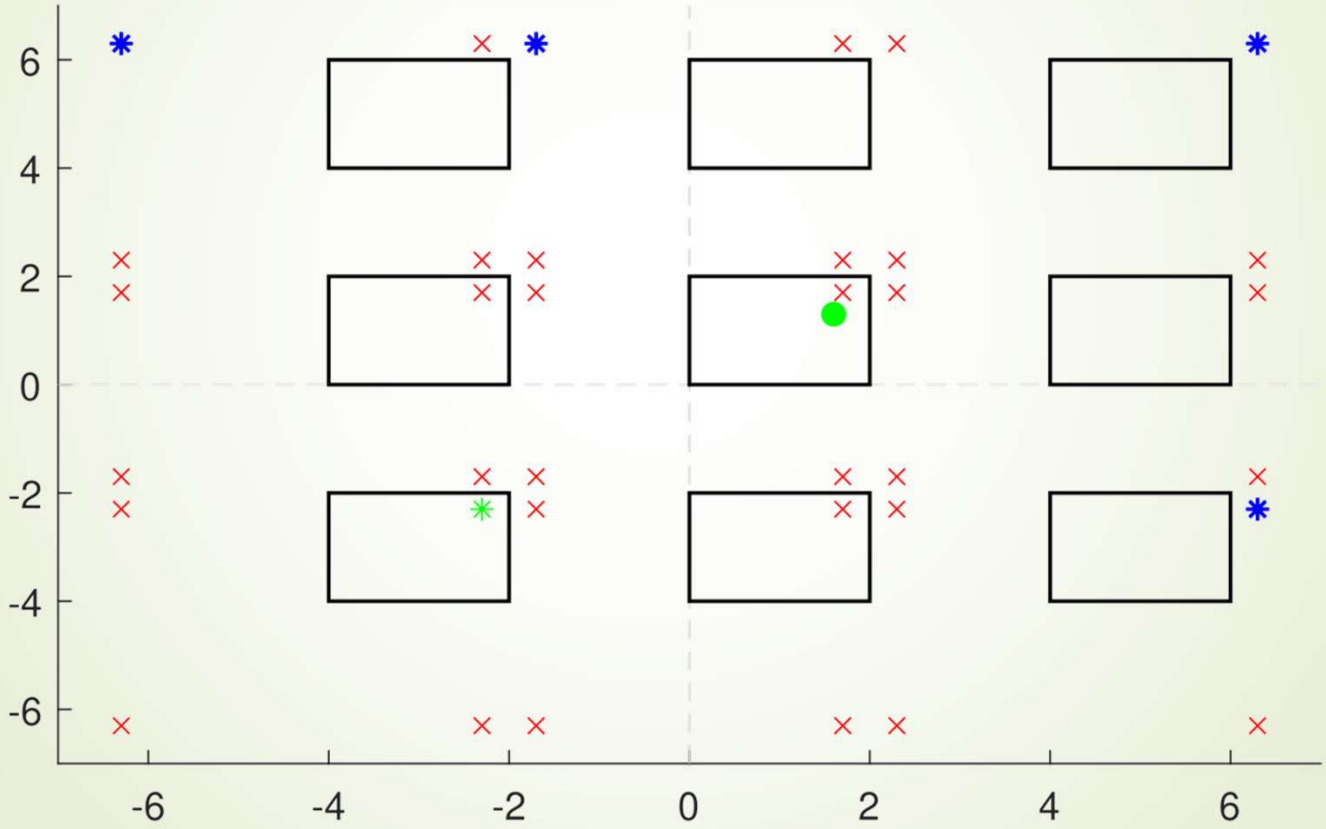
# Simulation methods for reverberations:

Image method with a source outside the room



# Simulation methods for reverberations:

Image method with a source outside the room

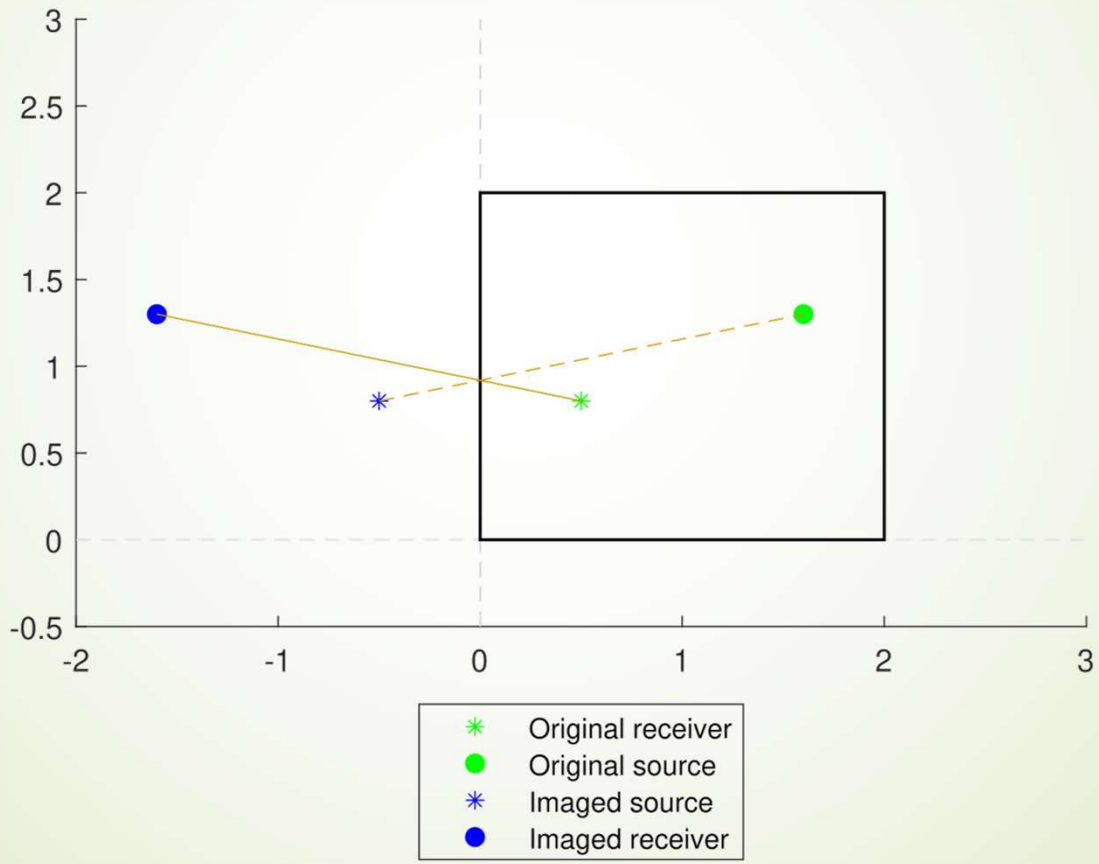


# Outline:

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

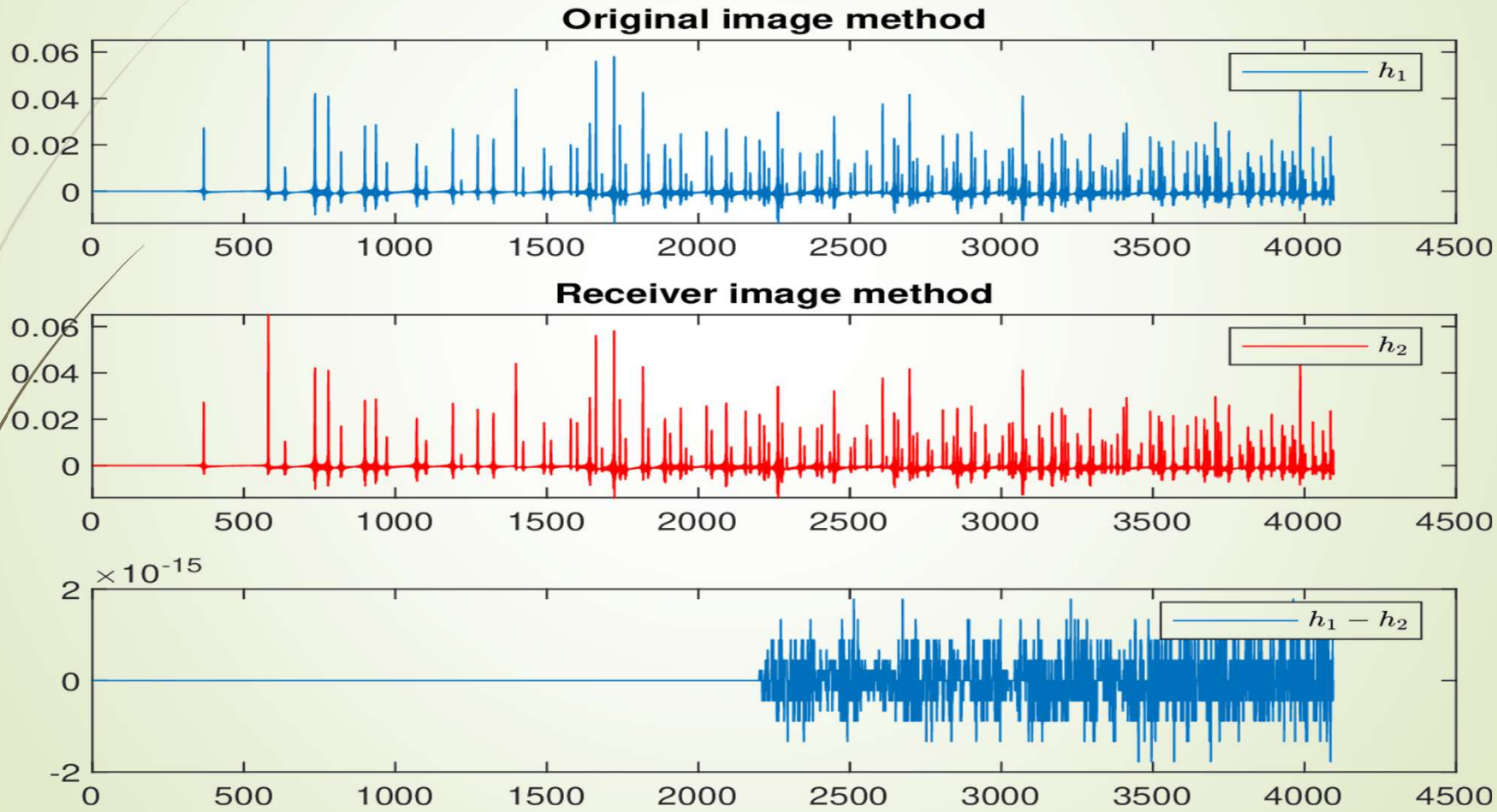
# StIM:

Receiver image method



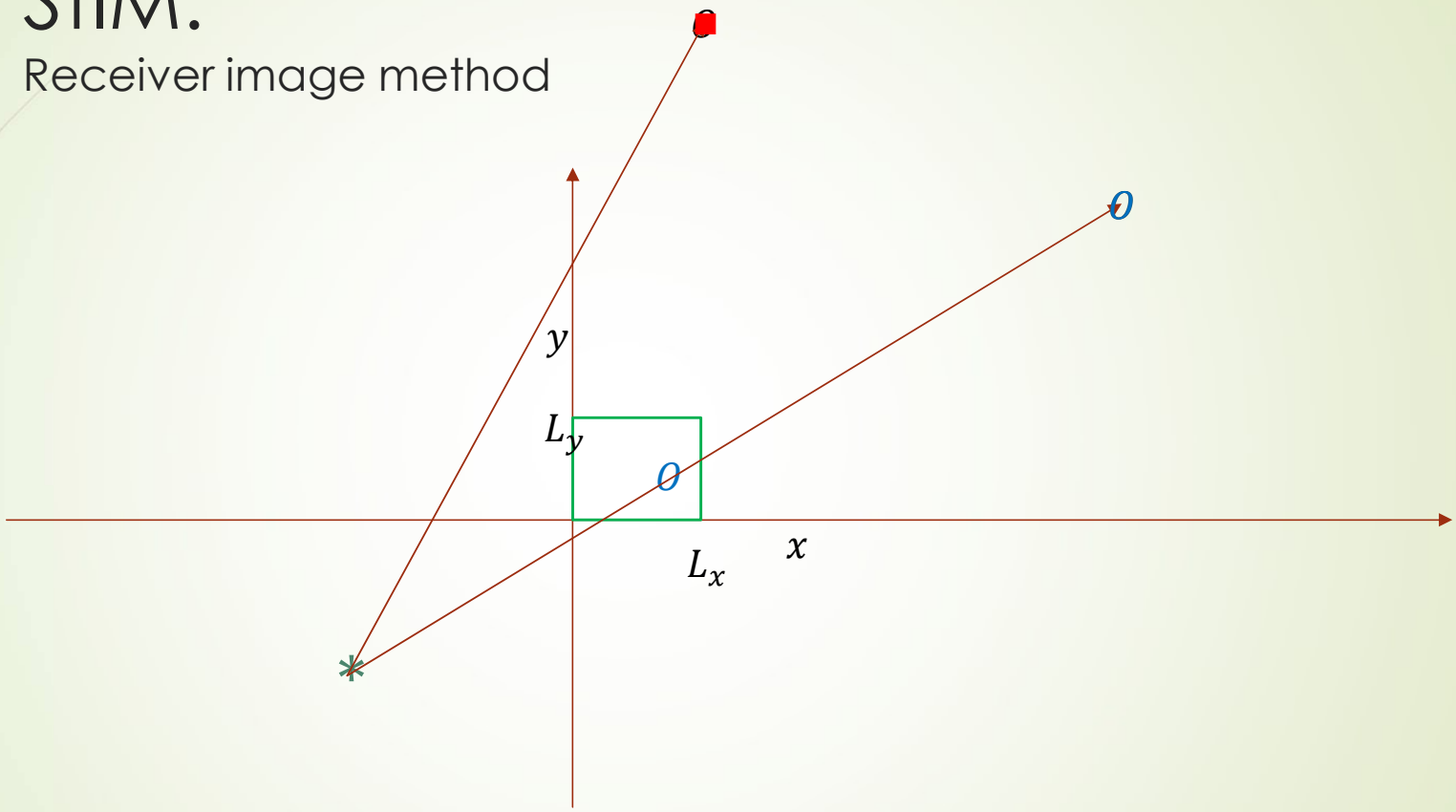
# StIM:

Receiver image method



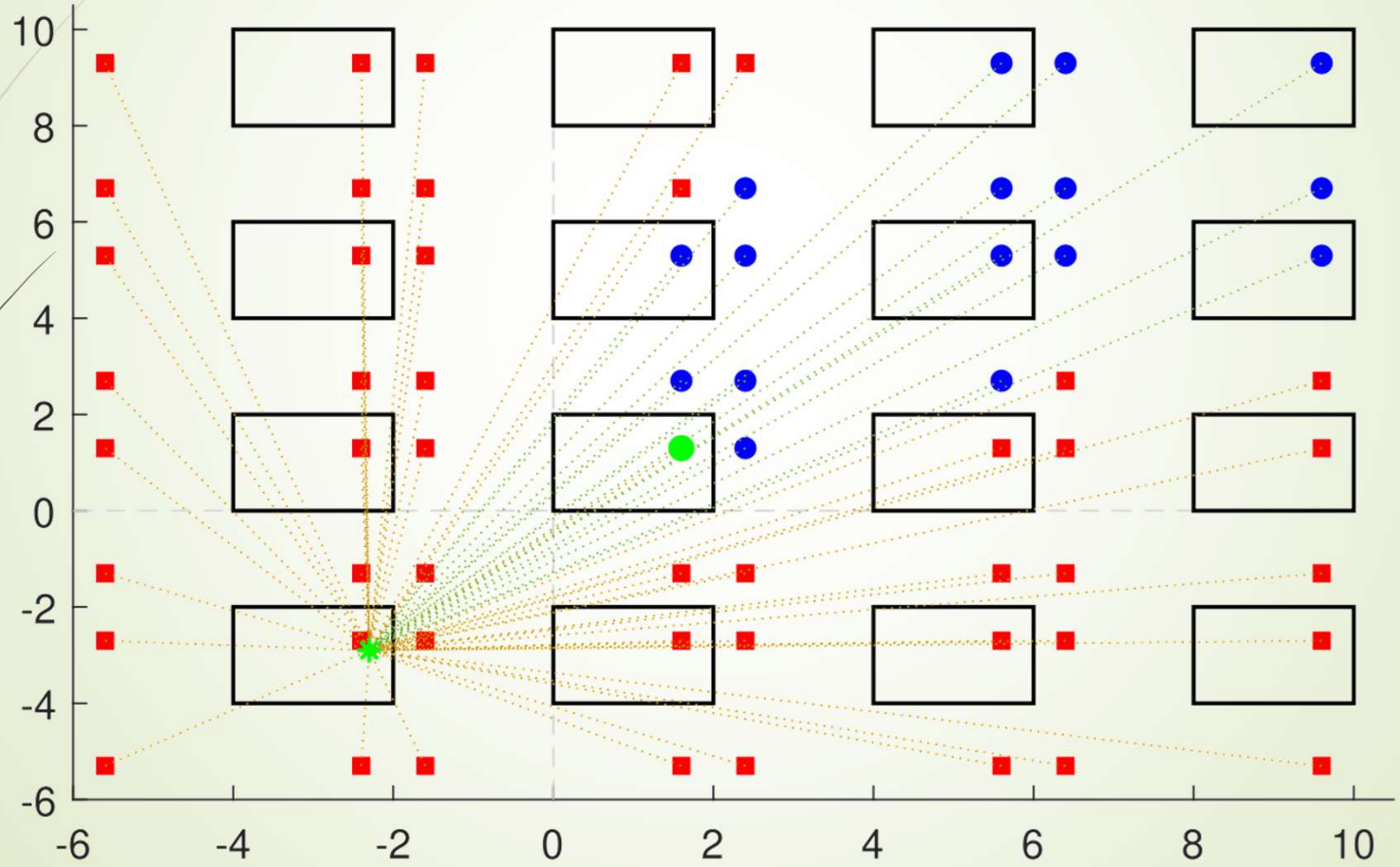
# StIM:

Receiver image method



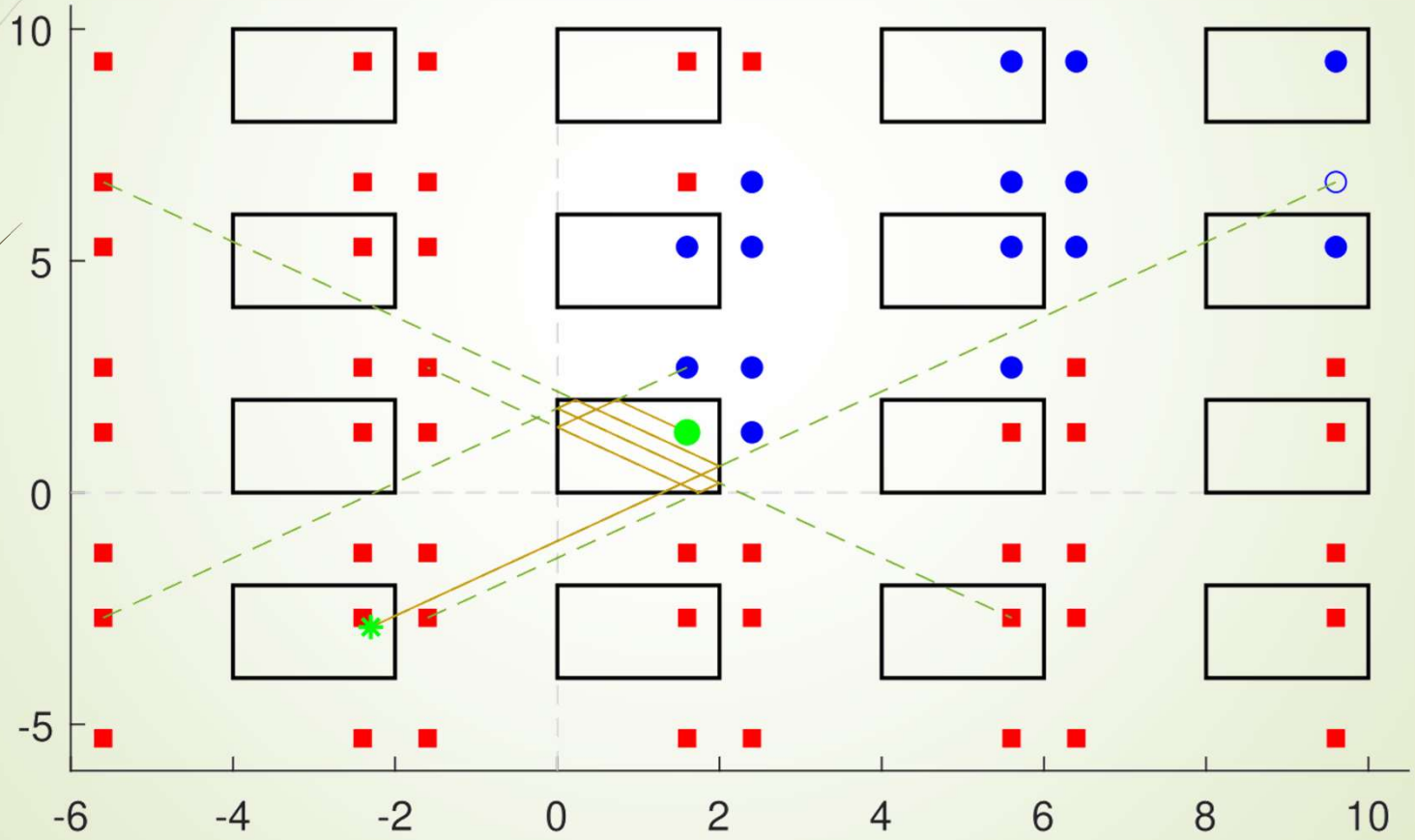
# StIM:

Receiver image method



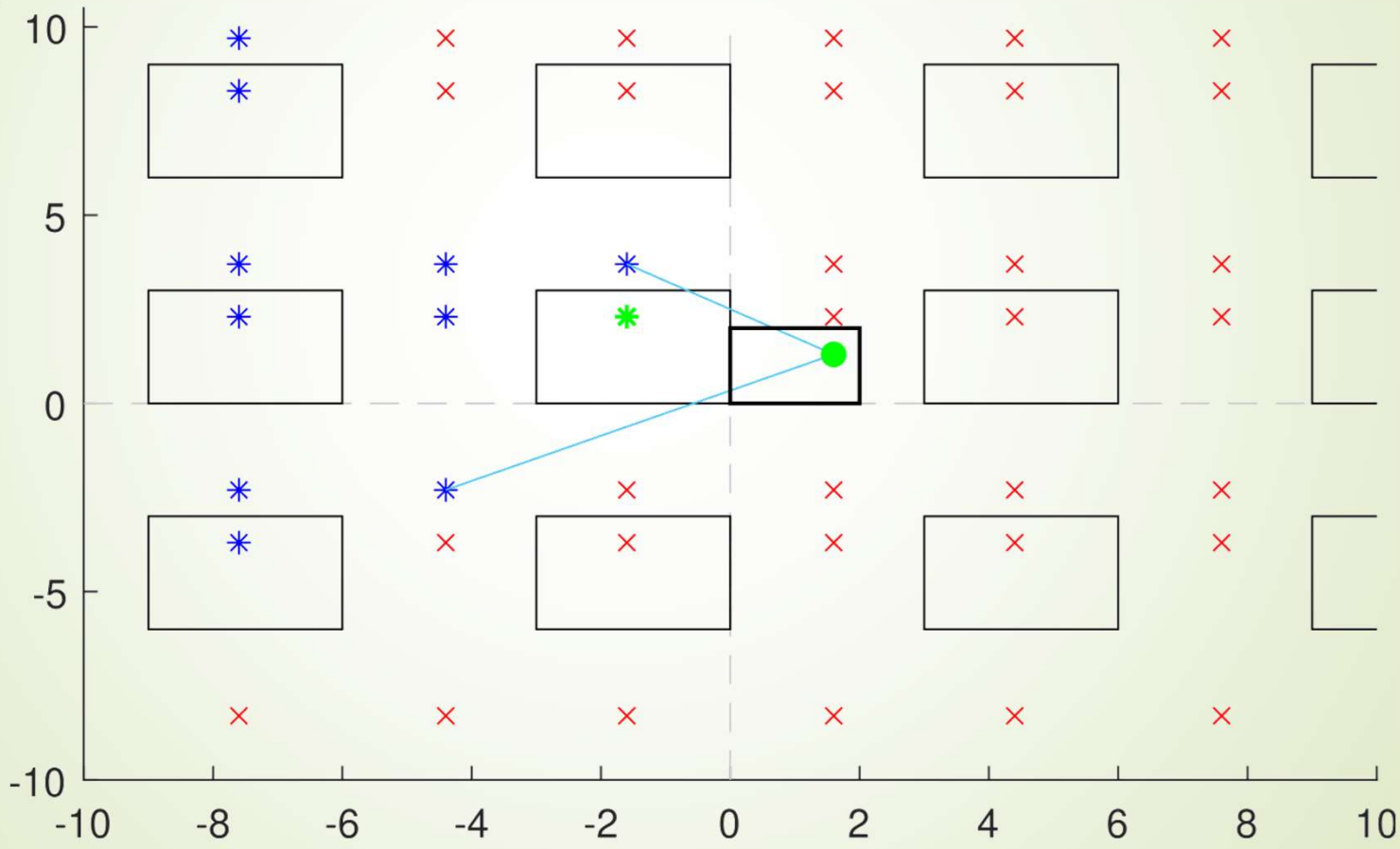
# StIM:

Receiver image method



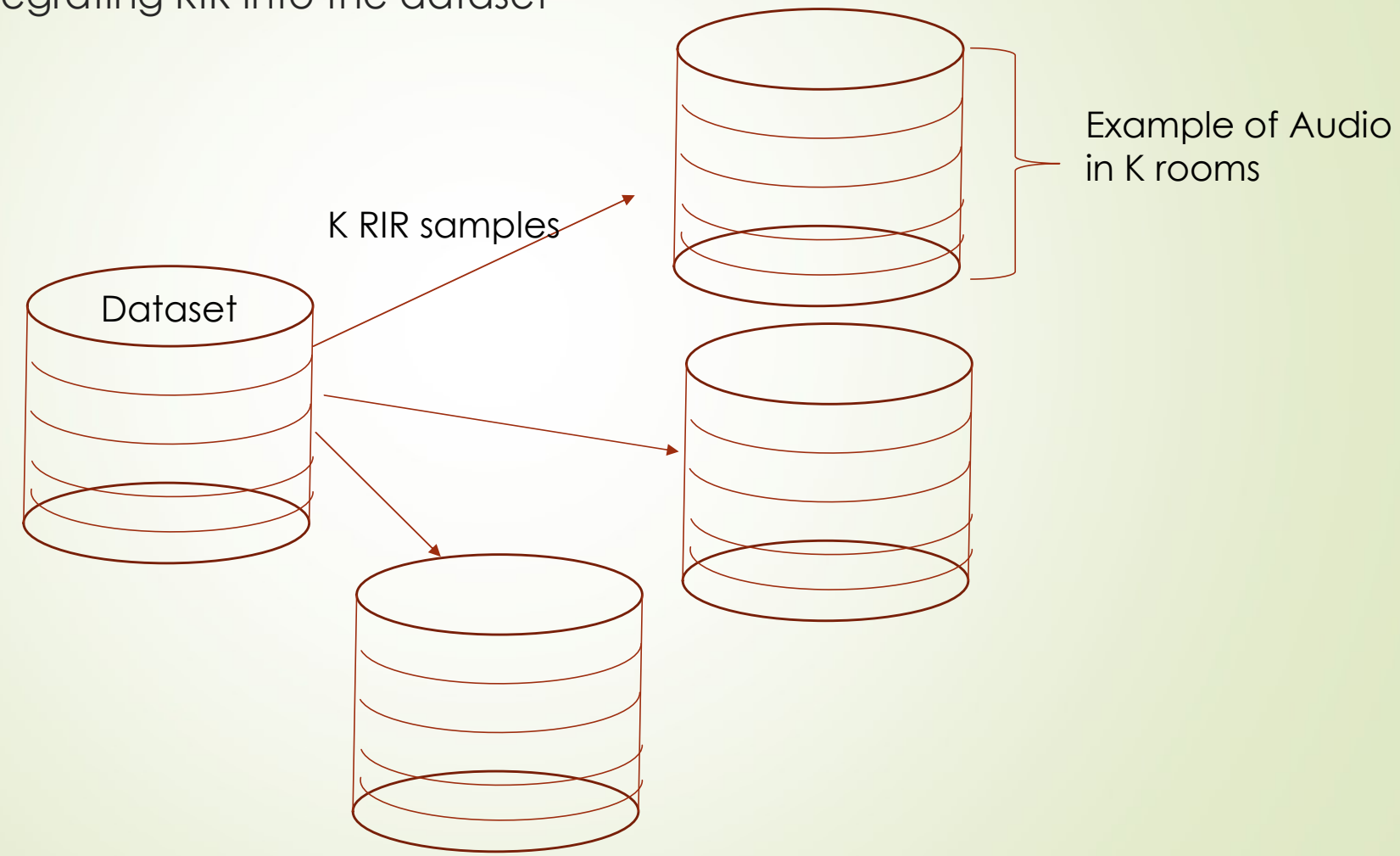
# StIM:

Structural image method



# Audio classification from another room:

## Integrating RIR into the dataset

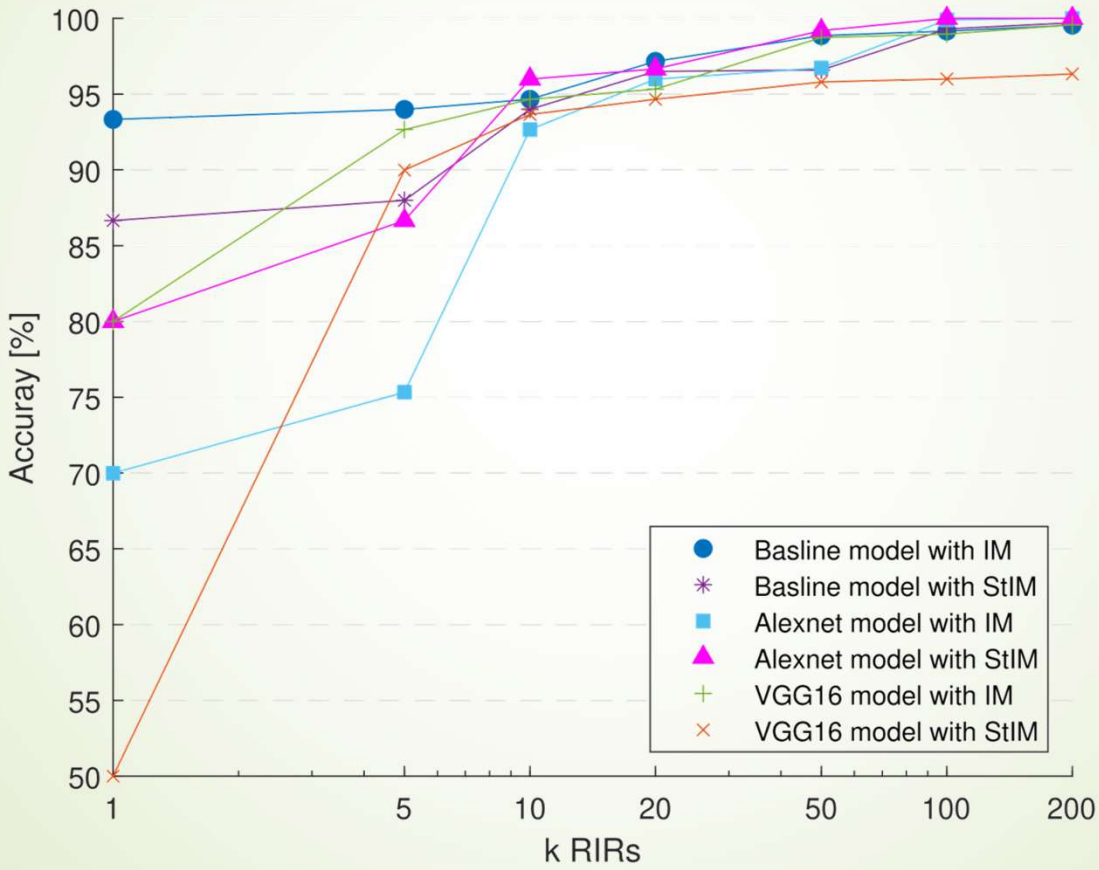


# Audio classification from another room:
## Models

- Baseline CNN model.
- AlexNet classifier.
- VGG16 classifier.

# Audio classification from another room:

## Integrating RIR into the dataset



# Audio classification from another room:
Performances

| Model | Clean | Clean | StIM | StIM |
|---|---|---|---|---|
| Base model | 100 % | 56.66 % | 98.8 % | 93.33 % |
| AlexNet | 96.66 % | 40 % | 99.81 % | 83.33 % |
| VGG16 | 100 % | 53.33 % | 95.85 % | 90 % |

# 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.

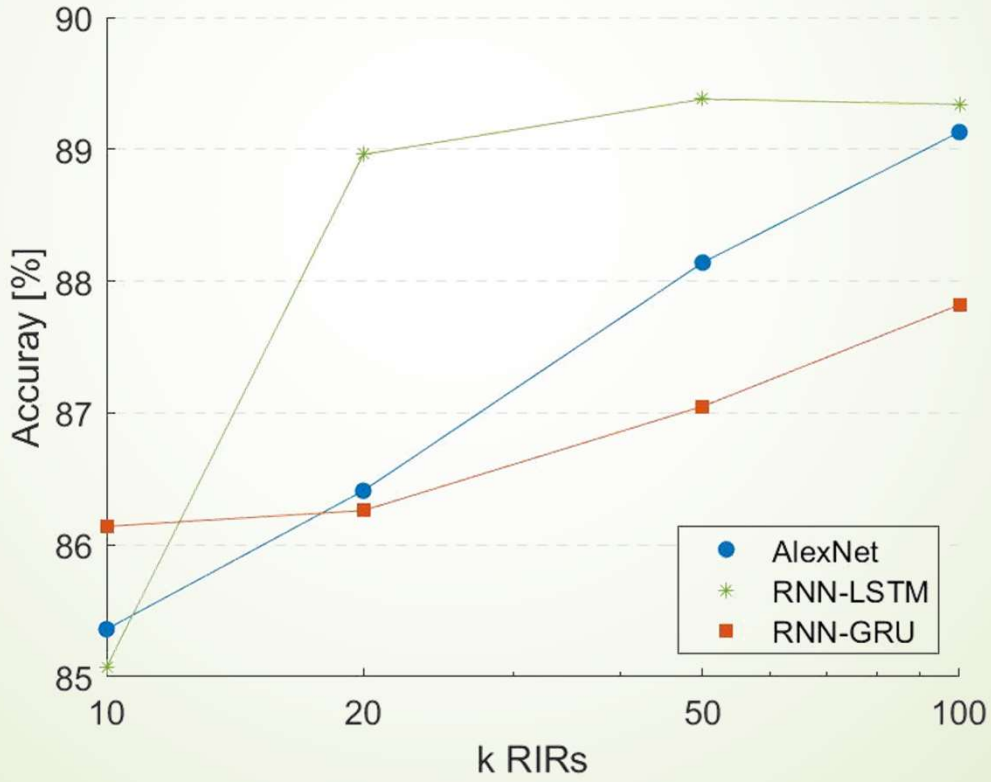
# Speech emotion recognition from beyond the wall:

Models

- AlexNet.
- LSTM-RNN.
- GRU-RNN.

# Speech emotion recognition from beyond the wall:

Results



# Speech emotion recognition from beyond the wall:

Results

| Model | Clean | Clean | StIM | StIM |
|---|---|---|---|---|
| AlexNet | 84.14 % | 20 % | 89.13 % | 50.66 % |
| RNN-LSTM | 83.74 % | 21.33 % | 89.34 % | 61.33 % |
| RNN-GRU | 85.94 % | 26.66 % | 87.82 % | 53.33 % |

# 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