Indian
Institute of
Technology
Mandi

RAJESH R (S21005) · SYNOPSIS SEMINAR · 08 FEBRUARY 2024

INTERFERENCE REDUCTION IN LIVE RECORDINGS FOR MUSIC SOURCE SEPARATION

MUSIC

MUSIC

How many instrument sources you can hear?

MUSIC

How many instrument sources you can hear?

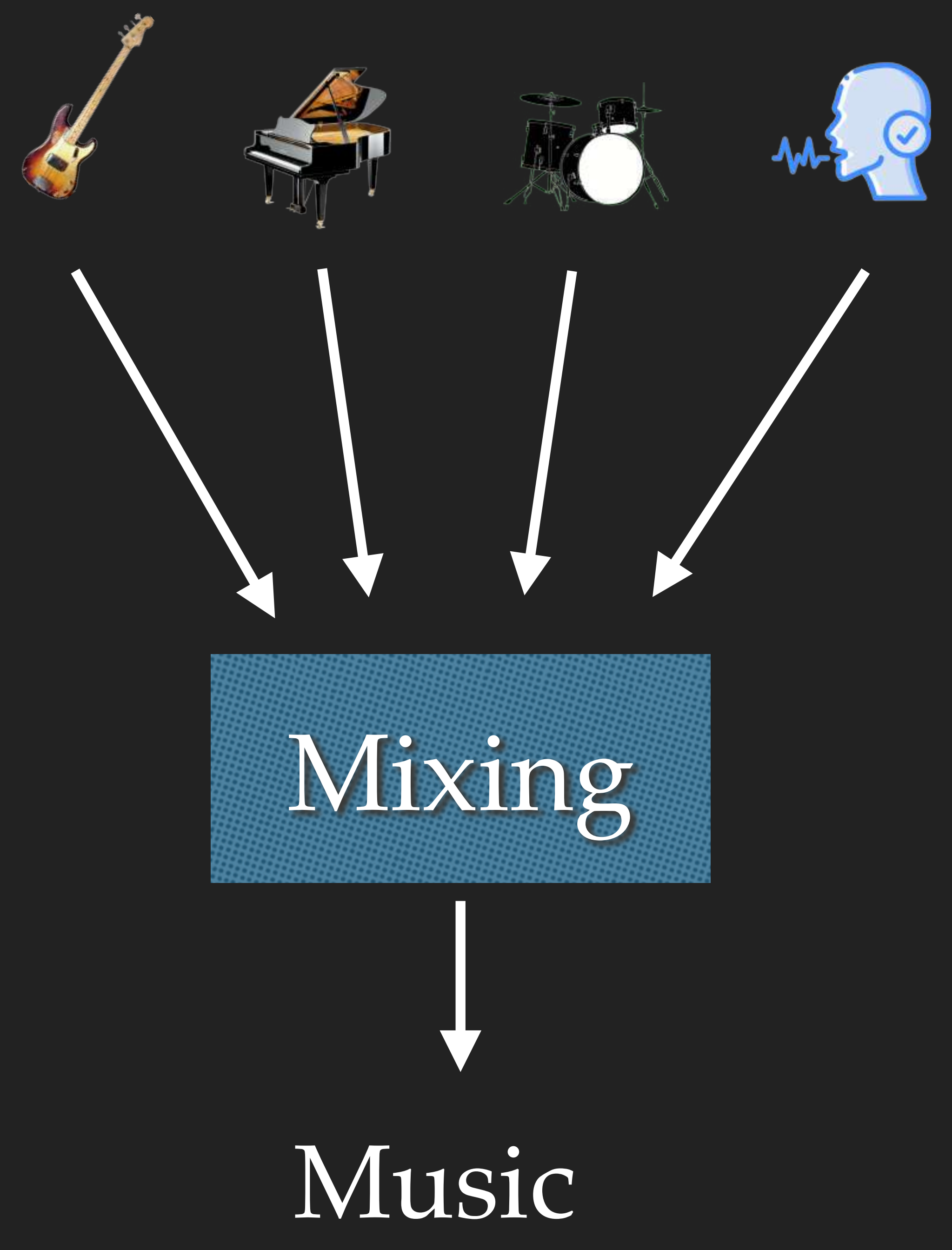
MUSIC

How many instrument sources you can hear?

5+

MUSIC PRODUCTION

Music remixing: DAW



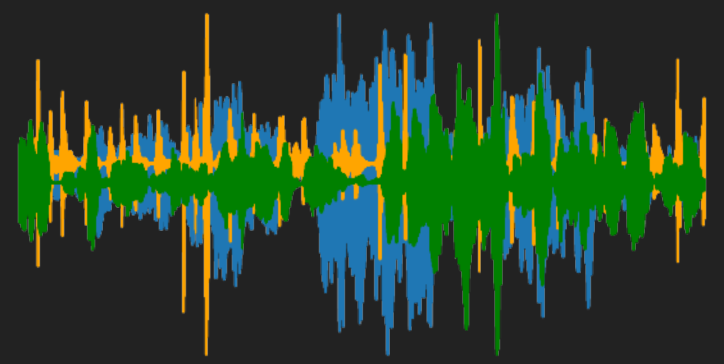
Mixing

Music

https://www.seekpng.com/ima/u2q8r5y3e6y3r5u2/

MUSIC SOURCE SEPARATION

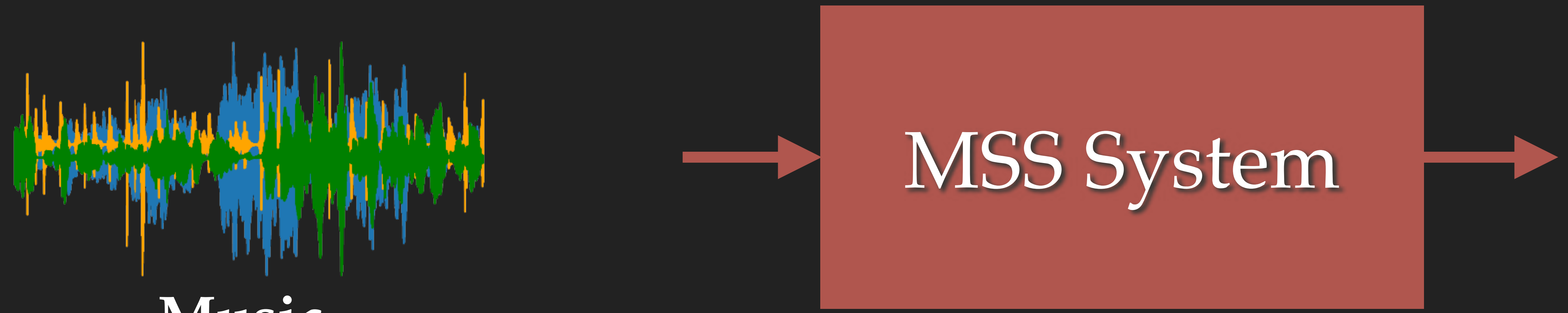
MUSIC SOURCE SEPARATION



Music

(Vocal(s), Bass, Drums and Other)

MUSIC SOURCE SEPARATION

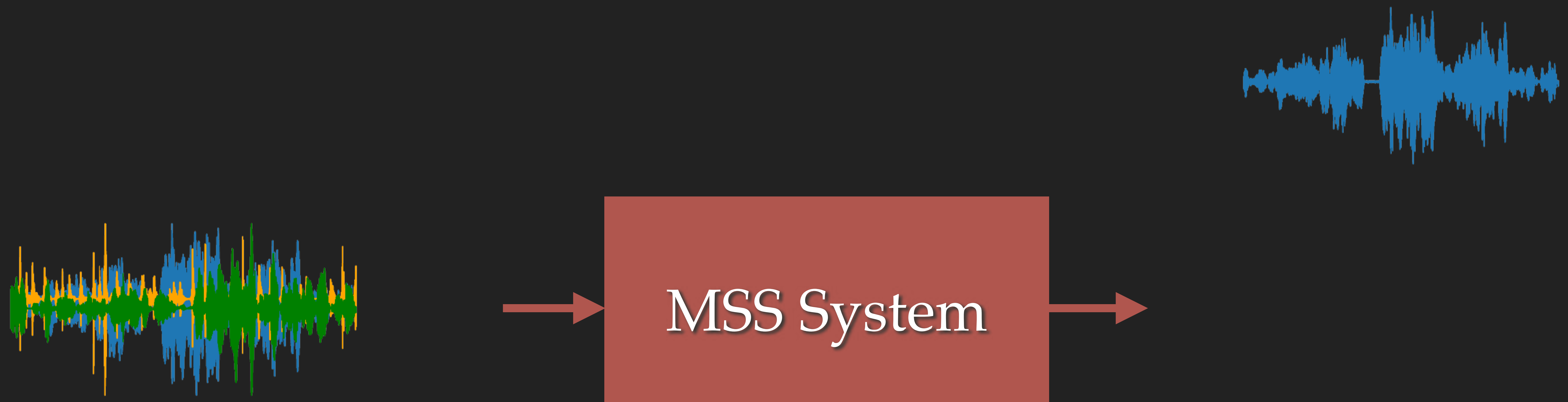


Music

(Vocal(s), Bass, Drums and Other)

MSS System

MUSIC SOURCE SEPARATION



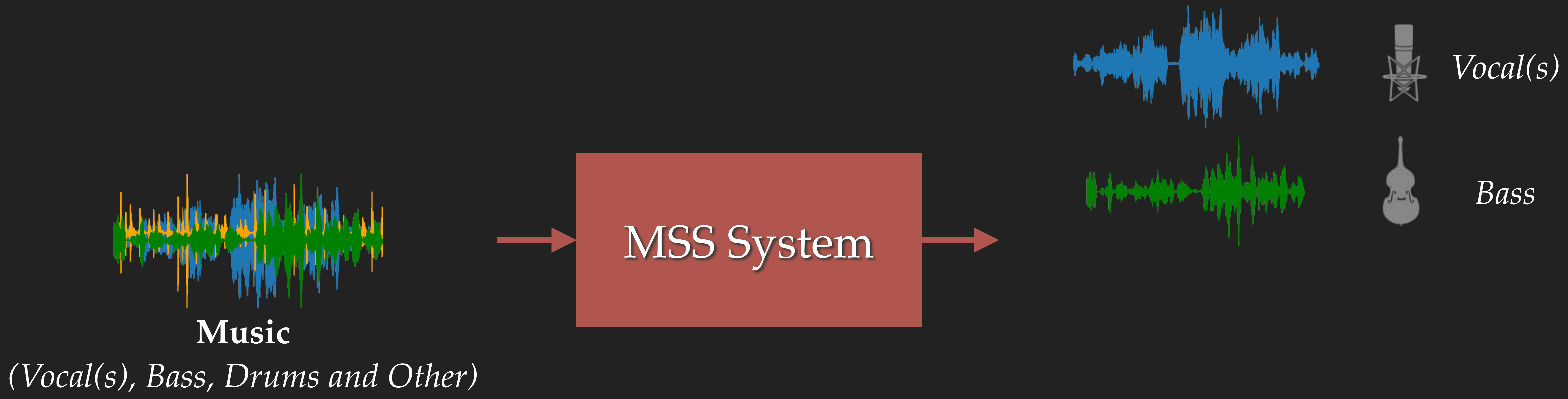
Music

(Vocal(s), Bass, Drums and Other)

MSS System

Vocal(s)

MUSIC SOURCE SEPARATION



Music

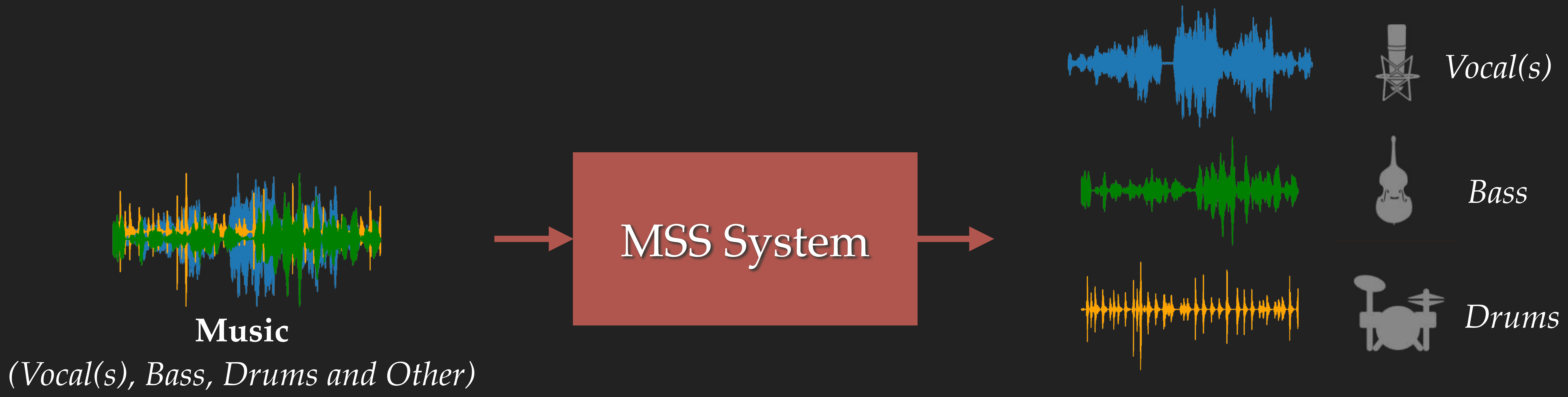
(Vocal(s), Bass, Drums and Other)

MSS System

Vocal(s)

Bass

MUSIC SOURCE SEPARATION



Music

(Vocal(s), Bass, Drums and Other)

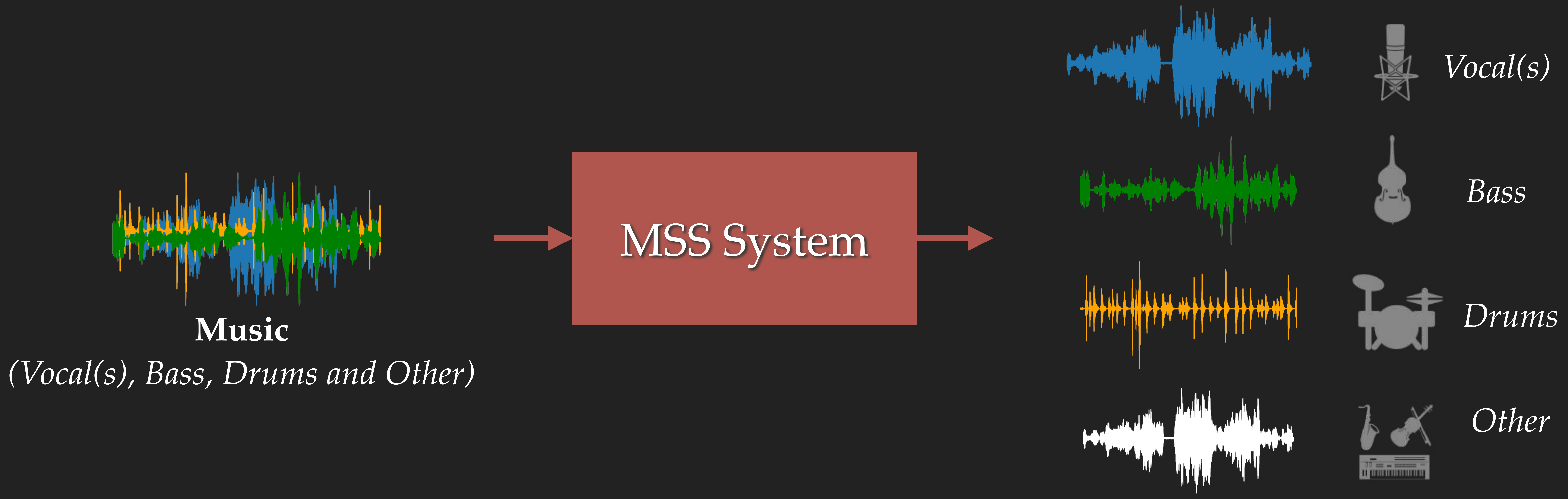
MSS System

Vocal(s)

Bass

Drums

MUSIC SOURCE SEPARATION



Music

(Vocal(s), Bass, Drums and Other)

MSS System

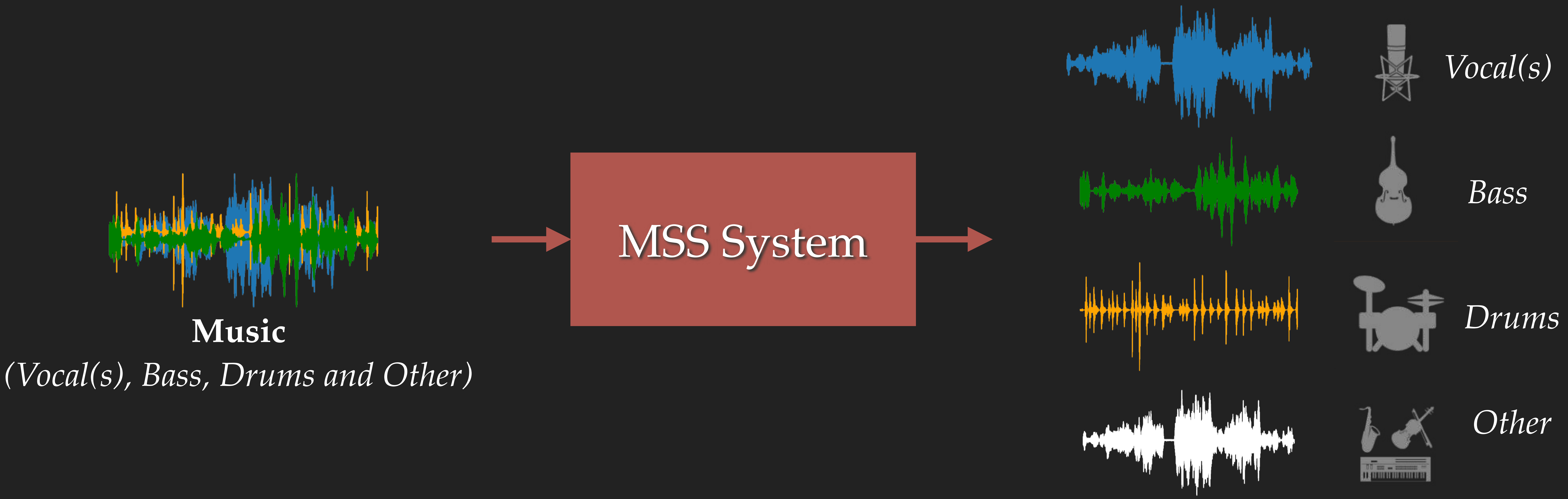
Vocal(s)

Bass

Drums

Other

MUSIC SOURCE SEPARATION



Music

(Vocal(s), Bass, Drums and Other)

MSS System

Vocal(s)

Bass

Drums

Other

WHY MUSIC SOURCE SEPARATION?

APPLICATIONS

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

APPLICATIONS

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

APPLICATIONS

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

APPLICATIONS

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

APPLICATIONS

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

Education & Learning

APPLICATIONS

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

APPLICATIONS

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

Education & Learning

Music Transcription

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

Education & Learning

APPLICATIONS

Music Transcription

Automatic Accompaniment Generation

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

Education & Learning

APPLICATIONS

Music Transcription

Automatic Accompaniment Generation

Health & Wellbeing

Speech Enhancement

WHY MUSIC SOURCE SEPARATION?

Music Production & Remixing

Automatic Music Tagging & Classification

Audio Restoration

Education & Learning

APPLICATIONS

Music Transcription

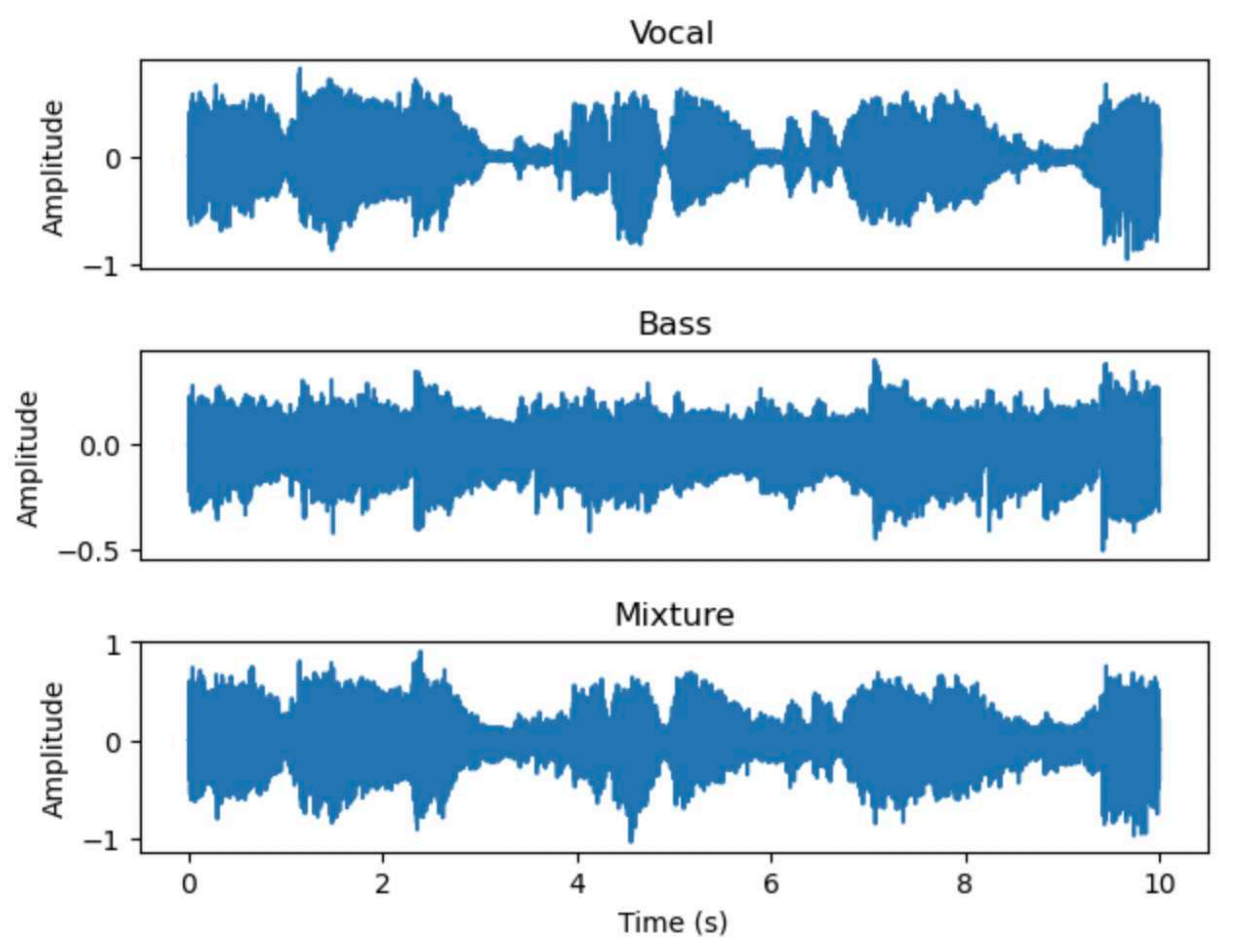
Automatic Accompaniment Generation

Health & Wellbeing

Music Information Retrieval

Speech Enhancement

IS HARD PROBLEM?



IS HARD PROBLEM?

IS HARD PROBLEM?

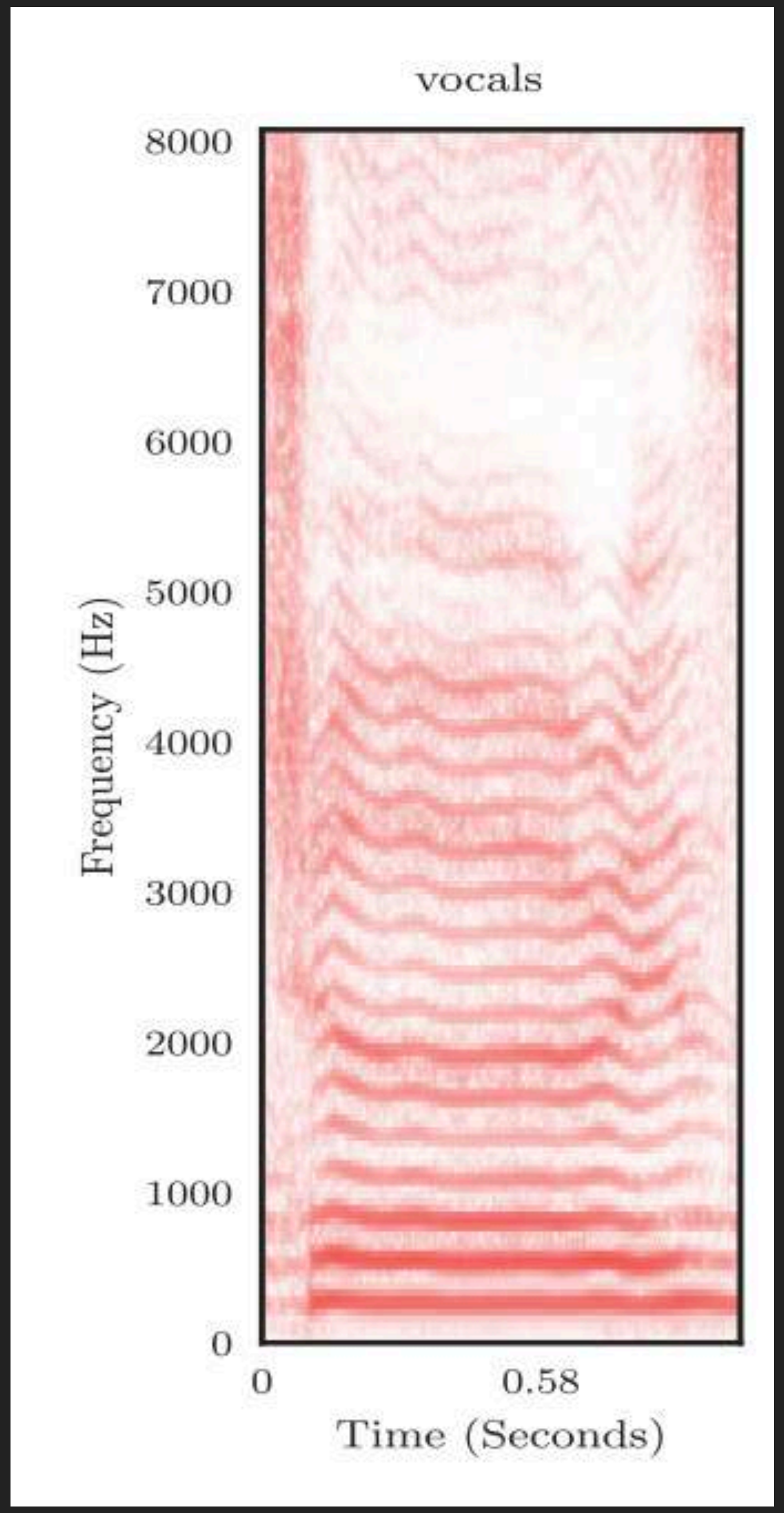


Figure taken from: *Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, Fabian-Robert Stöter. Musical Source Separation: An Introduction. IEEE Signal Processing Magazine, Institute of Electrical and Electronics Engineers, 2019, 36 (1), pp.31-40.*

IS HARD PROBLEM?

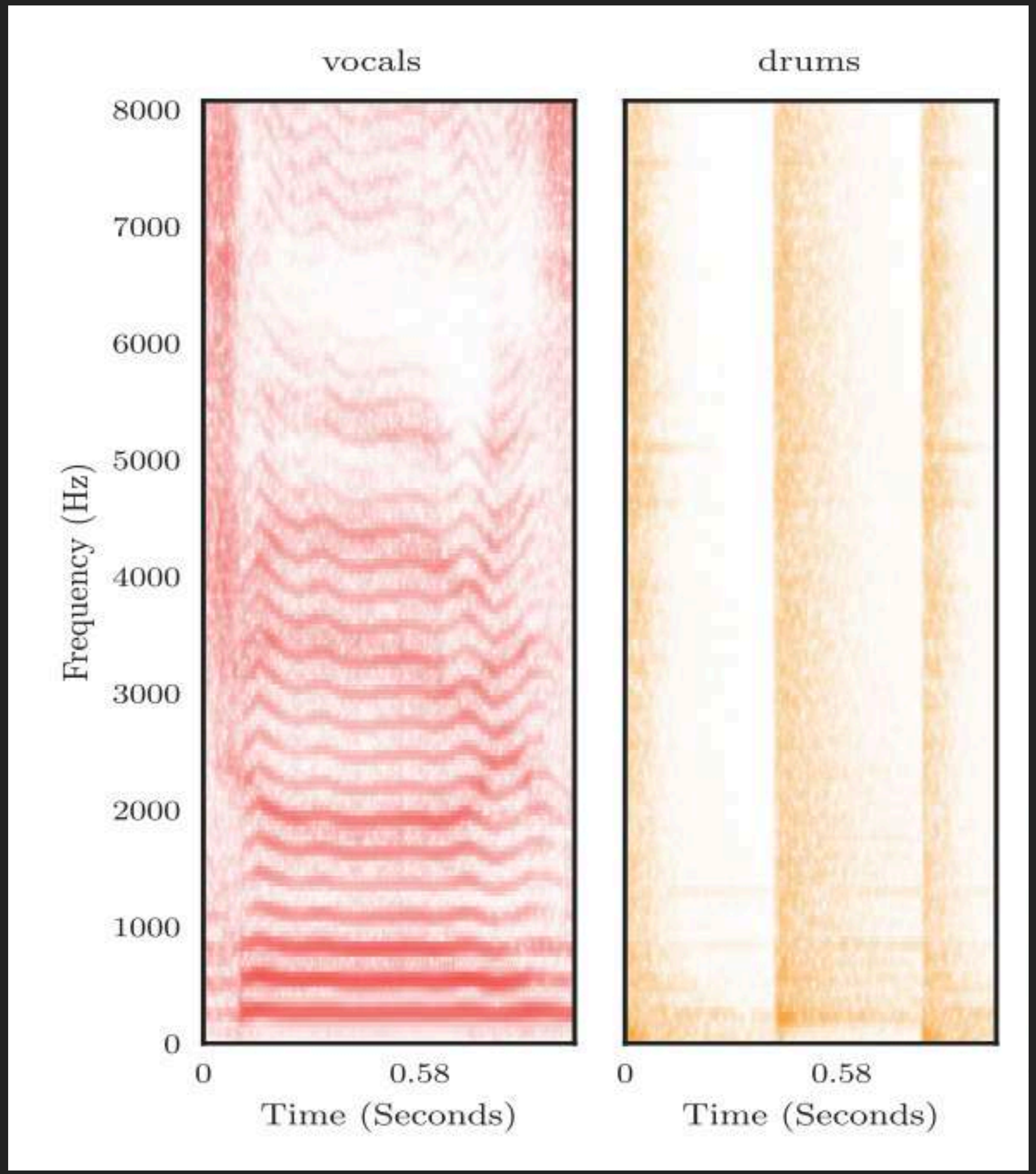


***Figure taken from**: Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, Fabian-Robert Stöter. Musical Source Separation: An Introduction. IEEE Signal Processing Magazine, Institute of Electrical and Electronics Engineers, 2019, 36 (1), pp.31-40.*

IS HARD PROBLEM?

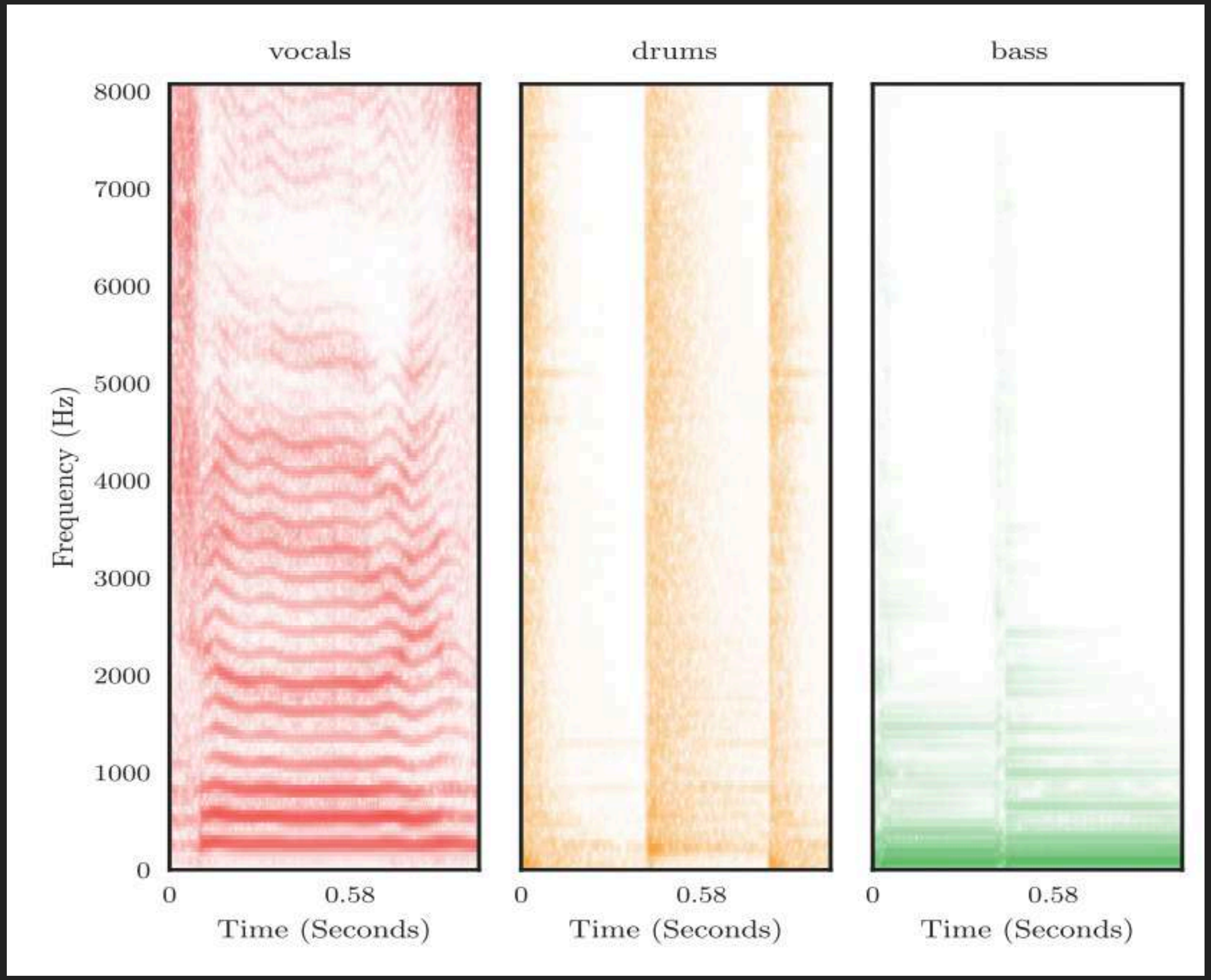


***Figure taken from**: Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, Fabian-Robert Stöter. Musical Source Separation: An Introduction. IEEE Signal Processing Magazine, Institute of Electrical and Electronics Engineers, 2019, 36 (1), pp.31-40.*

IS HARD PROBLEM?

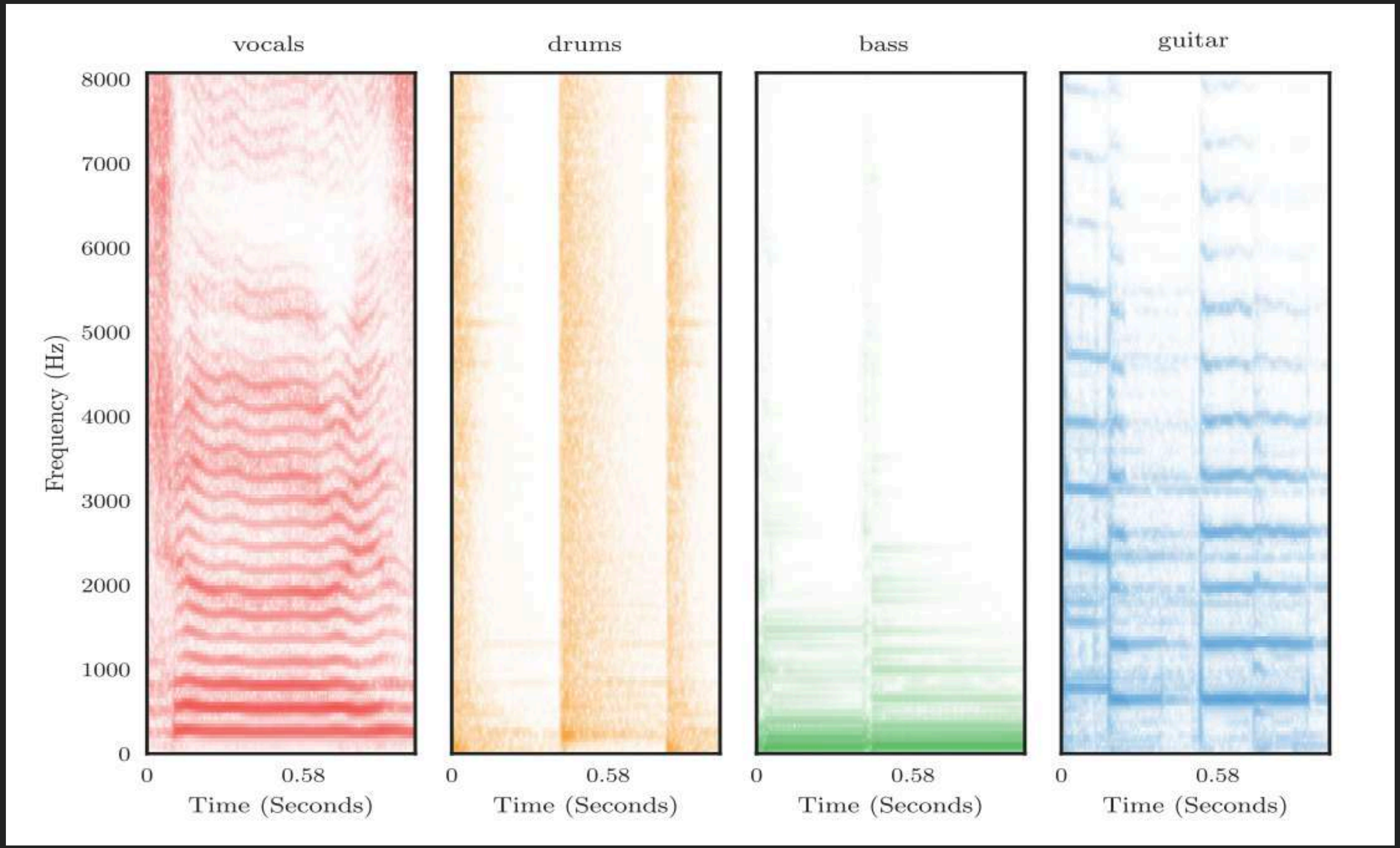


Figure taken from: *Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, Fabian-Robert Stöter. Musical Source Separation: An Introduction. IEEE Signal Processing Magazine, Institute of Electrical and Electronics Engineers, 2019, 36 (1), pp.31-40.*

IS HARD PROBLEM?

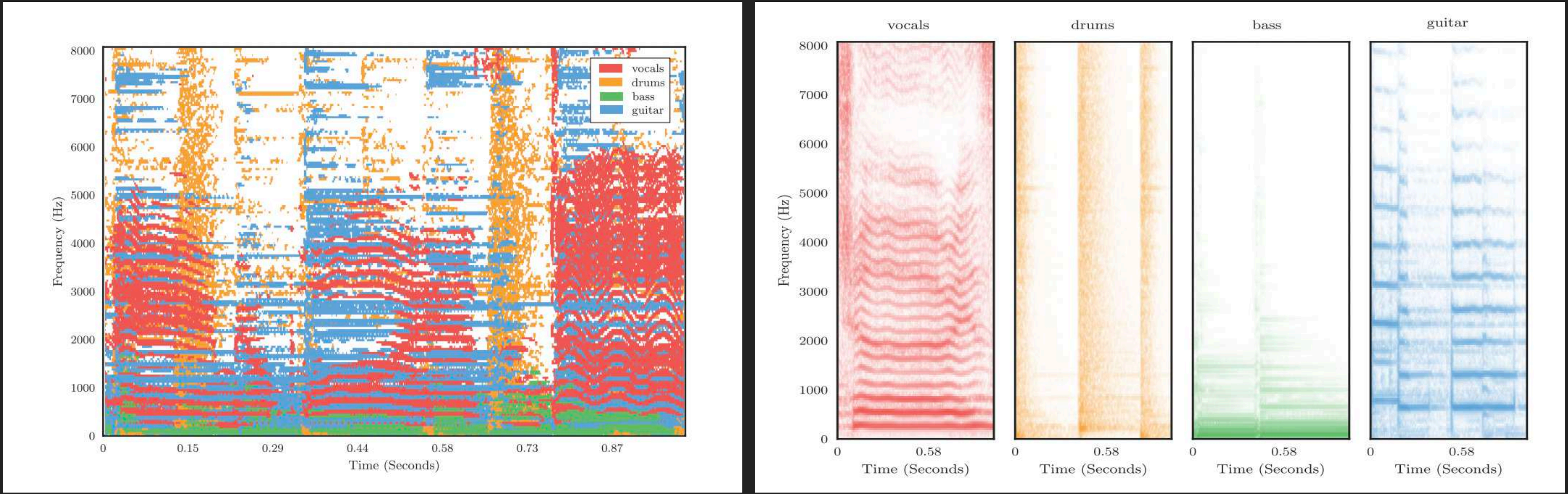


***Figure taken from**: Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, Fabian-Robert Stöter. Musical Source Separation: An Introduction. IEEE Signal Processing Magazine, Institute of Electrical and Electronics Engineers, 2019, 36 (1), pp.31-40.*

QUANTITATIVE METRICS

QUANTITATIVE METRICS

Source to Distortion Ratio

$$SDR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf} + e_{noise} + e_{arti}\|^2}$$

QUANTITATIVE METRICS

Source to Distortion Ratio

$$SDR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf} + e_{noise} + e_{arti}\|^2}$$

Source to Artifact Ratio

$$SAR = 10log_{10}\frac{\|S_{target} + e_{noise} + e_{arti}\|^2}{\|e_{interf}\|^2}$$

QUANTITATIVE METRICS

Source to Distortion Ratio

$$SDR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf} + e_{noise} + e_{arti}\|^2}$$

Source to Artifact Ratio

$$SAR = 10log_{10}\frac{\|S_{target} + e_{noise} + e_{arti}\|^2}{\|e_{interf}\|^2}$$

Source to Interference Ratio

$$SIR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf}\|^2}$$

QUANTITATIVE METRICS

Source to Distortion Ratio

$$SDR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf} + e_{noise} + e_{arti}\|^2}$$

Scale Invariant Source to Distortion Ratio

$$SI - SDR = 10log_{10}\frac{\|s\|^2}{\|s - \hat{s}\|^2} = 10log_{10}\frac{\|\alpha s\|^2}{\|\alpha s - \hat{s}\|^2}$$

$$\alpha = \frac{\hat{s}^T s}{\|s\|^2}$$

Source to Artifact Ratio

$$SAR = 10log_{10}\frac{\|S_{target} + e_{noise} + e_{arti}\|^2}{\|e_{interf}\|^2}$$

Source to Interference Ratio

$$SIR = 10log_{10}\frac{\|S_{target}\|^2}{\|e_{interf}\|^2}$$

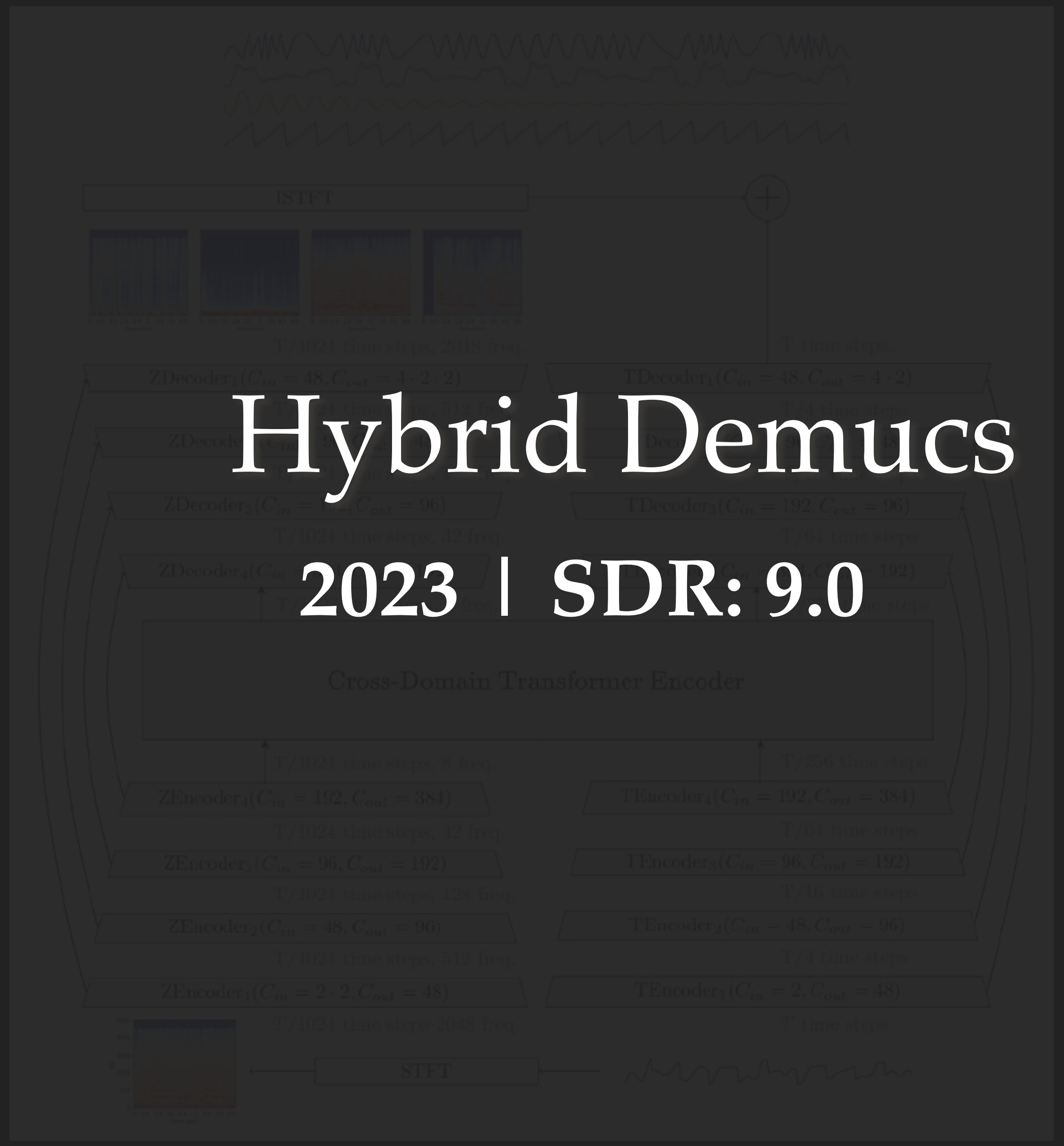
THE STATE-OF-THE-ART

Wave-U-Net

2018 | SDR: 3.2

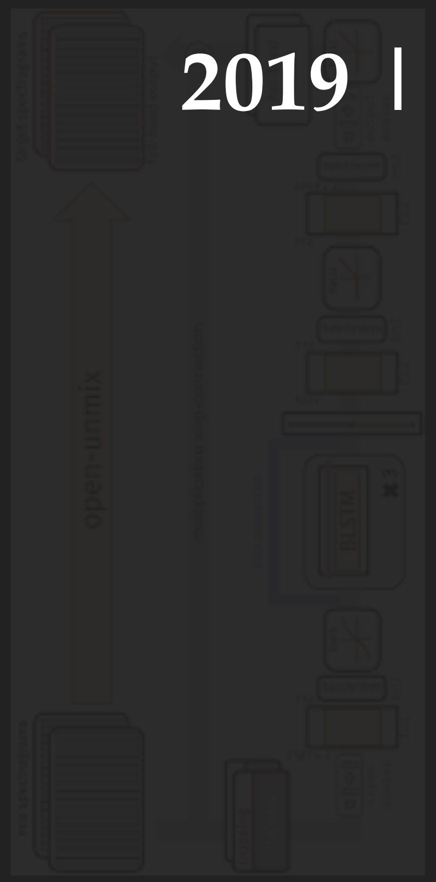
Band Split RNN

2023 | SDR: 9.0



Hybrid Demucs

2023 | SDR: 9.0



Open UnMix

2019 | SDR: 5.3

Spleeter

2020 | SDR: 5.9

Western Pop Music
MUSDB18

Spotify Google

facebook research

RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

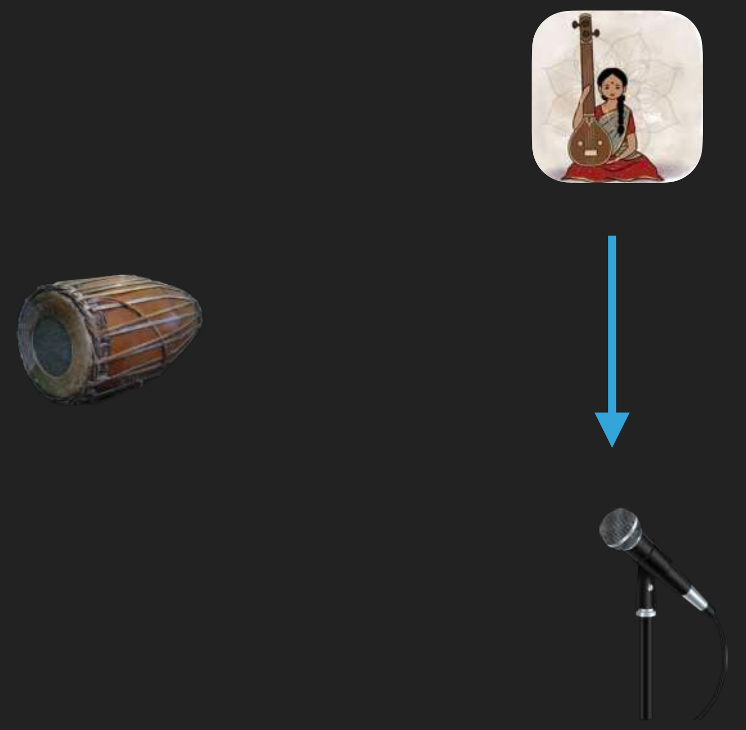
https://images.app.goo.gl/g9MPV2bNE5faJz4M7

RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

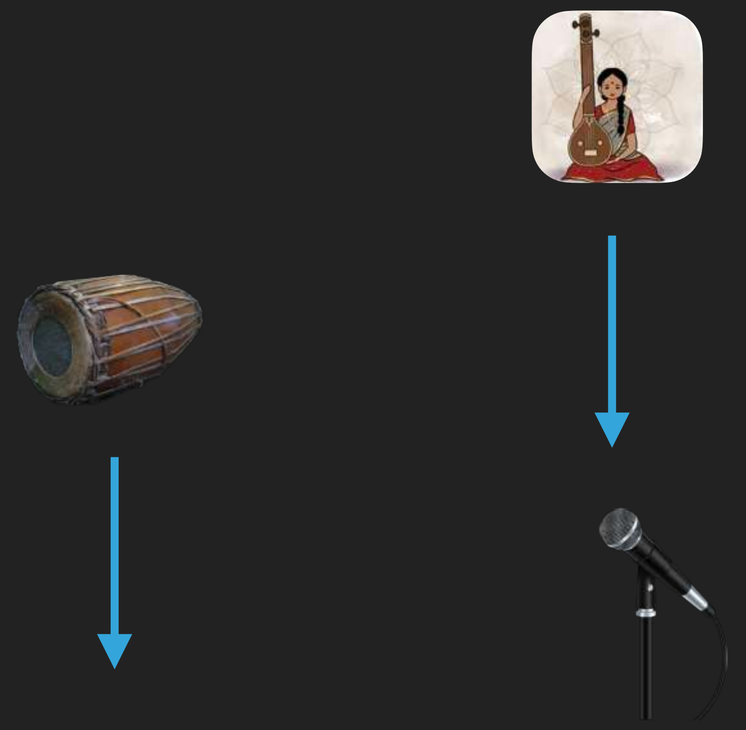
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



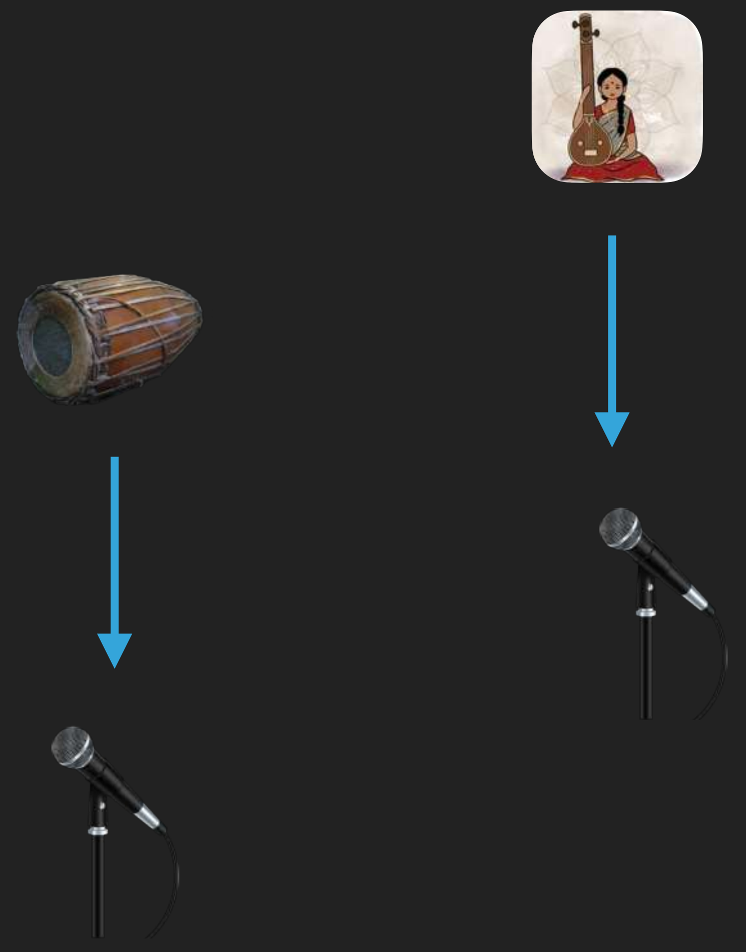
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



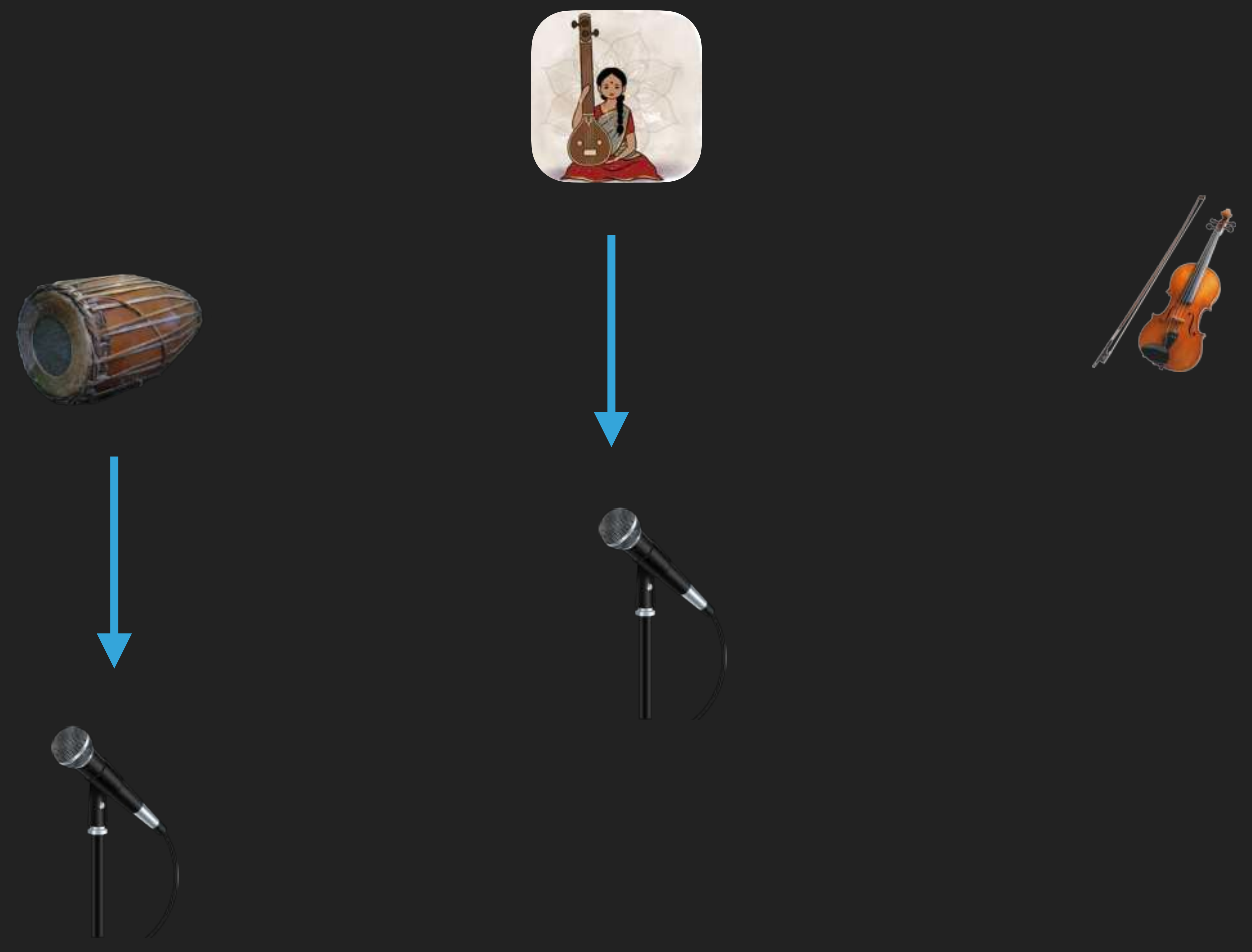
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



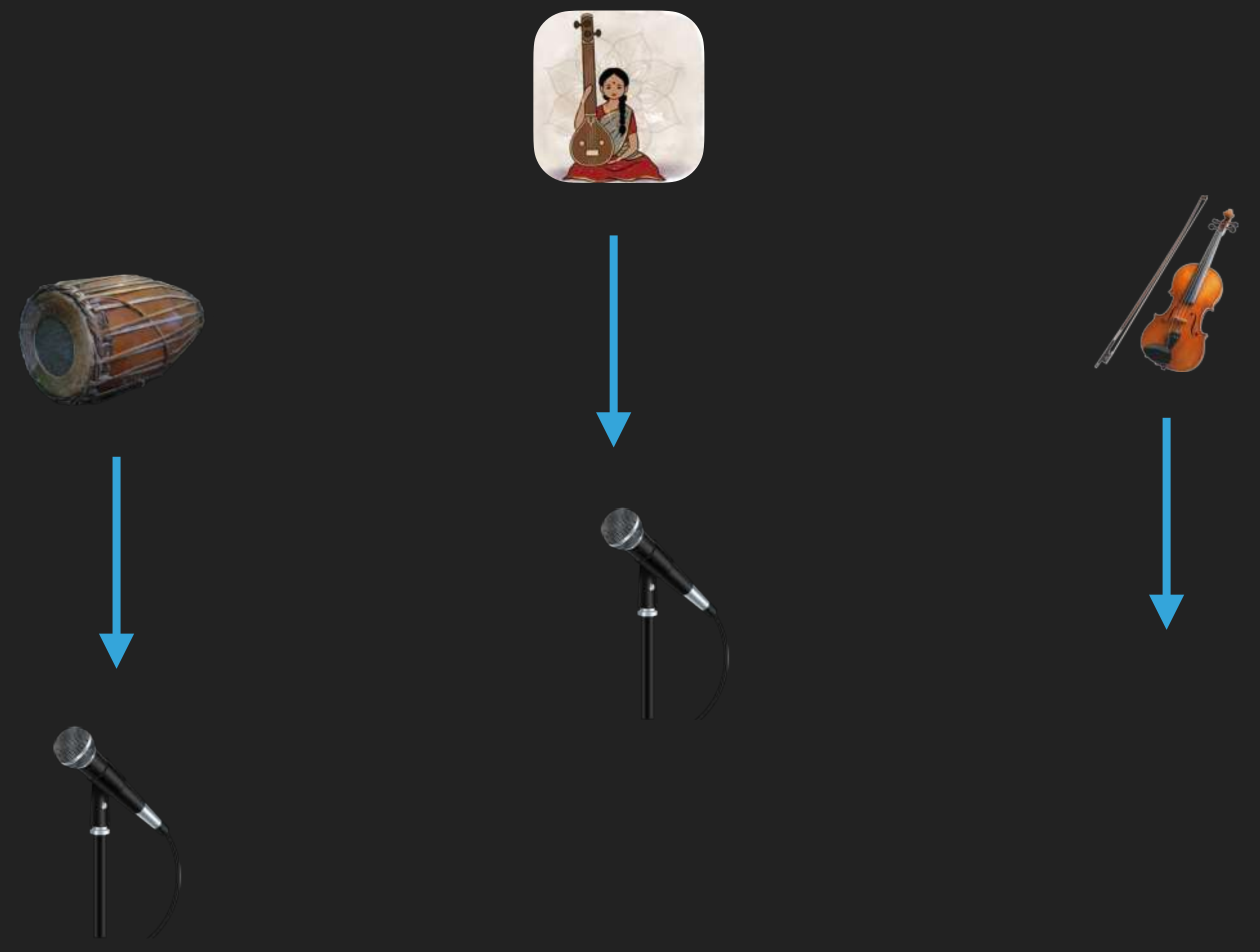
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



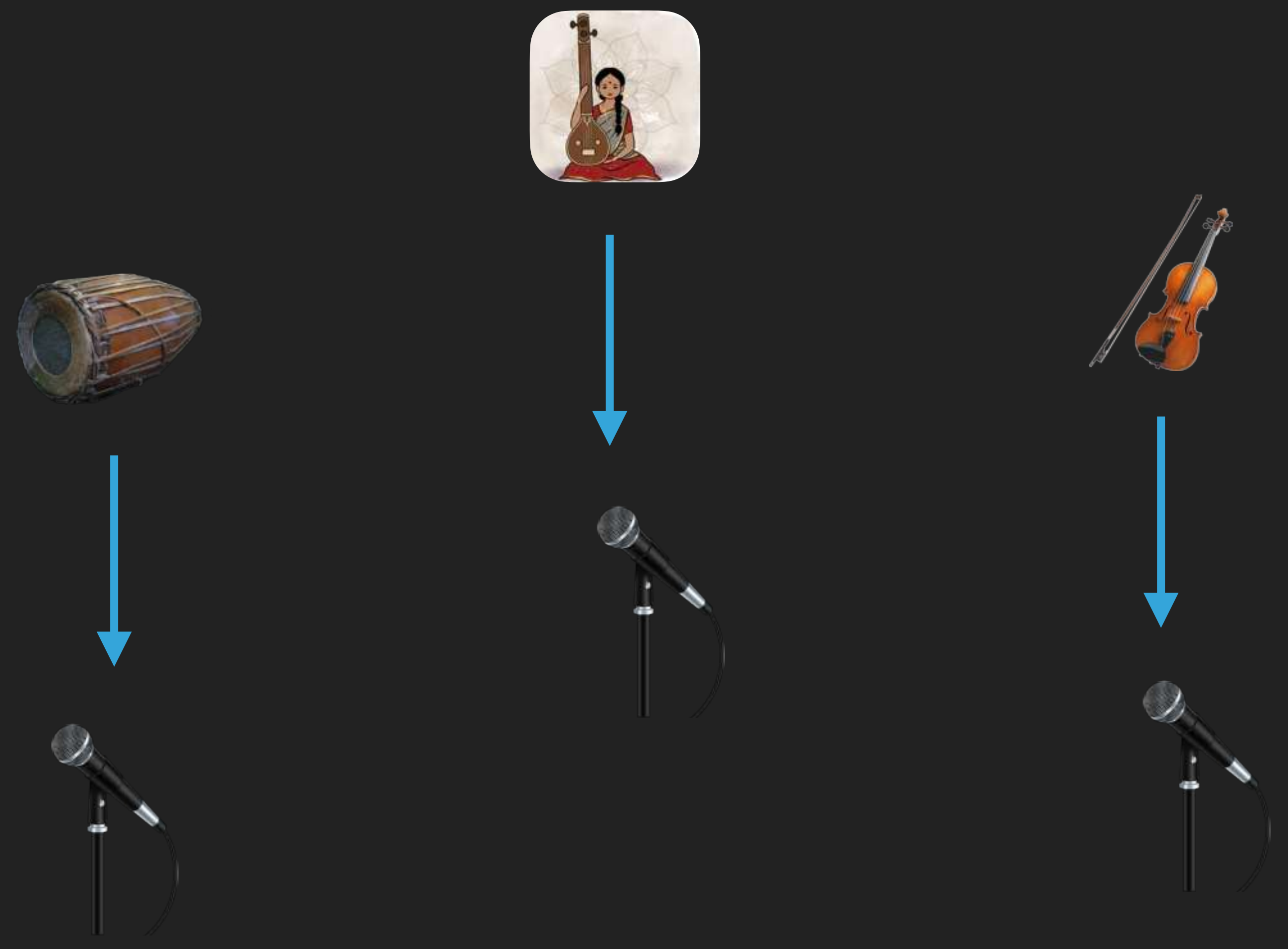
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



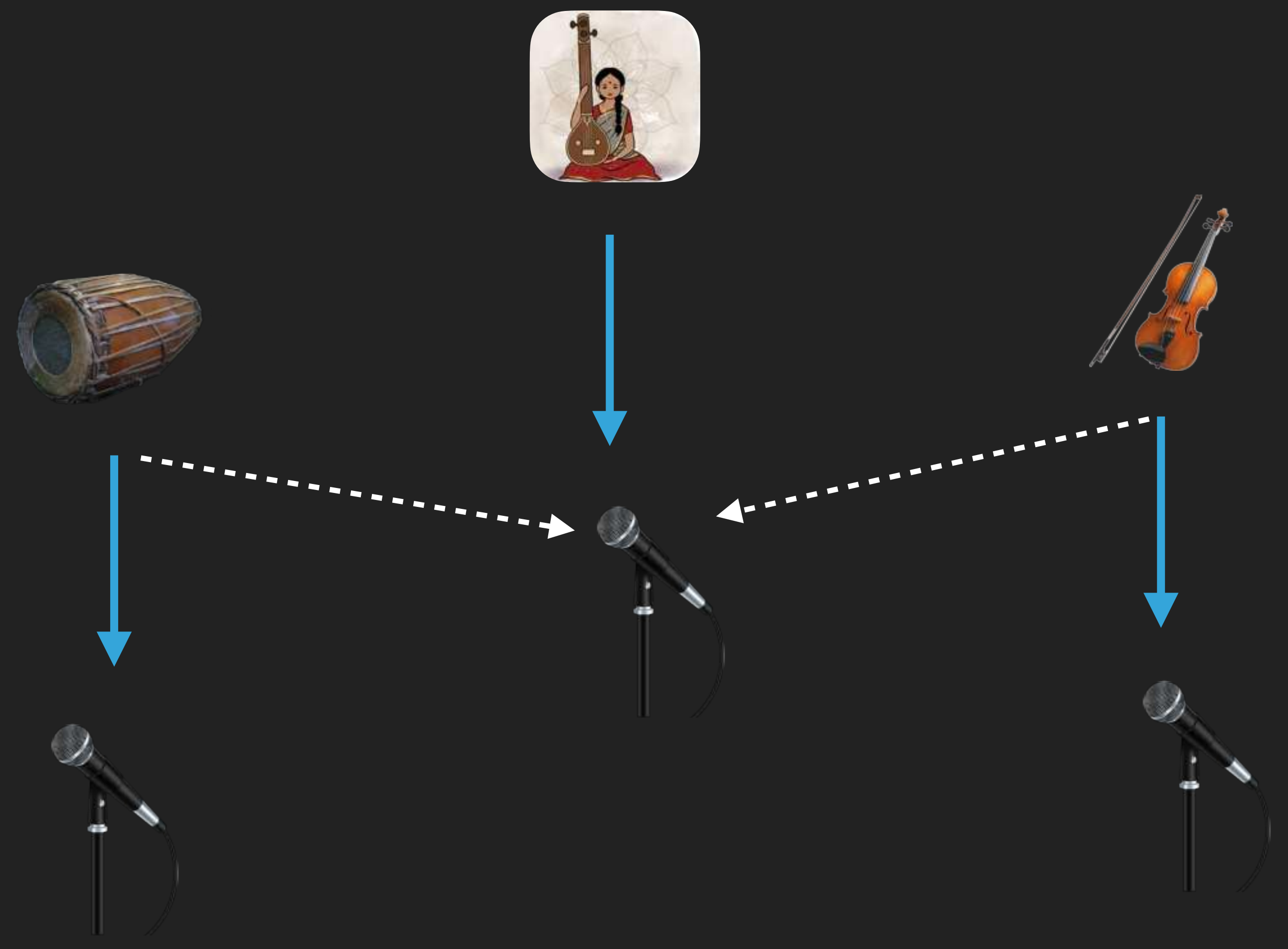
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



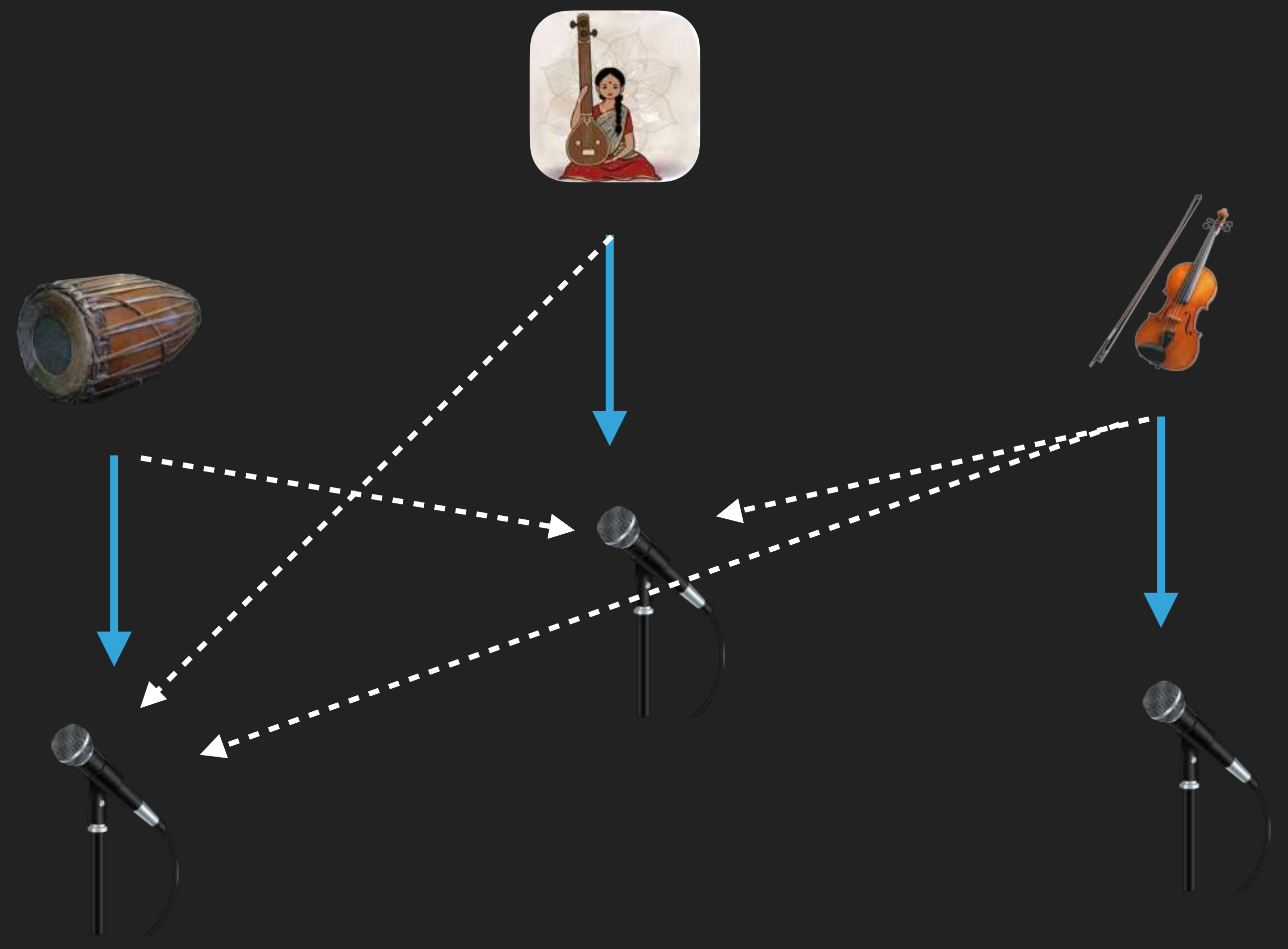
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



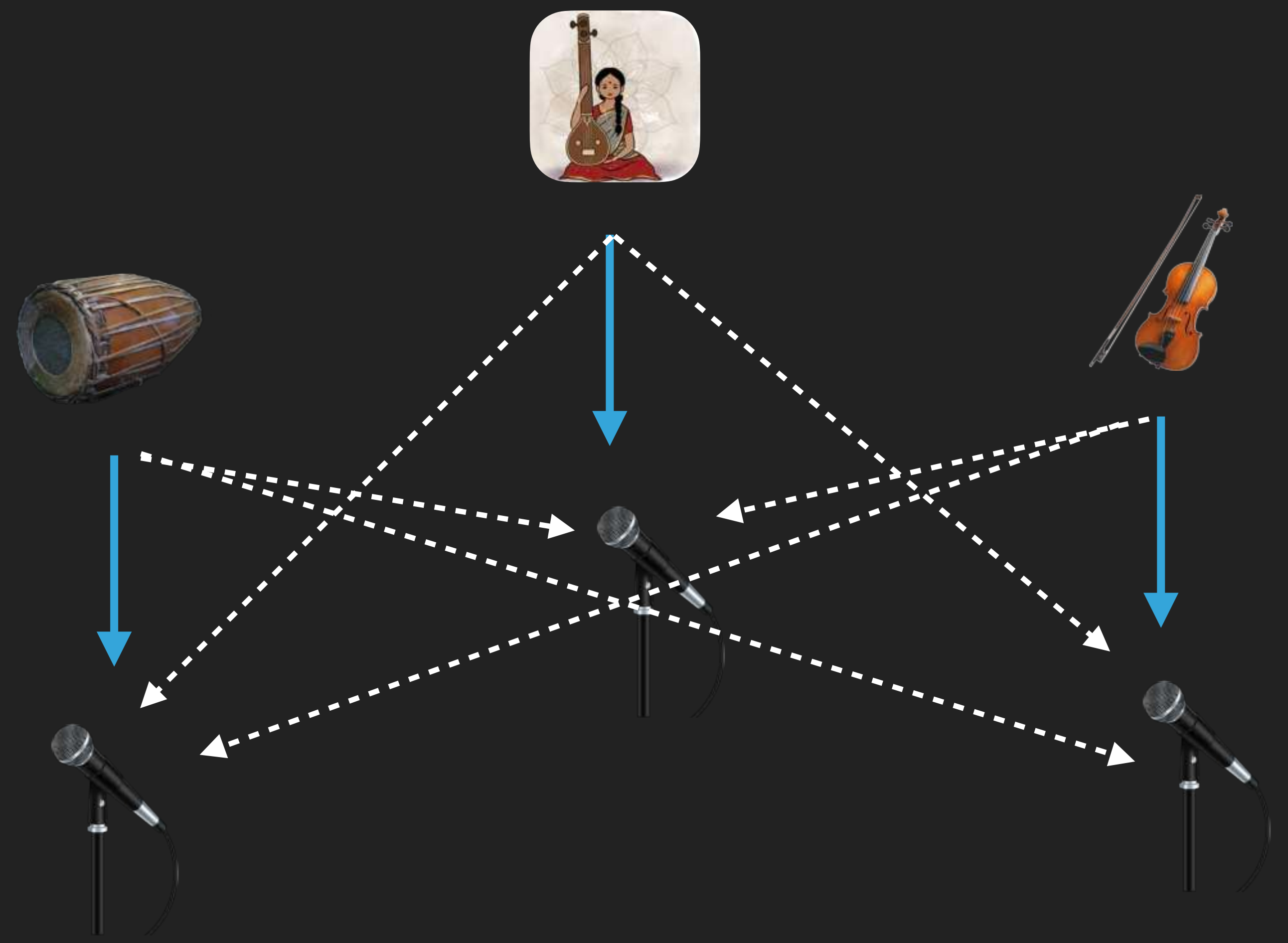
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



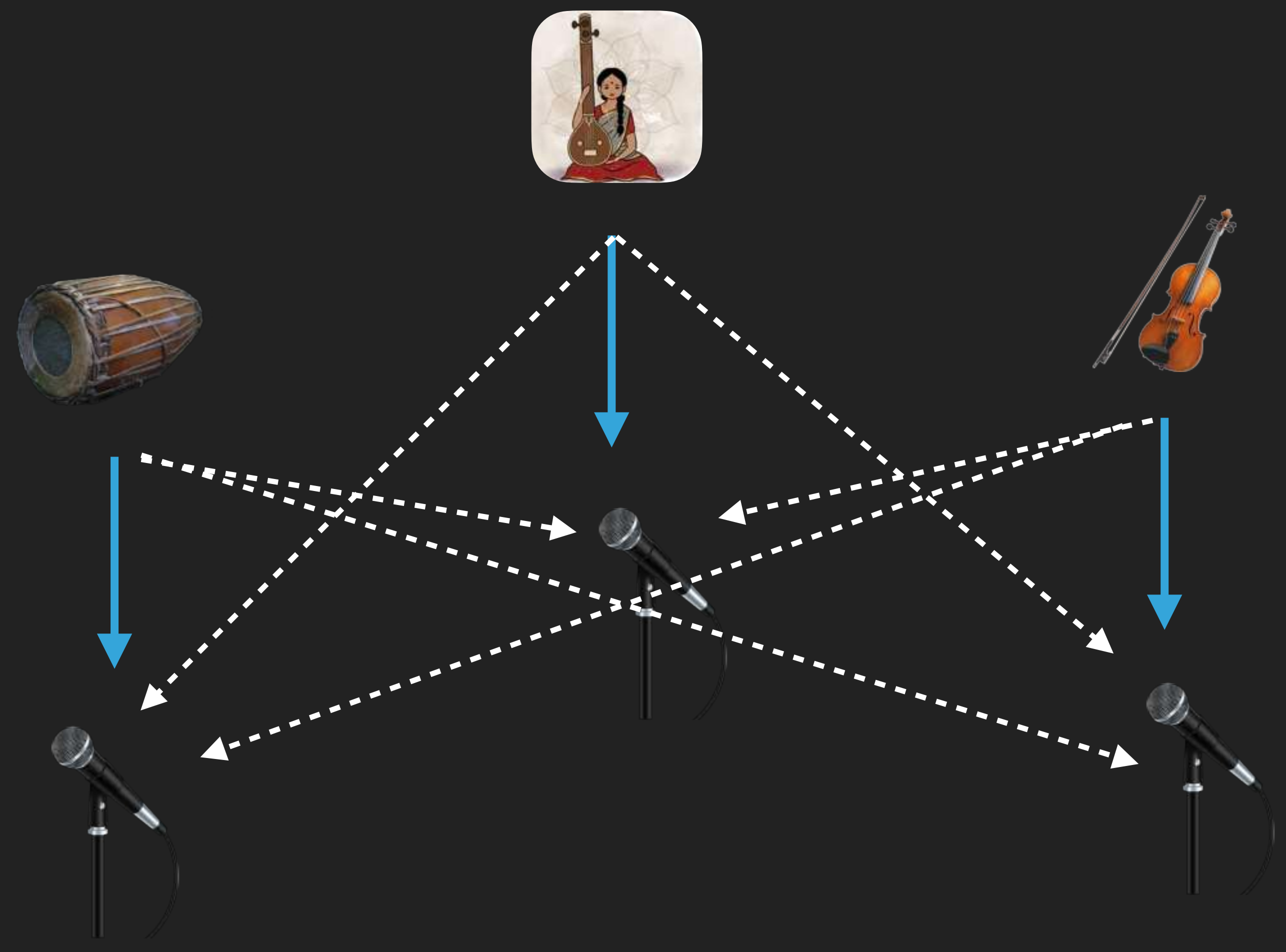
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

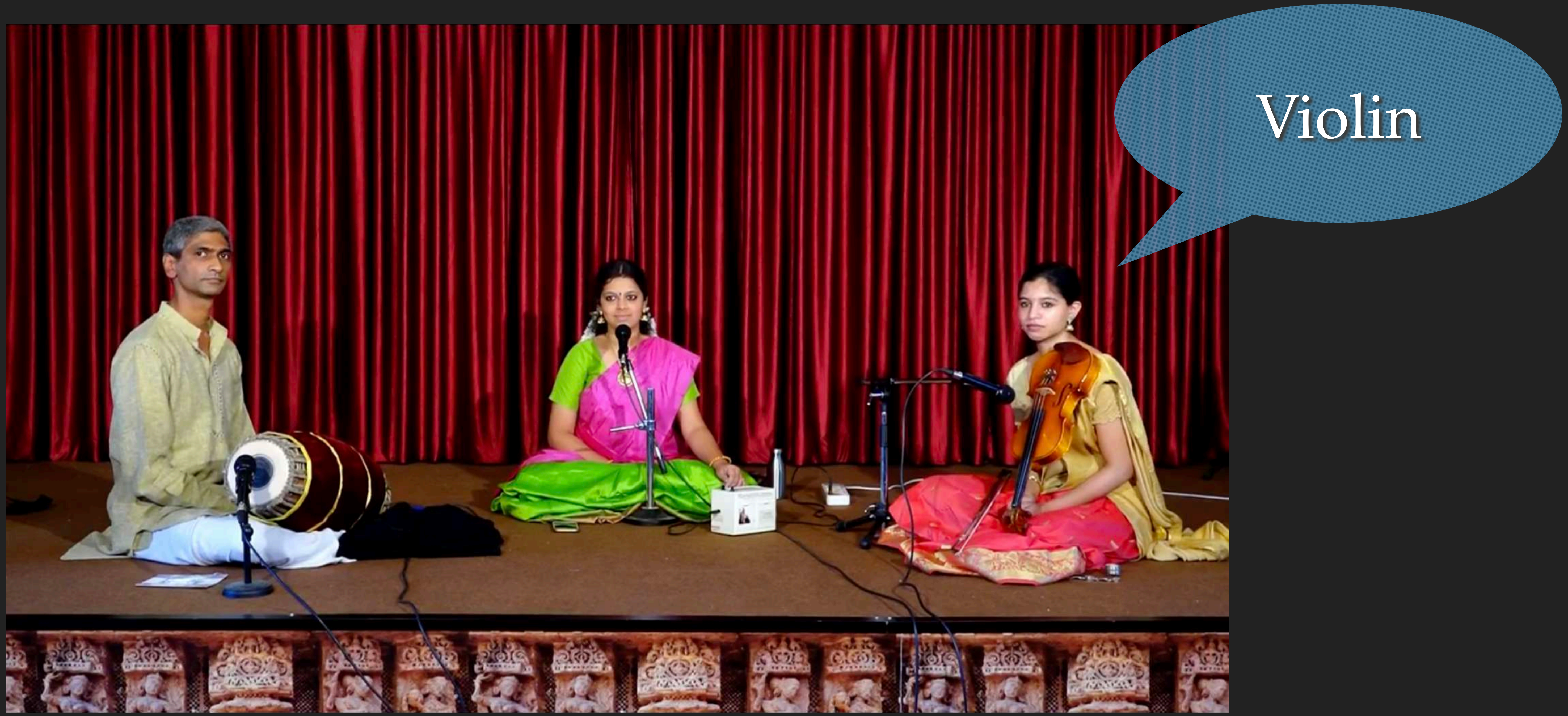


RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

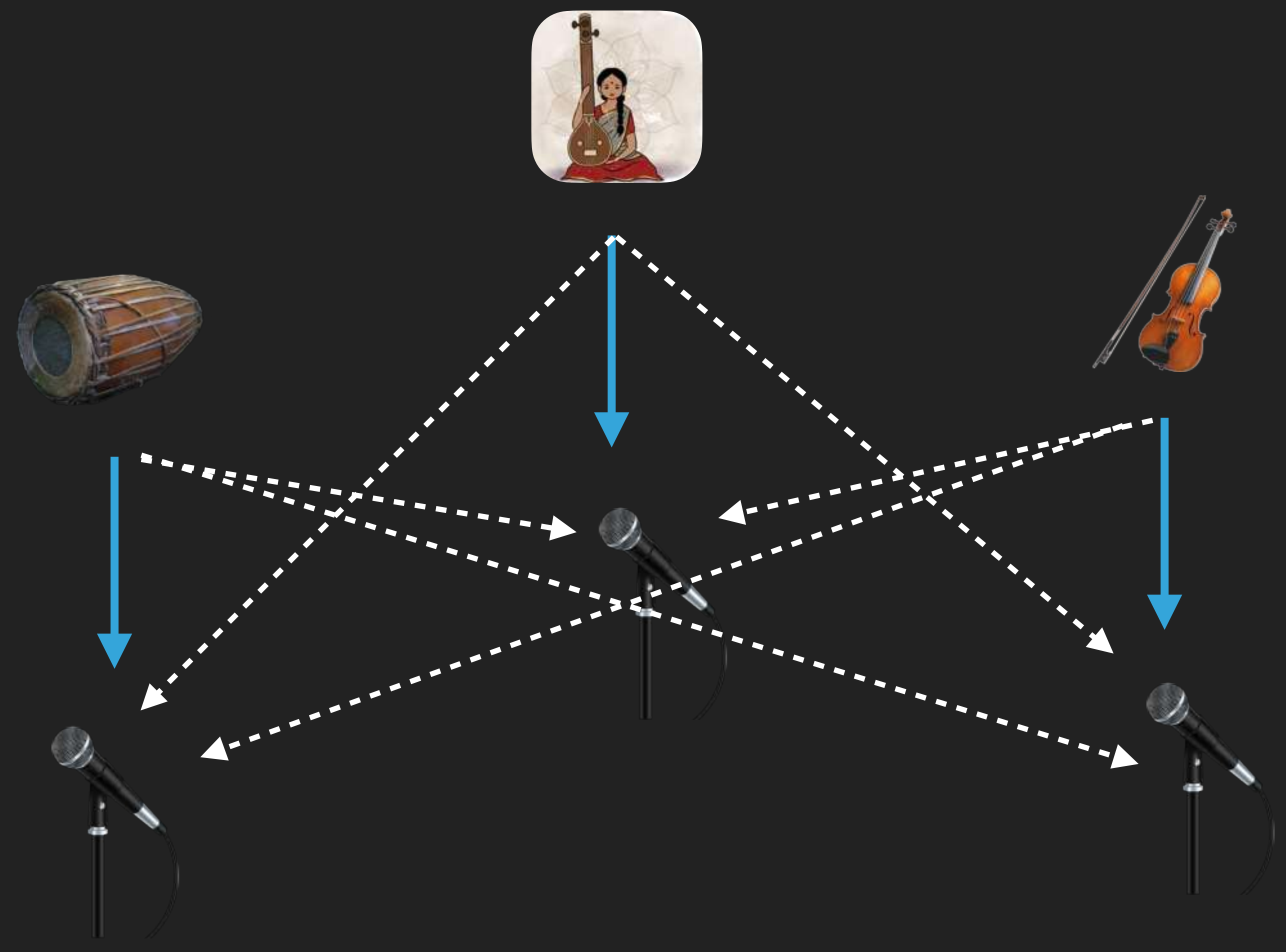
https://images.app.goo.gl/g9MPV2bNE5faJz4M7



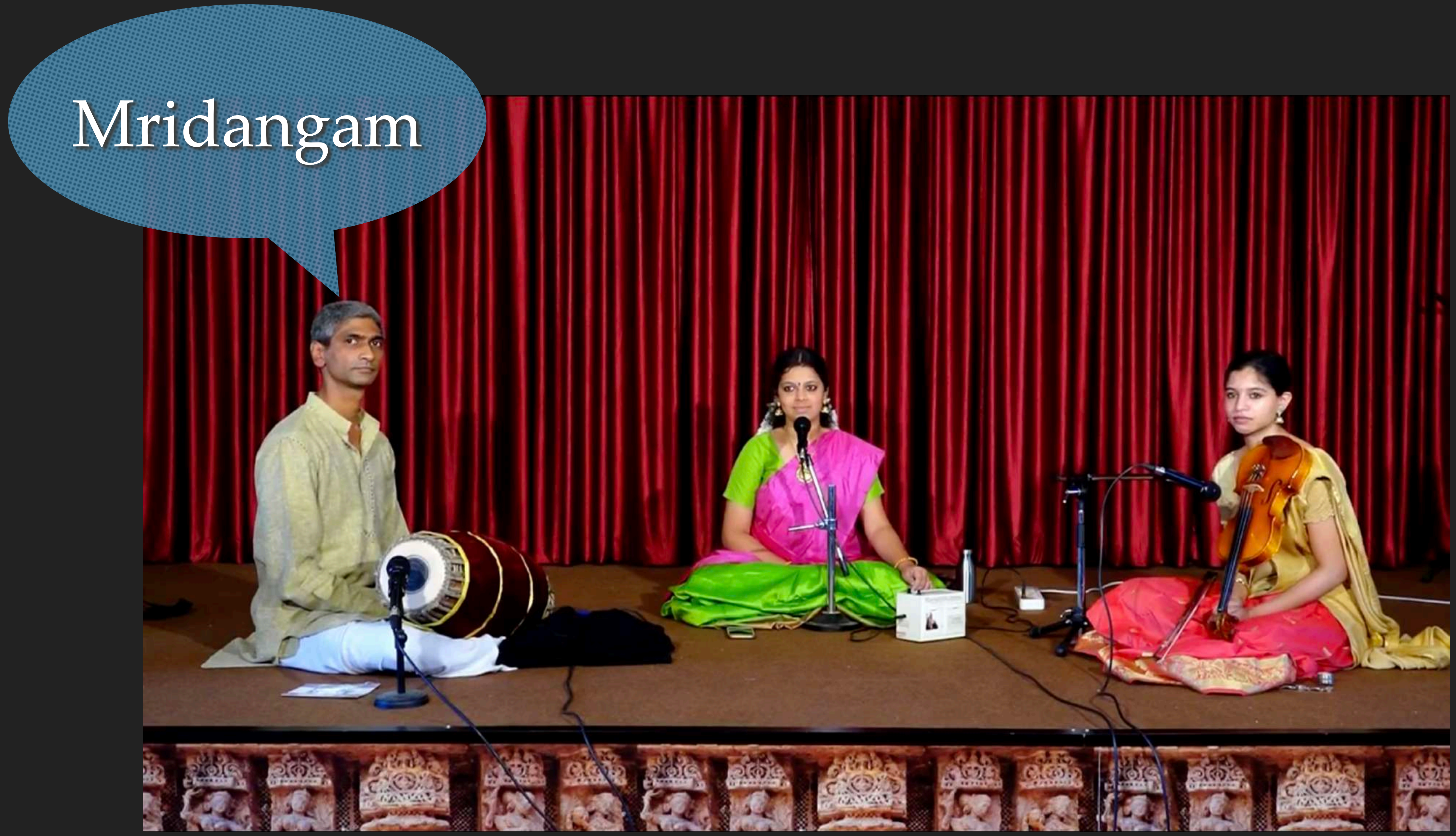
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)



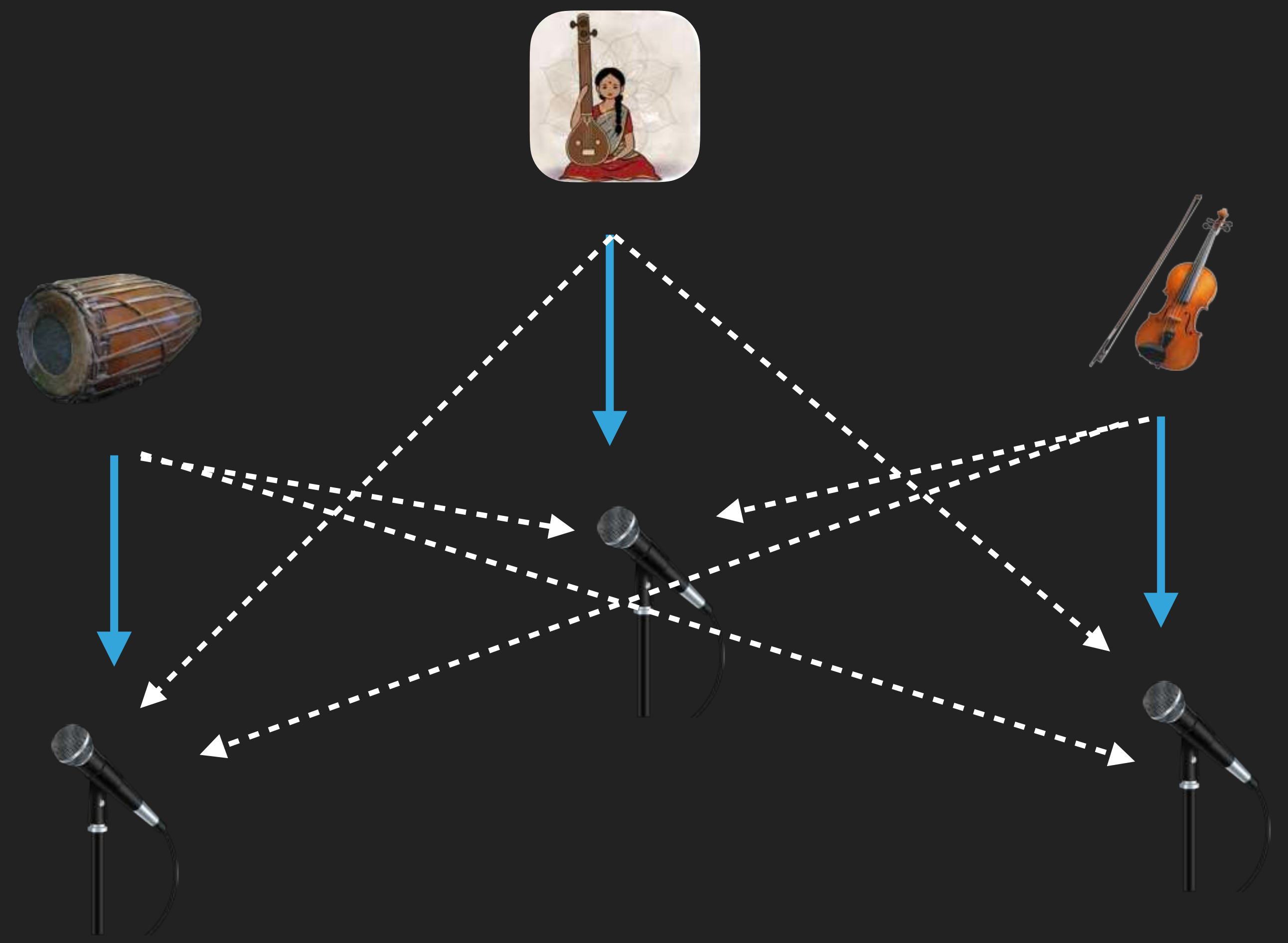
https://images.app.goo.gl/g9MPV2bNE5faJz4M7



RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

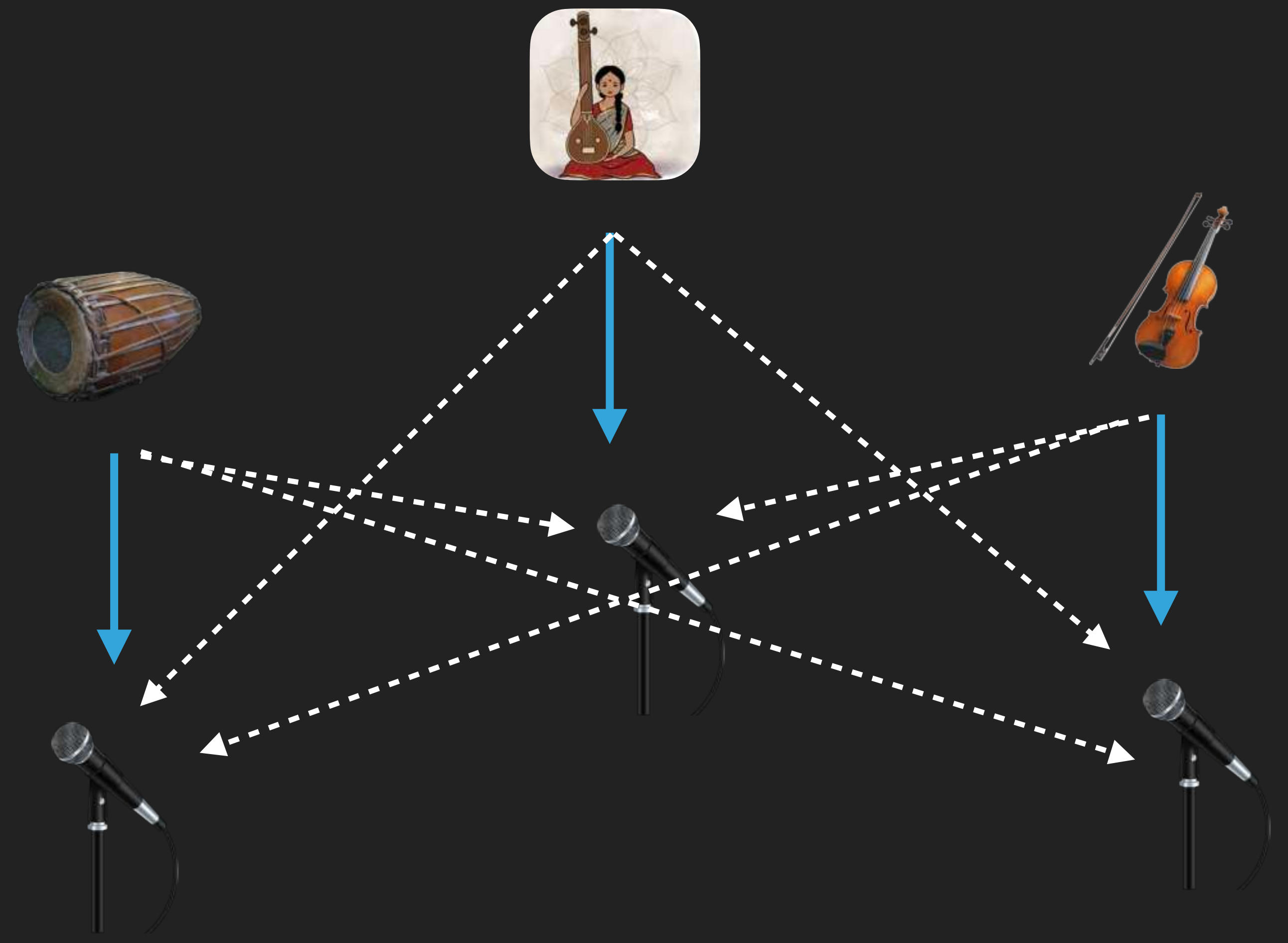


https://images.app.goo.gl/g9MPV2bNE5faJz4M7



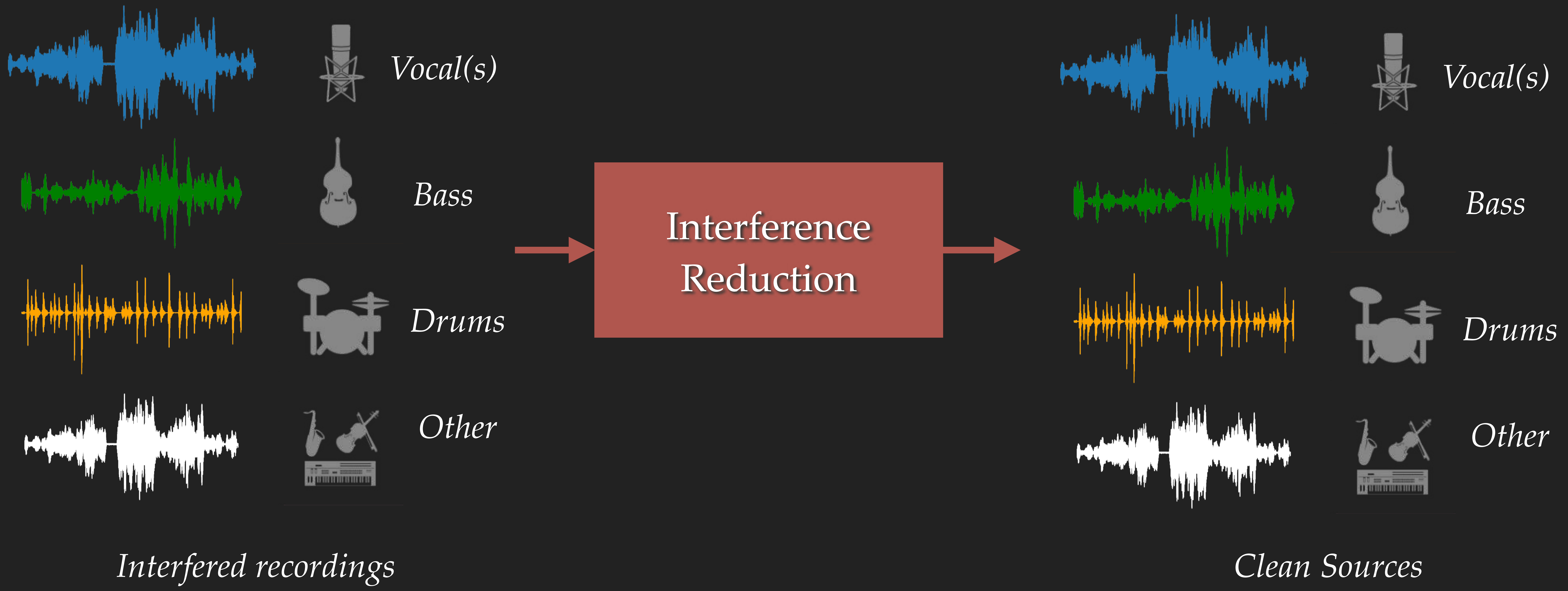
RECORDINGS FROM LIVE CONCERTS (CARNATIC CONCERT)

https://images.app.goo.gl/g9MPV2bNE5faJz4M7



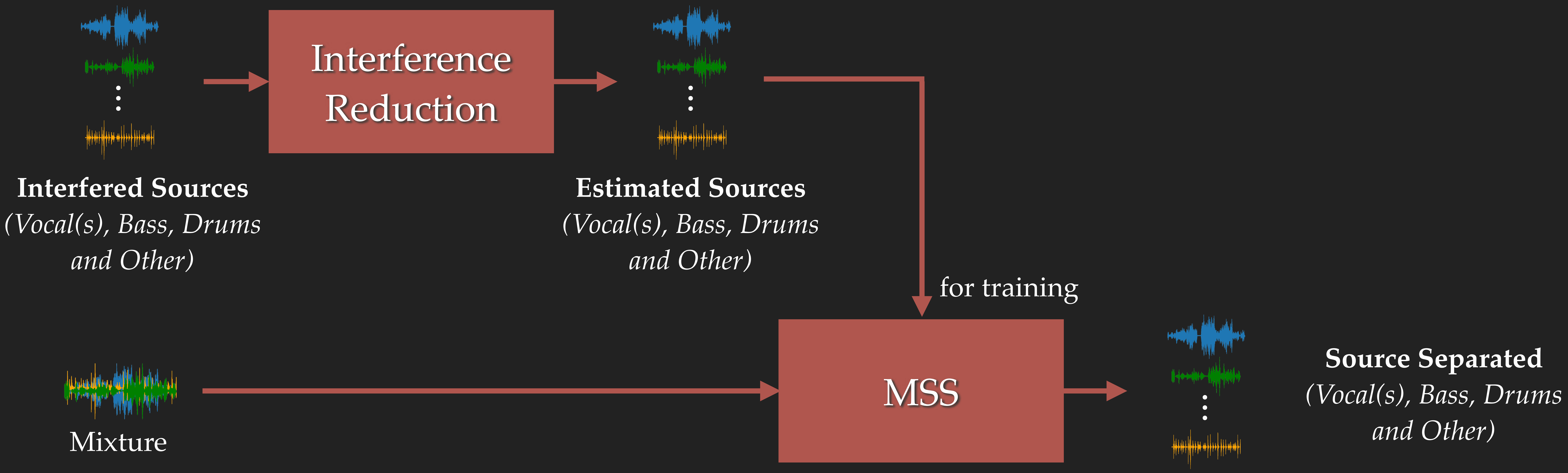
- Live recordings lacks acoustic shielding
- Microphone intended to pick specific source picks up the other sources as well

PRIMARY OBJECTIVE



Interference Reduction System

THE GOAL



Interference Reduction

Interfered Sources
(Vocal(s), Bass, Drums and Other)

Estimated Sources
(Vocal(s), Bass, Drums and Other)

for training

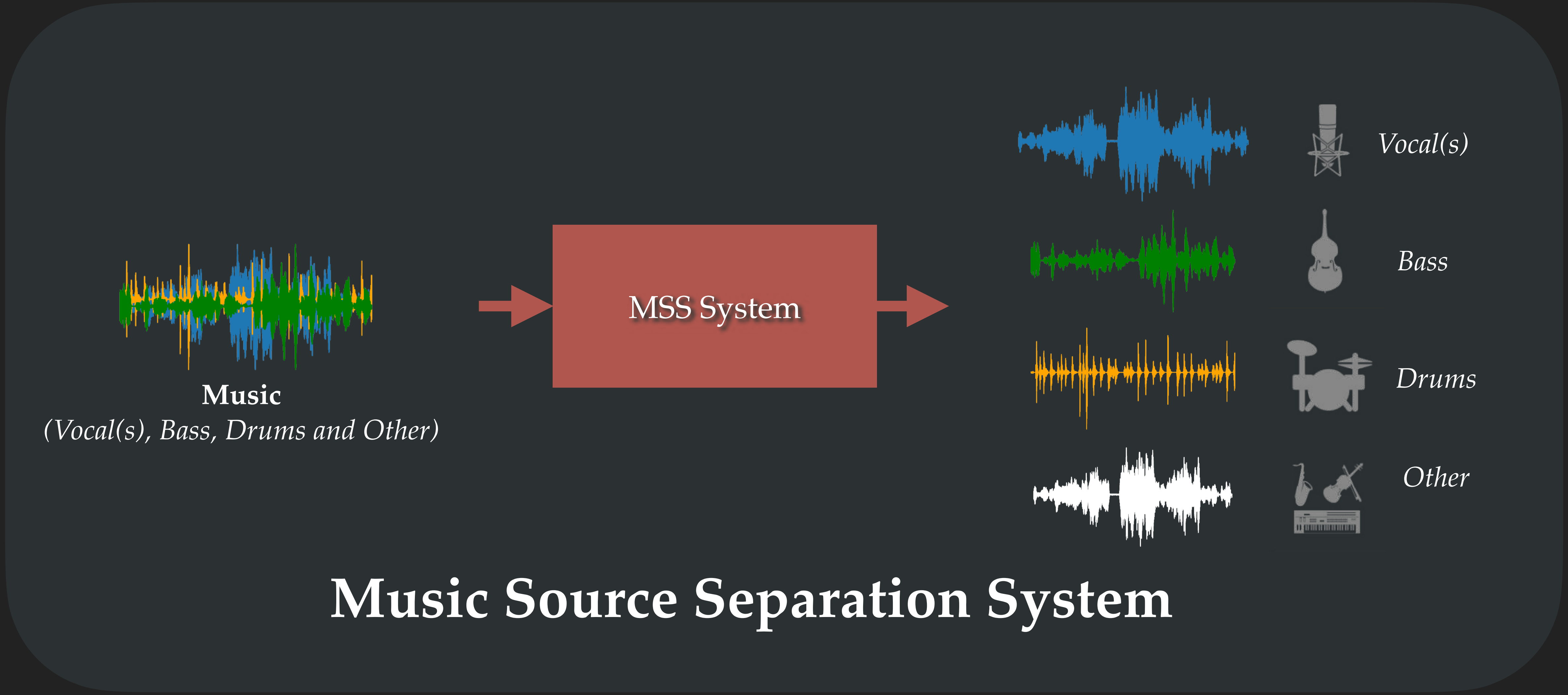
Mixture

MSS

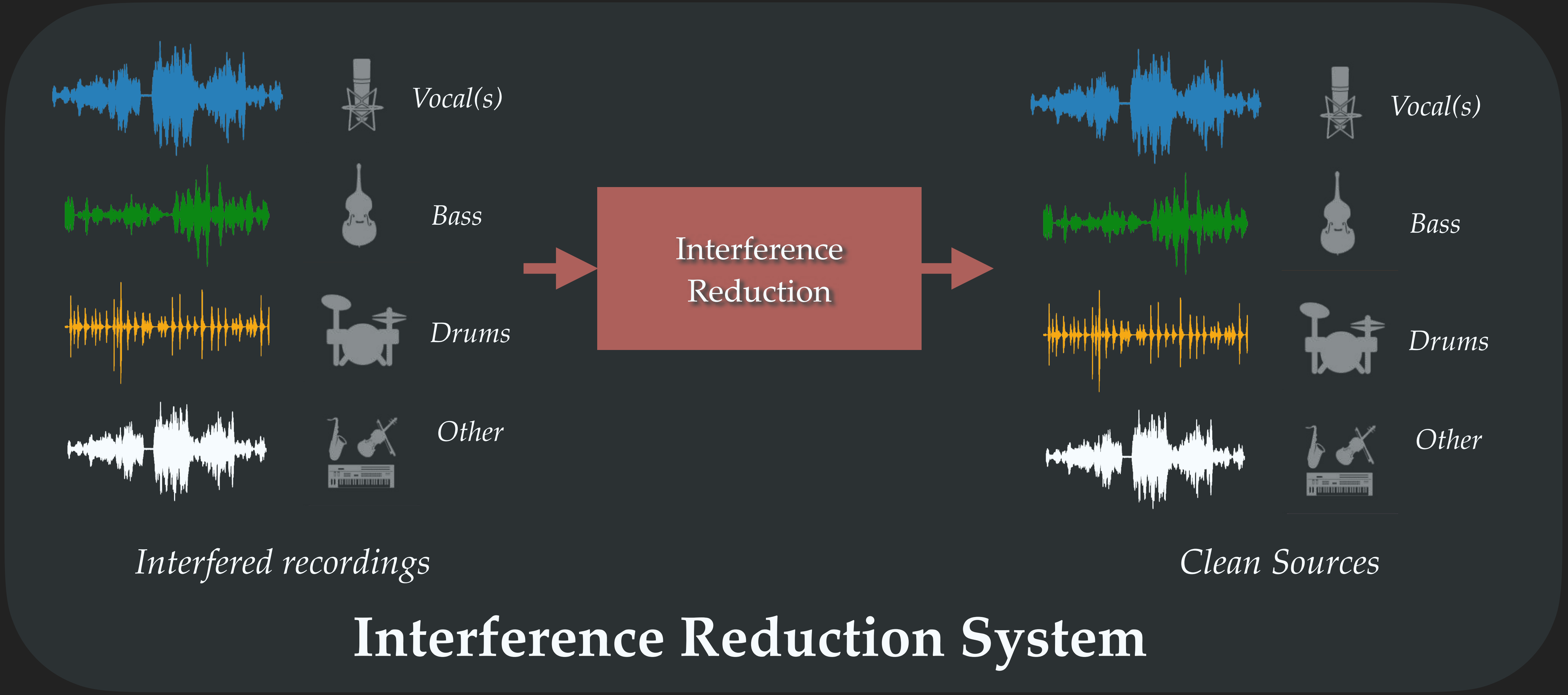
Source Separated
(Vocal(s), Bass, Drums and Other)

MSS VS INTERFERENCE REDUCTION

- Interference reduction: Special type of source separation
- Aim: Clean microphone recordings



Music Source Separation System



Interference Reduction System

INTERFERENCE REDUCTION

TRENDS SO FAR

TRENDS SO FAR

- No neural network-based techniques proposed, due to dataset?

TRENDS SO FAR

- No neural network-based techniques proposed, due to dataset?
- DSP Algorithms: **KAMIR**[1] (Kernel Additive Modelling for Interference Reduction) - the state-of-the-art [2015]

---

[1] *T. Pratzlich, R. M. Bittner, A. Liutkus, and M. Muller, "Kernel additive modeling for interference reduction in multi-channel music recordings," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 584–588*

TRENDS SO FAR

- No neural network-based techniques proposed, due to dataset?
- DSP Algorithms: **KAMIR**[1] (Kernel Additive Modelling for Interference Reduction) - the state-of-the-art [2015]
- MIRA (Multitrack Interference Reduction Algorithm) & FastMIRA[2] are the advancement of KAMIR

---

[1] *T. Pratzlich, R. M. Bittner, A. Liutkus, and M. Muller, "Kernel additive modeling for interference reduction in multi-channel music recordings," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 584–588*

[2] *Di Carlo, Diego, Antoine Liutkus, and Ken Déguemel. "Interference reduction on full-length live recordings." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018*

CONTRIBUTIONS

CONTRIBUTIONS

- Learning free Optimisation Algorithm

CONTRIBUTIONS

- Learning free Optimisation Algorithm
- Convolutional Autoencoders (CAEs)

CONTRIBUTIONS

- Learning free Optimisation Algorithm
- Convolutional Autoencoders (CAEs)
- Truncated UNet (t-UNet)

CONTRIBUTIONS

- Learning free Optimisation Algorithm
- Convolutional Autoencoders (CAEs)
- Truncated UNet (t-UNet)
- Dilated full Wave-U-Net (dfUNet) with Graph Attentions

ASSUMPTIONS

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

ASSUMPTIONS

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

- Each source has at least one dedicated microphones.

ASSUMPTIONS

https://images.app.goo.gl/g9MPV2bNE5faJz4M7

- Each source has at least one dedicated microphones.
- At least a single source is dominant in its dedicated microphone.

INTERFERENCE AS NOISE

Treating interference as a noise,

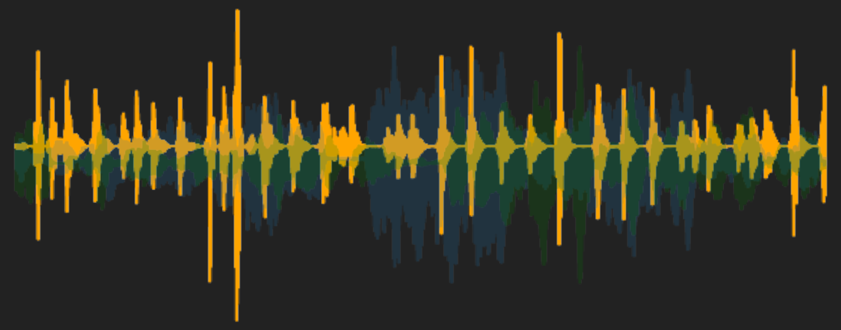
$$x(t) = s(t) + n(t)$$

INTERFERENCE AS NOISE

Treating interference as a noise,

$$x(t) = s(t) + n(t)$$

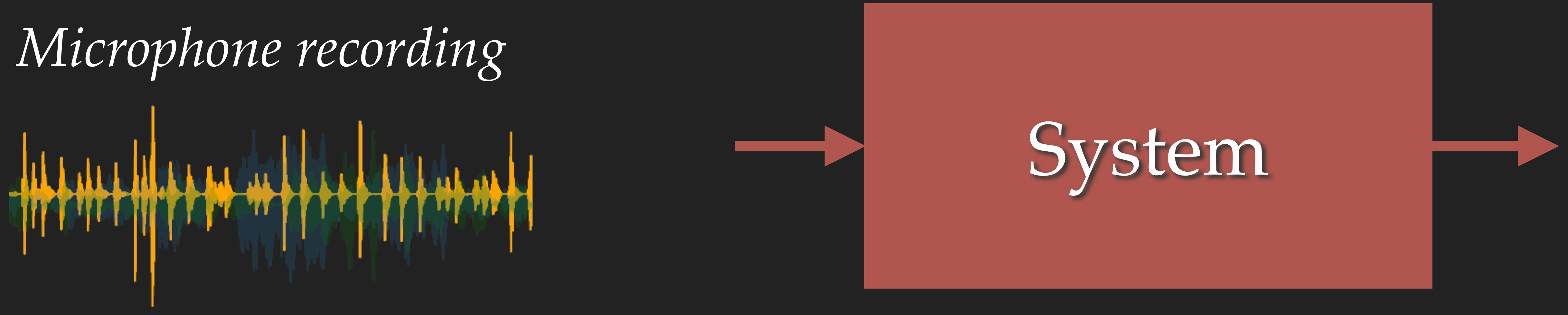
Microphone recording



INTERFERENCE AS NOISE

Treating interference as a noise,

$$x(t) = s(t) + n(t)$$



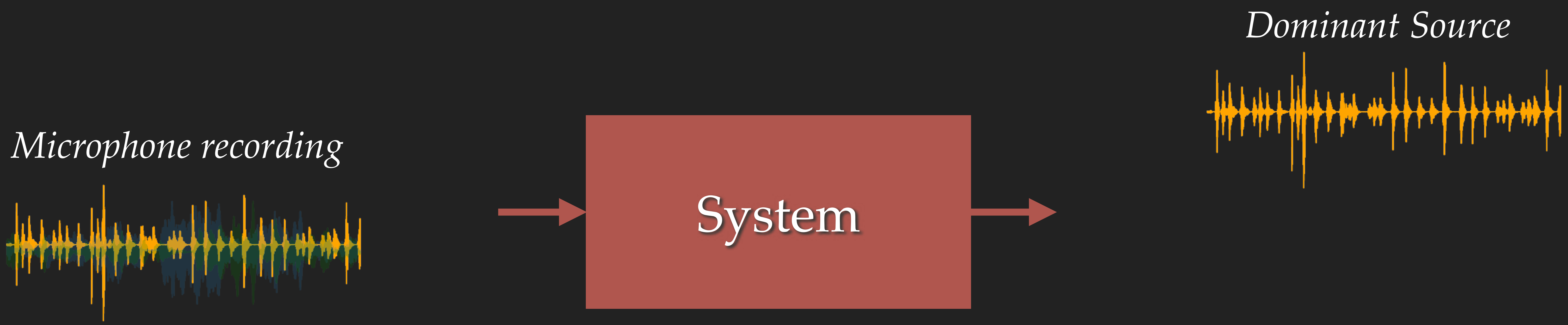
Microphone recording

System

INTERFERENCE AS NOISE

Treating interference as a noise,

$$x(t) = s(t) + n(t)$$



Dominant Source

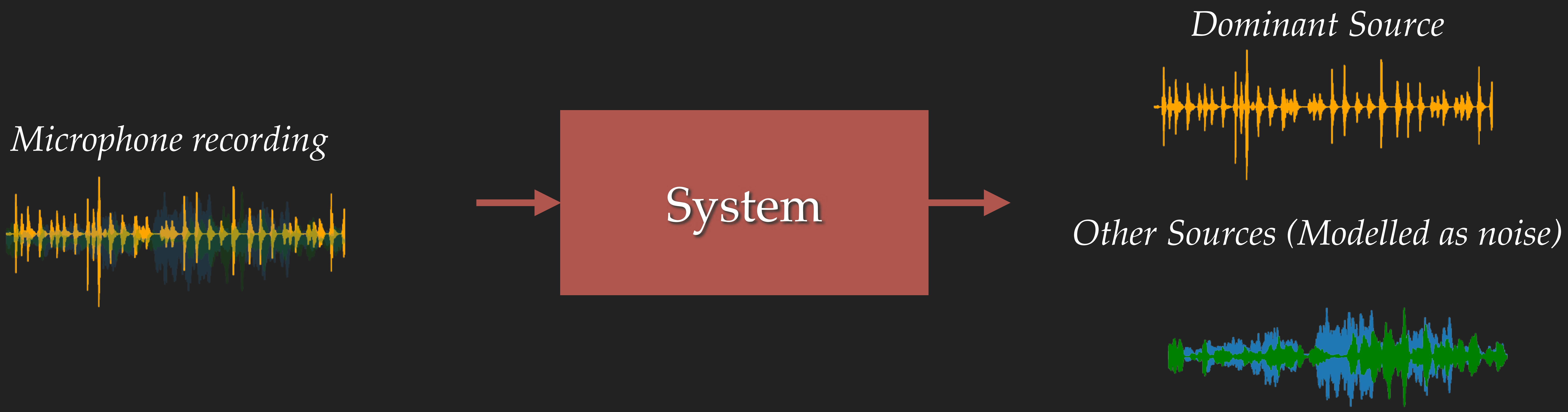
Microphone recording

System

INTERFERENCE AS NOISE

Treating interference as a noise,

$$x(t) = s(t) + n(t)$$



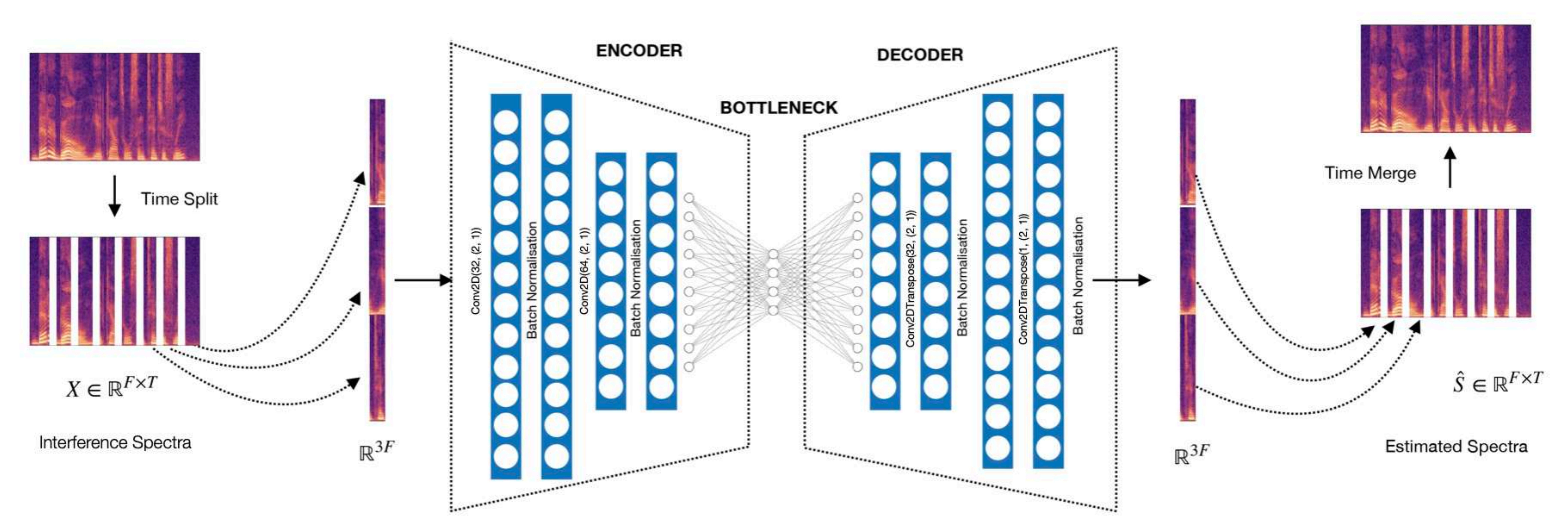
Microphone recording

System

Dominant Source

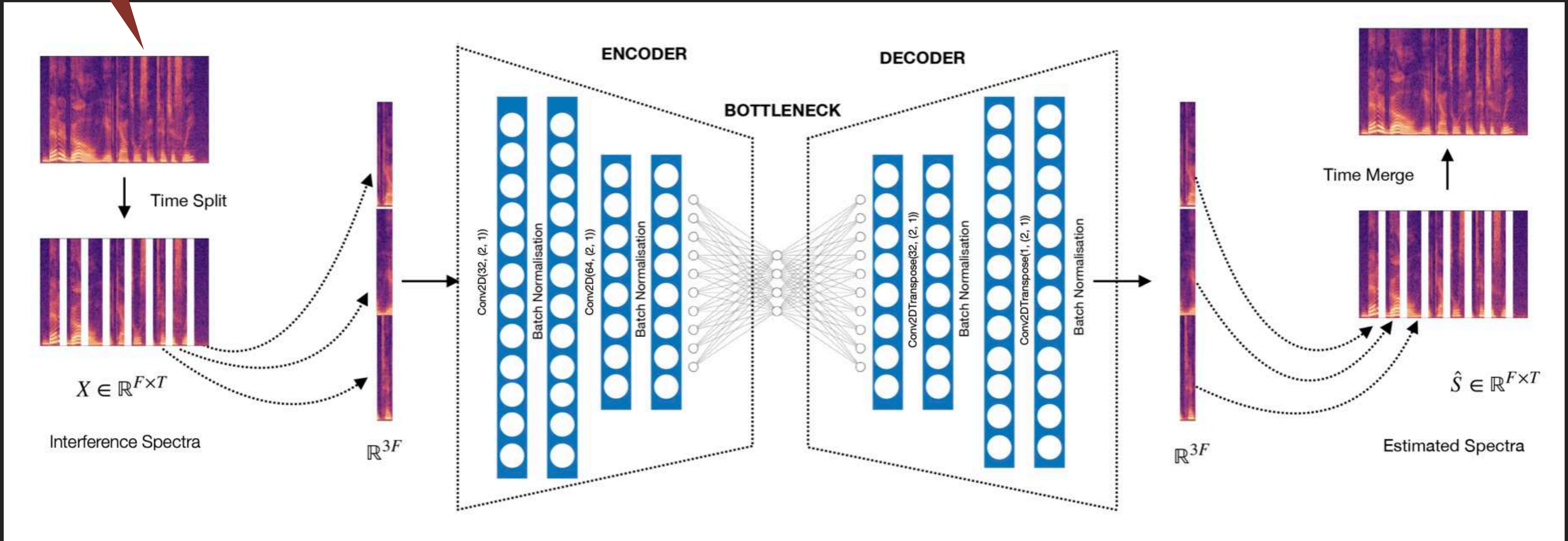
Other Sources (Modelled as noise)

THE CONVOLUTIONAL AUTOENCODER (CAE)



THE CONVOLUTIONAL AUTOENCODER (CAE)

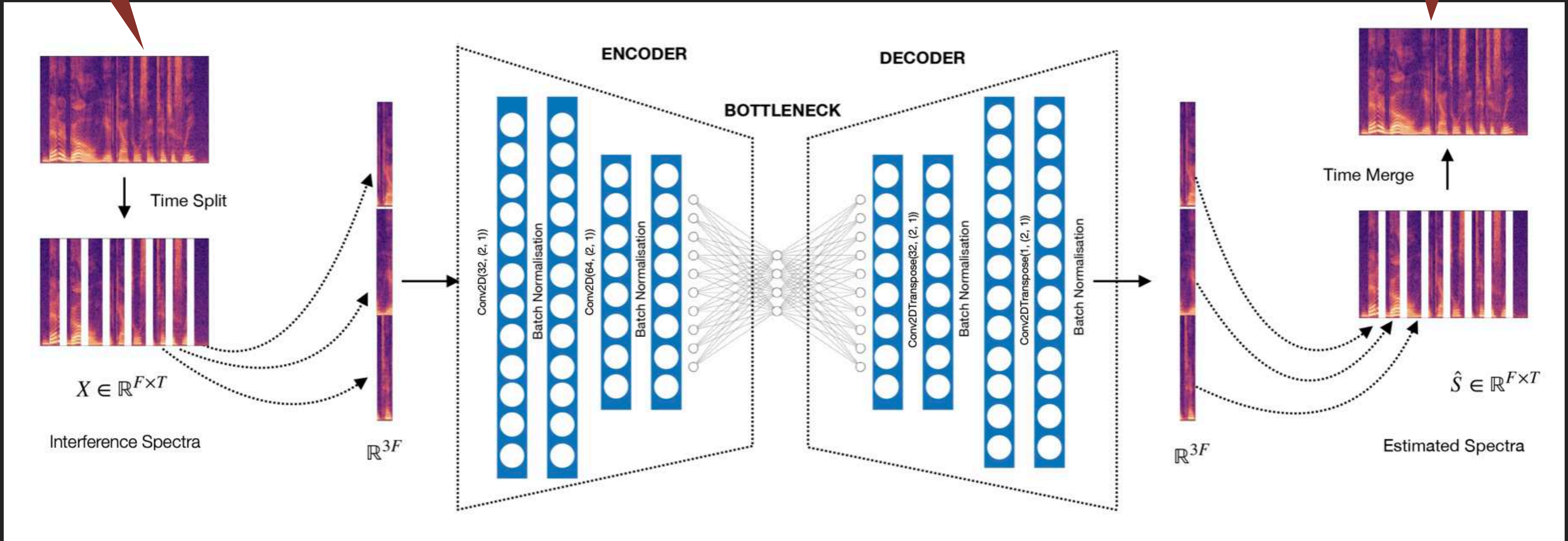
Microphone Recordings



THE CONVOLUTIONAL AUTOENCODER (CAE)

Microphone Recordings

Estimated Sources



SHORTCOMINGS OF THE APPROACH

- Poor generalisability
- Thus, for each source there should be dedicated trained CAEs
- Phase information issues

HIDDEN INFORMATION

Interfered Vocal(s)

Dominant Vocal(s)

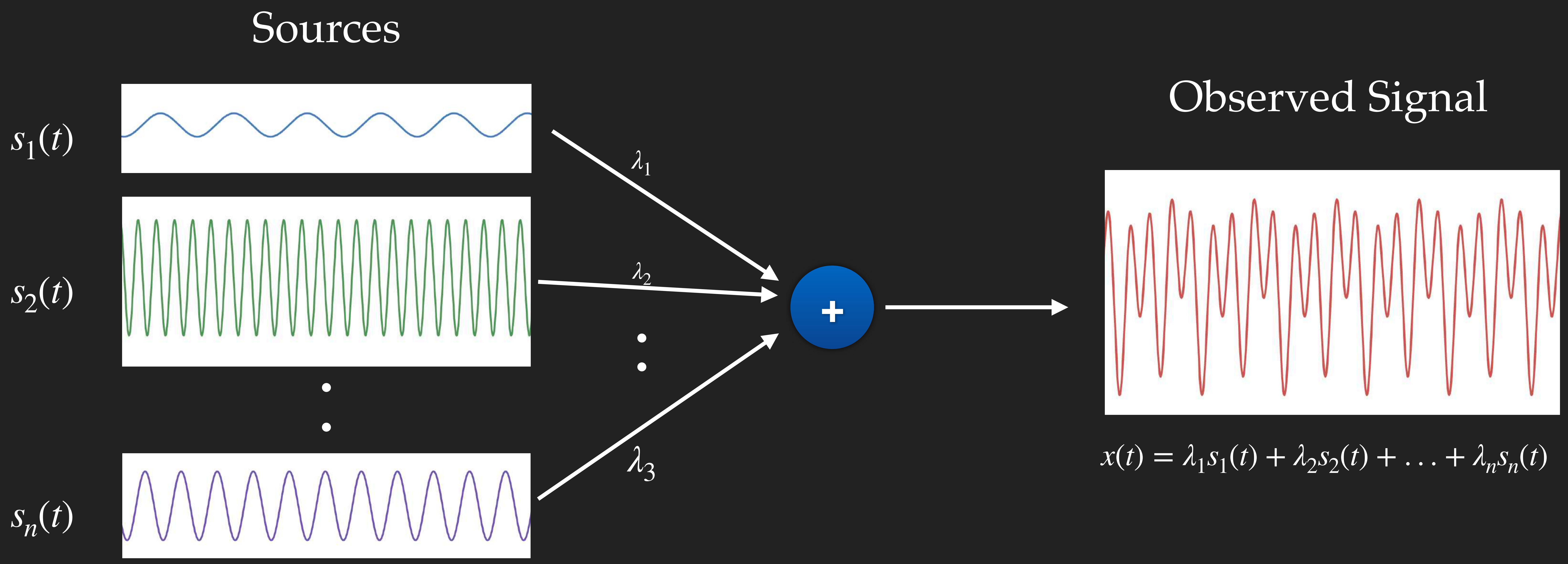
bass in background

drums in background

Other in background

We
have this
information !!!

MATHEMATICAL FORMULATION



Sources

$s_1(t)$

$s_2(t)$

$s_n(t)$

λ_1

λ_2

λ_3

Observed Signal

$$x(t) = \lambda_1 s_1(t) + \lambda_2 s_2(t) + \ldots + \lambda_n s_n(t)$$

MATHEMATICAL FORMULATION

MATHEMATICAL FORMULATION

For k microphones and n sources,

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

$$X = \Lambda S$$

$$X = [x_1(t), x_2(t), \ldots, x_k(t)]^T$$

$$S = [s_1(t), s_2(t), \ldots, s_n(t)]^T$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

$$X = \Lambda S$$

Microphone
Recordings

$$X = [x_1(t), x_2(t), \ldots, x_k(t)]^T$$

$$S = [s_1(t), s_2(t), \ldots, s_n(t)]^T$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

$$X = \Lambda S$$

Microphone Recordings

Mixing Matrix

$$X = [x_1(t), x_2(t), \ldots, x_k(t)]^T$$

$$S = [s_1(t), s_2(t), \ldots, s_n(t)]^T$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

$$X = \Lambda S$$

Microphone Recordings

Mixing Matrix

Source Signals

$$X = [x_1(t), x_2(t), \ldots, x_k(t)]^T$$

$$S = [s_1(t), s_2(t), \ldots, s_n(t)]^T$$

MATHEMATICAL FORMULATION

For k microphones and n sources,

$$x_1(t) = \lambda_{11}s_1(t) + \lambda_{12}s_2(t) + .. + \lambda_{1n}s_n(t)$$

$$x_2(t) = \lambda_{21}s_1(t) + \lambda_{22}s_2(t) + .. + \lambda_{2n}s_n(t)$$

$$\vdots$$

$$x_k(t) = \lambda_{k1}s_1(t) + \lambda_{k2}s_2(t) + .. + \lambda_{kn}s_n(t)$$

$$\underset{\substack{\nearrow \\ \text{Microphone} \\ \text{Recordings}}}{X} = \underset{\substack{\uparrow \\ \text{Mixing} \\ \text{Matrix}}}{\Lambda}\underset{\substack{\nwarrow \\ \text{Source} \\ \text{Signals}}}{S}$$

$$X = [x_1(t), x_2(t), \ldots, x_k(t)]^T$$

$$S = [s_1(t), s_2(t), \ldots, s_n(t)]^T$$

Similarly for mixture signal,

$$m(t) = \sum_{i=0}^{n} \beta_i s_i(t) = b^T S$$

ISSUES: WHAT'S NEXT?

Equations: $X = \Lambda S$ and $m = b^T S$

ISSUES: WHAT'S NEXT?

Equations: $X = \Lambda S$ and $m = b^T S$

- $X = \Lambda S$ is an over-determined or over-constrained problem

ISSUES: WHAT'S NEXT?

Equations: $X = \Lambda S$ and $m = b^T S$

- $X = \Lambda S$ is an over-determined or over-constrained problem
- No unique solution, multiple solution exists

LEARNING FREE OPTIMISATION ALGORITHM

Equations: $X = \Lambda S$ and $m = b^T S$

LEARNING FREE OPTIMISATION ALGORITHM

Equations: $X = \Lambda S$ and $m = b^T S$

With guidance of
Dr. Siddhartha Sarma

LEARNING FREE OPTIMISATION ALGORITHM

Equations: $X = \Lambda S$ and $m = b^T S$

Problem statement: *minimise* $\|X - \Lambda S\|^2 + \|m - b^T S\|^2$ *with respect to* Λ, S *and* b *subject to constraints:*

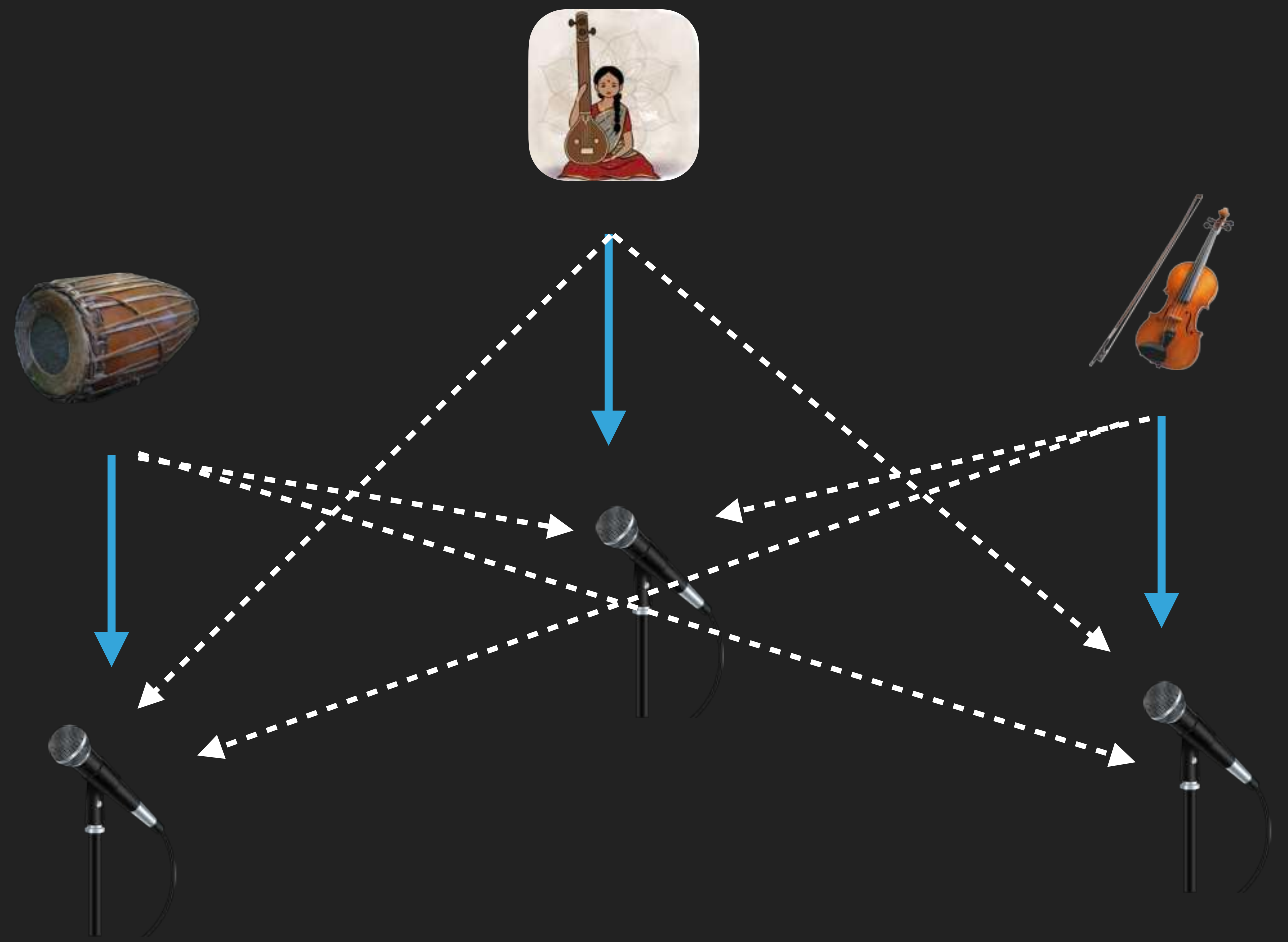
1. $\Lambda \neq I$
2. $\lambda_{ii} > \lambda_{ij}$
3. $\gamma_1 \leq \lambda_{ij} \leq \gamma_2, \forall i \neq j$

LEARNING FREE OPTIMISATION ALGORITHM

Equations: $X = \Lambda S$ and $m = b^T S$

Problem statement: *minimise* $\|X - \Lambda S\|^2 + \|m - b^T S\|^2$ *with respect to* Λ, S *and* b *subject to constraints:*

1. $\Lambda \neq I$
2. $\lambda_{ii} > \lambda_{ij}$
3. $\gamma_1 \leq \lambda_{ij} \leq \gamma_2, \forall i \neq j$



ALTERNATE MINIMISATION SOLUTION

- Non convex problem, global minima does not exist
- Alternate minimisation approach
- Derived the update rule for Λ, S *and* b.

ALTERNATE MINIMISATION SOLUTION

- Non convex problem, global minima does not exist
- Alternate minimisation approach
- Derived the update rule for Λ, S *and* b.

Update Rules:

$$\Lambda = (XSS^T)(SS^T + \eta I)^{-1}$$

$$S = (\Lambda^T \Lambda + bb^T)^{-1}(bm + \Lambda^T X)$$

$$b = (SS^T + \eta I)^{-1}(Sm^T)$$

ALTERNATE MINIMISATION SOLUTION

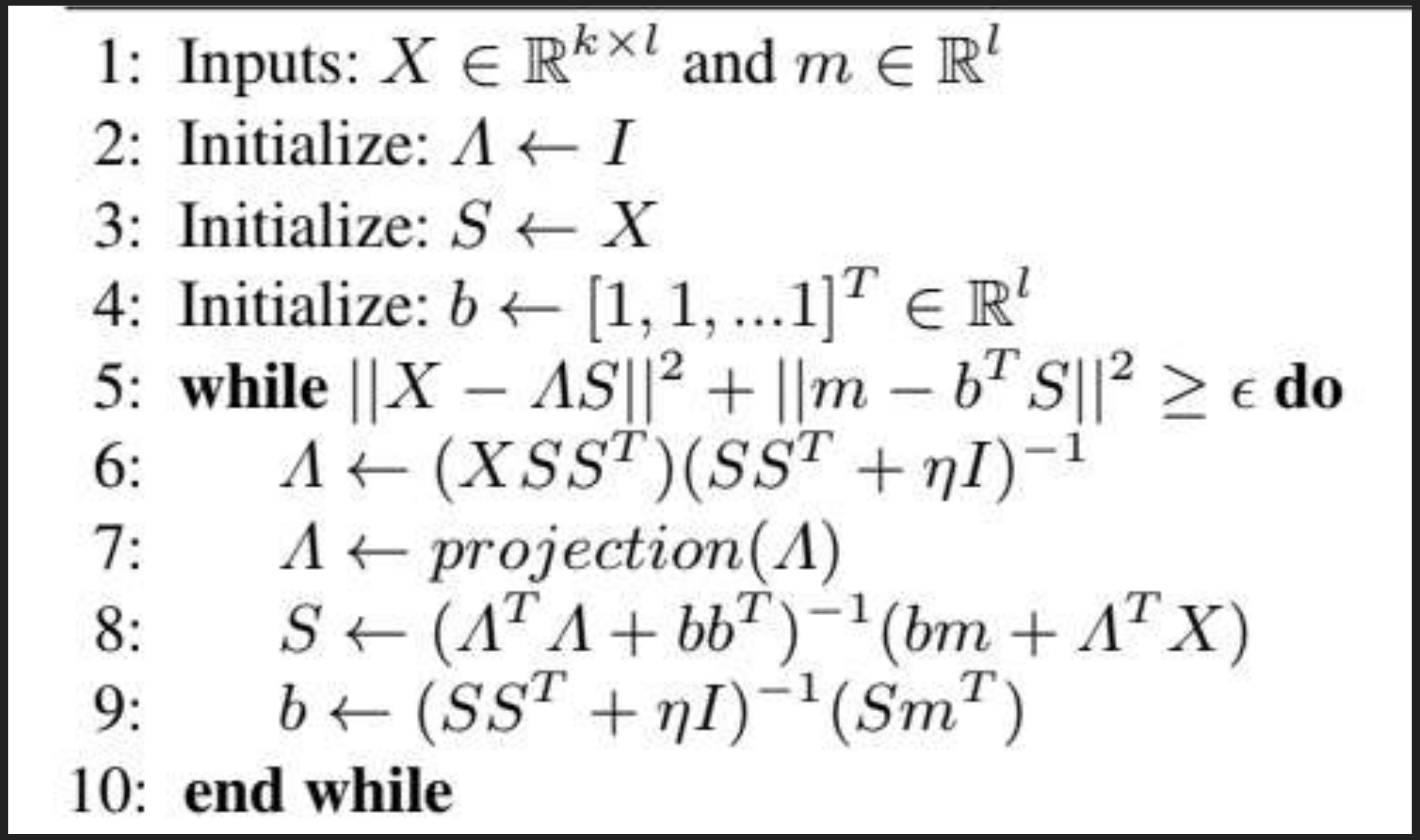
- Non convex problem, global minima does not exist
- Alternate minimisation approach
- Derived the update rule for Λ, S *and* b.

Update Rules:

$$\Lambda = (XSS^T)(SS^T + \eta I)^{-1}$$

$$S = (\Lambda^T \Lambda + bb^T)^{-1}(bm + \Lambda^T X)$$

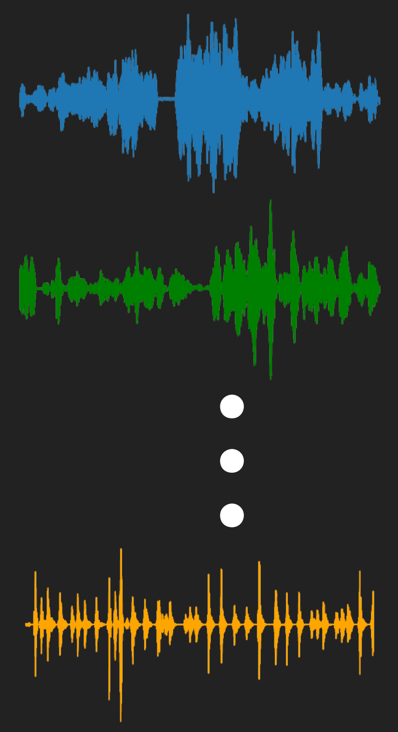
$$b = (SS^T + \eta I)^{-1}(Sm^T)$$



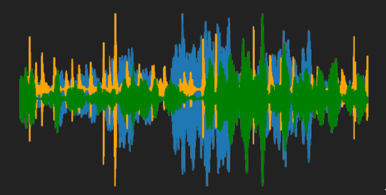
```
Inputs: $X \in \mathbb{R}^{k \times l}$ and $m \in \mathbb{R}^l$
Initialize: $\Lambda \leftarrow I$
Initialize: $S \leftarrow X$
Initialize: $b \leftarrow [1, 1, ...1]^T \in \mathbb{R}^l$
while $||X - \Lambda S||^2 + ||m - b^T S||^2 \geq \epsilon$ do
    $\Lambda \leftarrow (XSS^T)(SS^T + \eta I)^{-1}$
    $\Lambda \leftarrow projection(\Lambda)$
    $S \leftarrow (\Lambda^T \Lambda + bb^T)^{-1}(bm + \Lambda^T X)$
    $b \leftarrow (SS^T + \eta I)^{-1}(Sm^T)$
end while
```

OVERALL PROCEDURE

OVERALL PROCEDURE

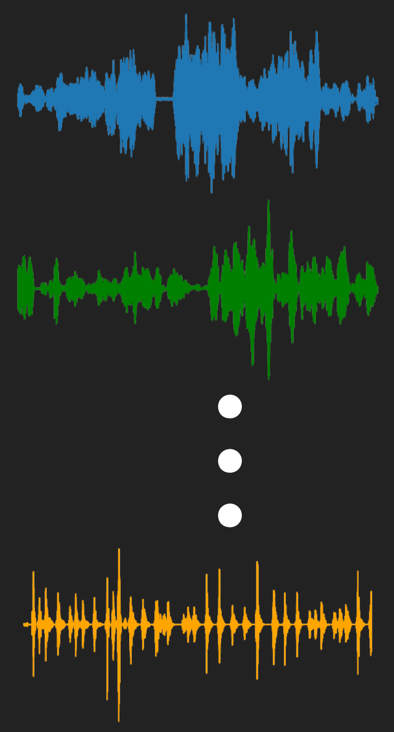


$x(t) \in \mathbb{R}^l$



$m(t) \in \mathbb{R}^l$

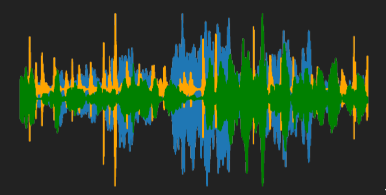
OVERALL PROCEDURE



$x(t) \in \mathbb{R}^l$

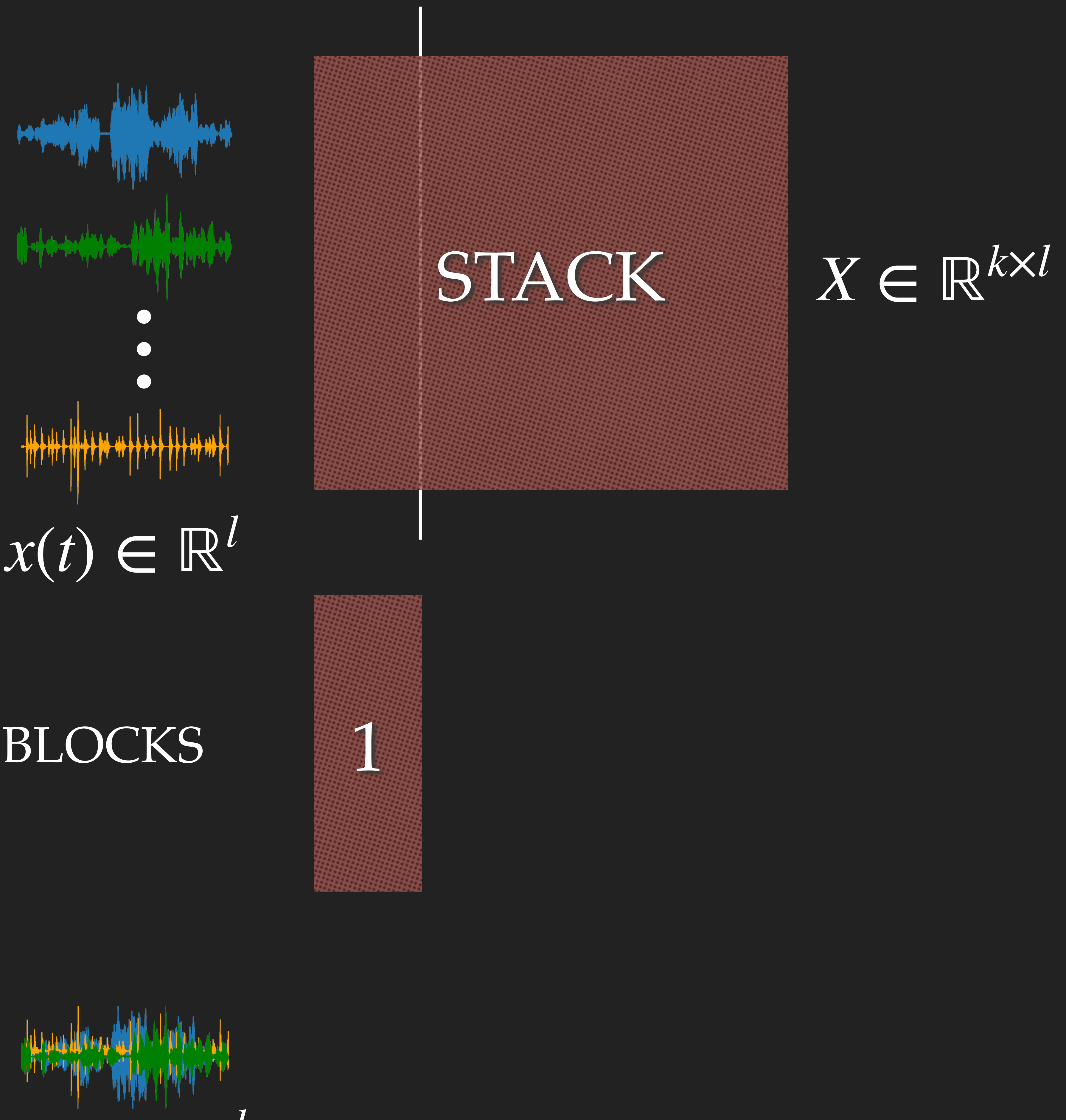
STACK

$X \in \mathbb{R}^{k \times l}$

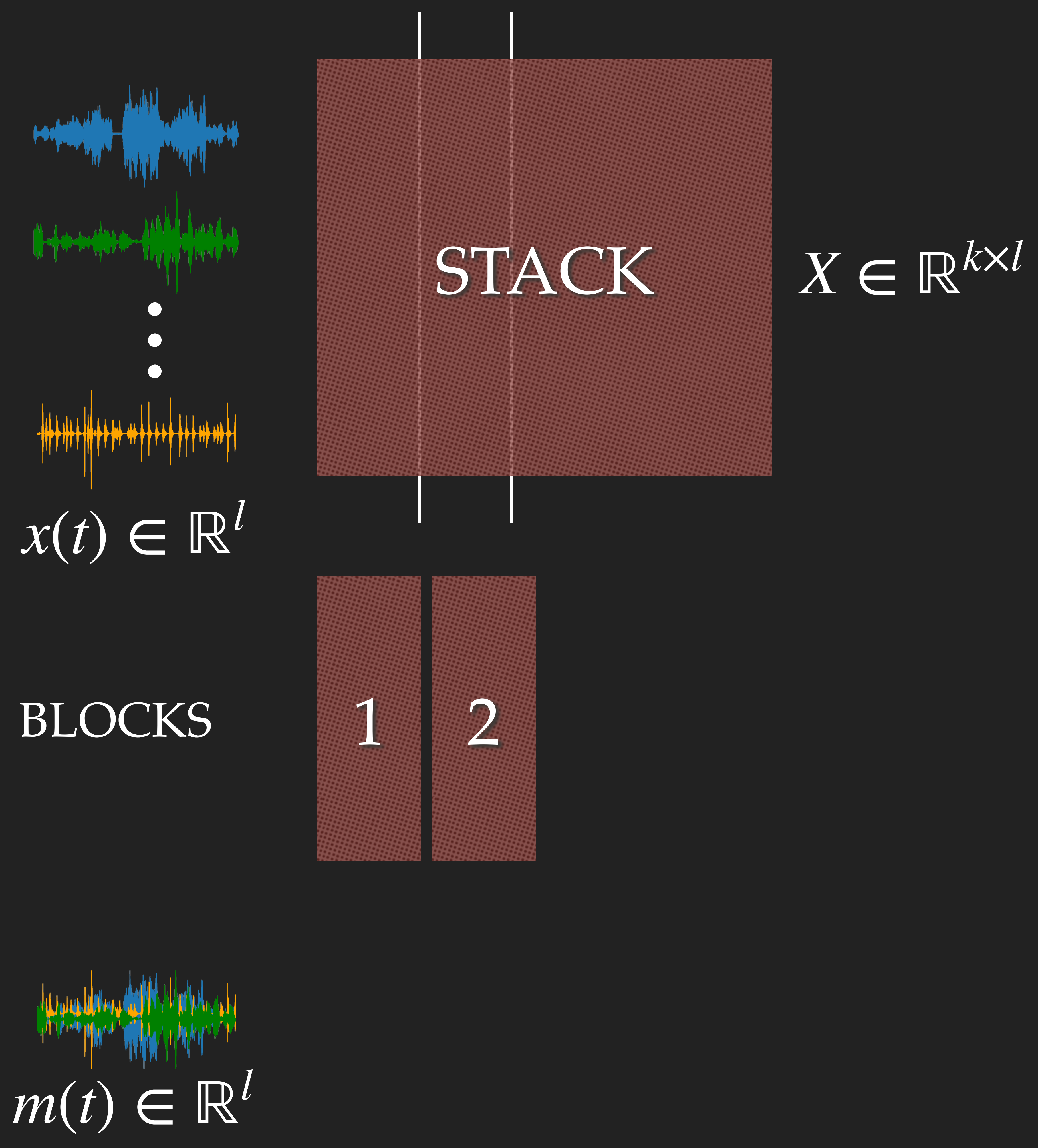


$m(t) \in \mathbb{R}^l$

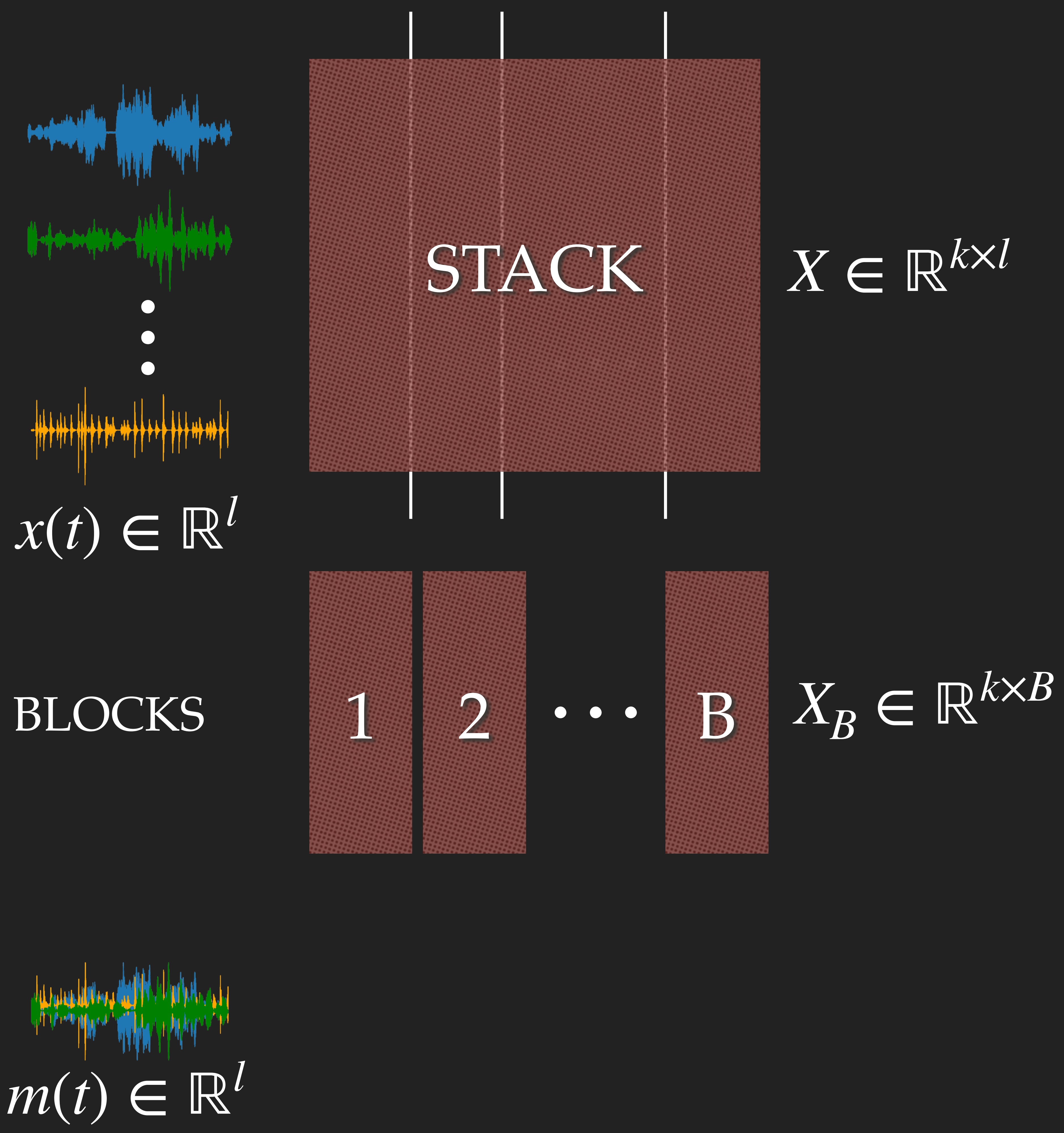
OVERALL PROCEDURE



OVERALL PROCEDURE



OVERALL PROCEDURE



STACK

$X \in \mathbb{R}^{k \times l}$

$x(t) \in \mathbb{R}^{l}$

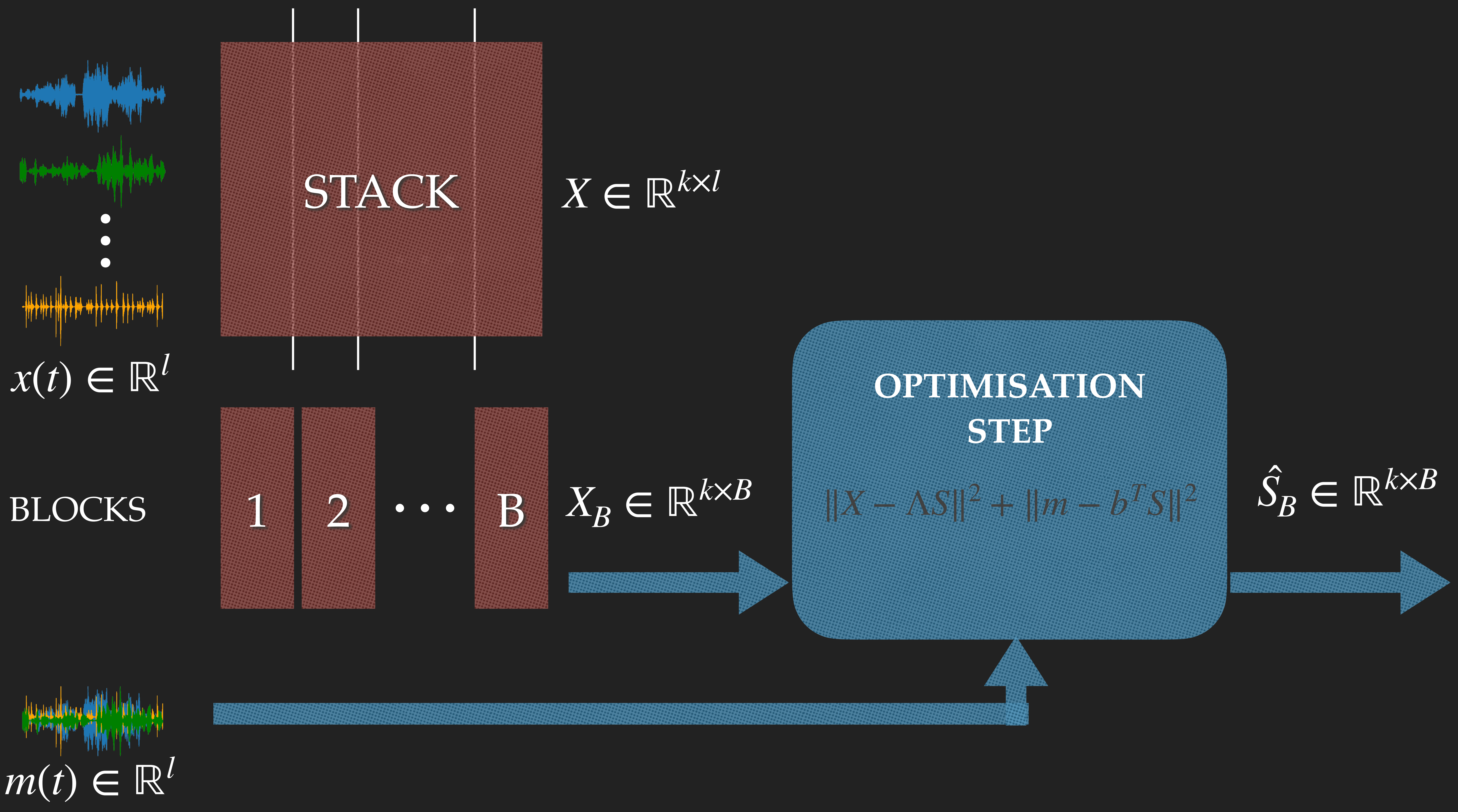
BLOCKS

1 2 $\cdots$ B

$X_B \in \mathbb{R}^{k \times B}$

$m(t) \in \mathbb{R}^{l}$

OVERALL PROCEDURE



STACK

$X \in \mathbb{R}^{k \times l}$

$x(t) \in \mathbb{R}^{l}$

BLOCKS

1 2 $\cdots$ B

$X_B \in \mathbb{R}^{k \times B}$

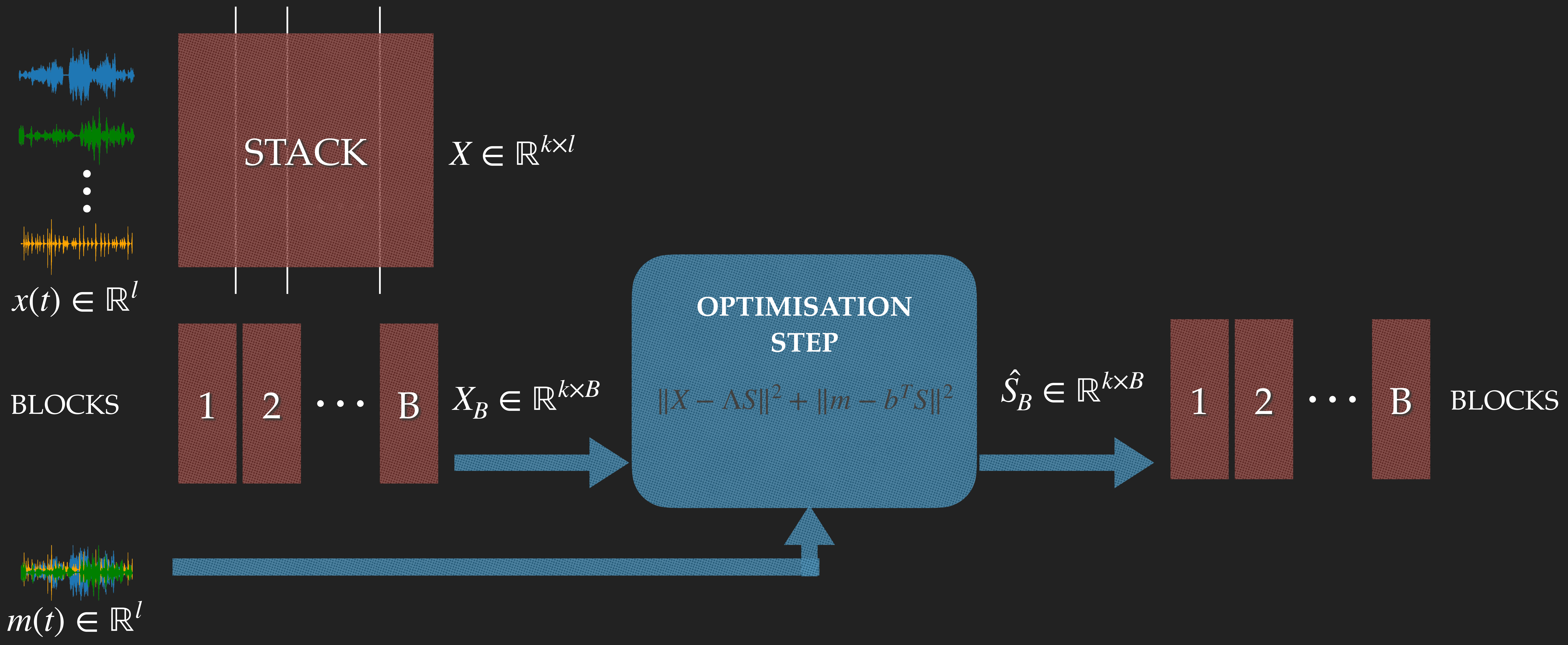
OPTIMISATION STEP

$\|X - AS\|^2 + \|m - b^T S\|^2$

$\hat{S}_B \in \mathbb{R}^{k \times B}$

$m(t) \in \mathbb{R}^{l}$

OVERALL PROCEDURE



STACK

$X \in \mathbb{R}^{k \times l}$

$x(t) \in \mathbb{R}^l$

BLOCKS

1 2 $\cdots$ B

$X_B \in \mathbb{R}^{k \times B}$

OPTIMISATION STEP

$$\|X - \Lambda S\|^2 + \|m - b^T S\|^2$$

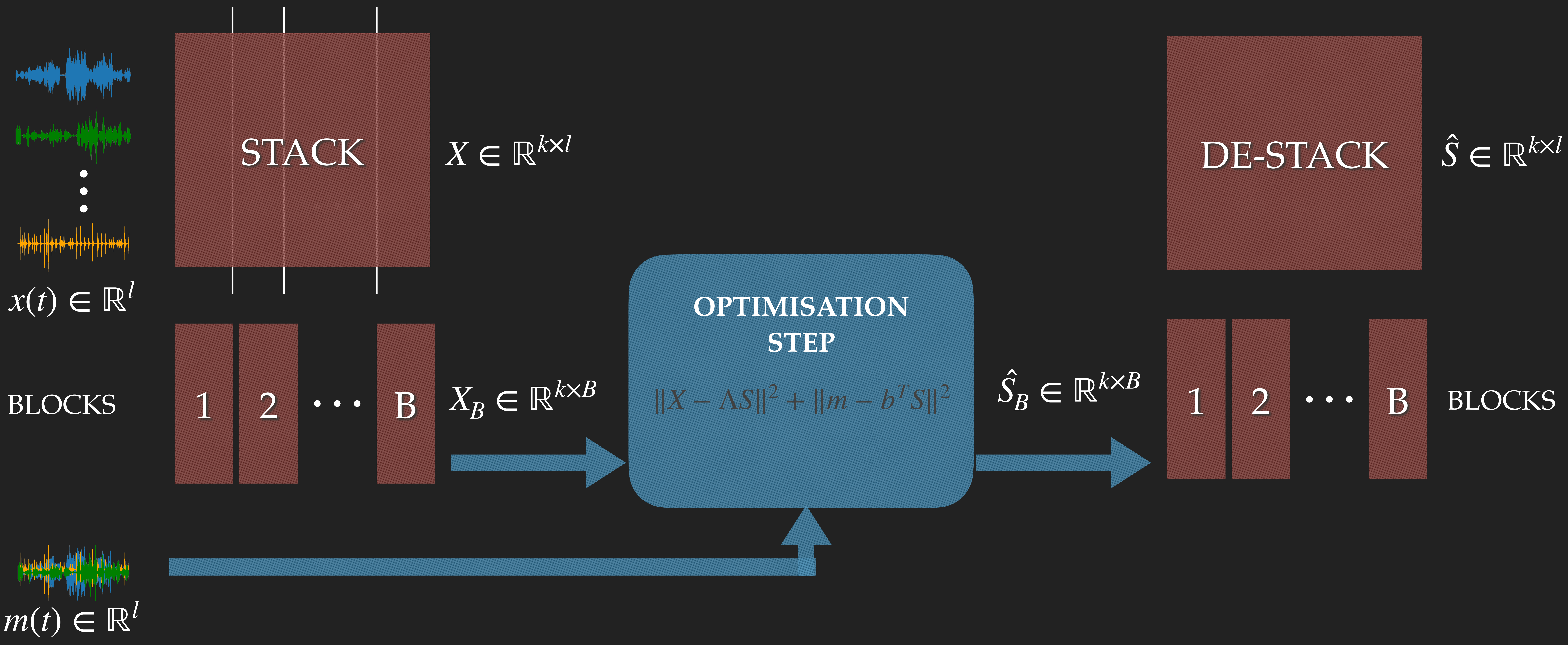
$\hat{S}_B \in \mathbb{R}^{k \times B}$

1 2 $\cdots$ B

BLOCKS

$m(t) \in \mathbb{R}^l$

OVERALL PROCEDURE



STACK

$X \in \mathbb{R}^{k \times l}$

DE-STACK

$\hat{S} \in \mathbb{R}^{k \times l}$

$x(t) \in \mathbb{R}^{l}$

BLOCKS

1 2 $\cdots$ B

$X_B \in \mathbb{R}^{k \times B}$

OPTIMISATION STEP

$\|X - AS\|^2 + \|m - b^T S\|^2$

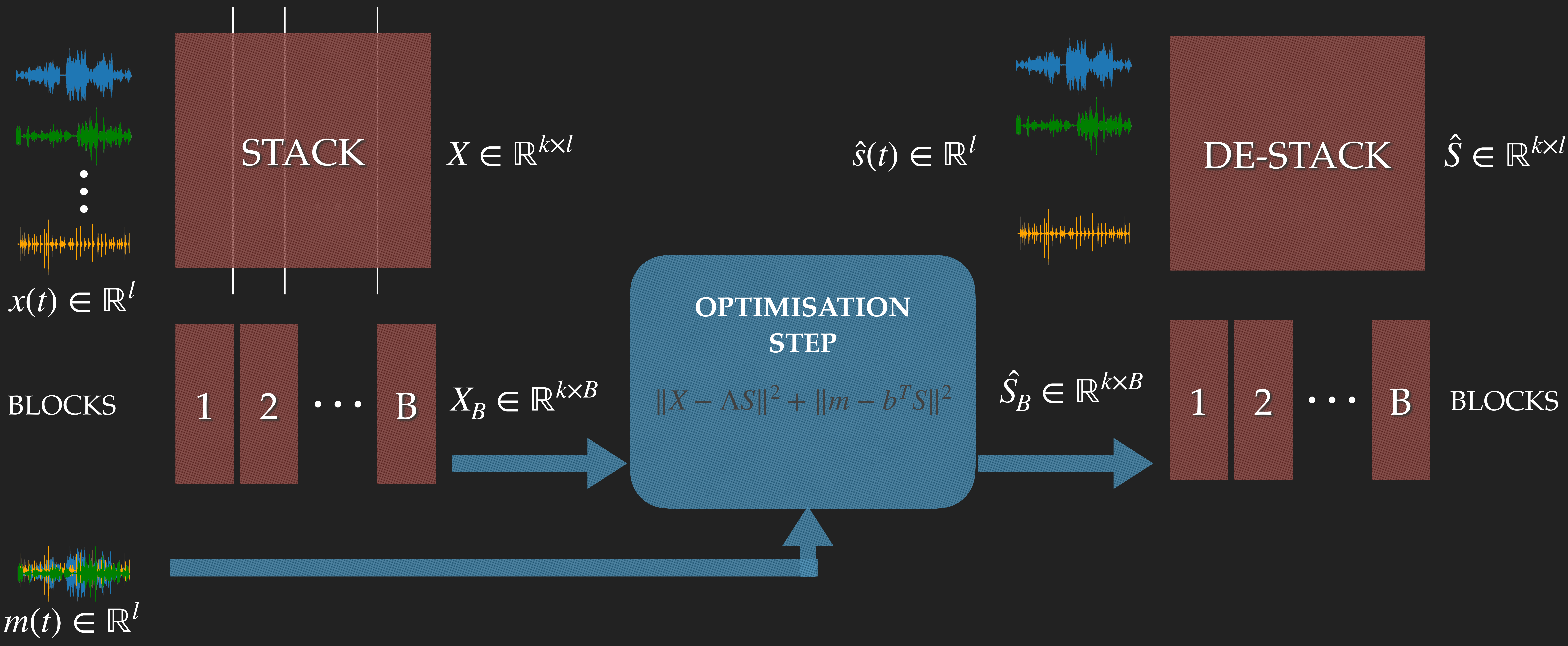
$\hat{S}_B \in \mathbb{R}^{k \times B}$

1 2 $\cdots$ B

BLOCKS

$m(t) \in \mathbb{R}^{l}$

OVERALL PROCEDURE



STACK

$X \in \mathbb{R}^{k \times l}$

$x(t) \in \mathbb{R}^l$

BLOCKS

1 2 $\cdots$ B

$X_B \in \mathbb{R}^{k \times B}$

$m(t) \in \mathbb{R}^l$

OPTIMISATION STEP

$\|X - AS\|^2 + \|m - b^T S\|^2$

$\hat{s}(t) \in \mathbb{R}^l$

DE-STACK

$\hat{S} \in \mathbb{R}^{k \times l}$

$\hat{S}_B \in \mathbb{R}^{k \times B}$

1 2 $\cdots$ B

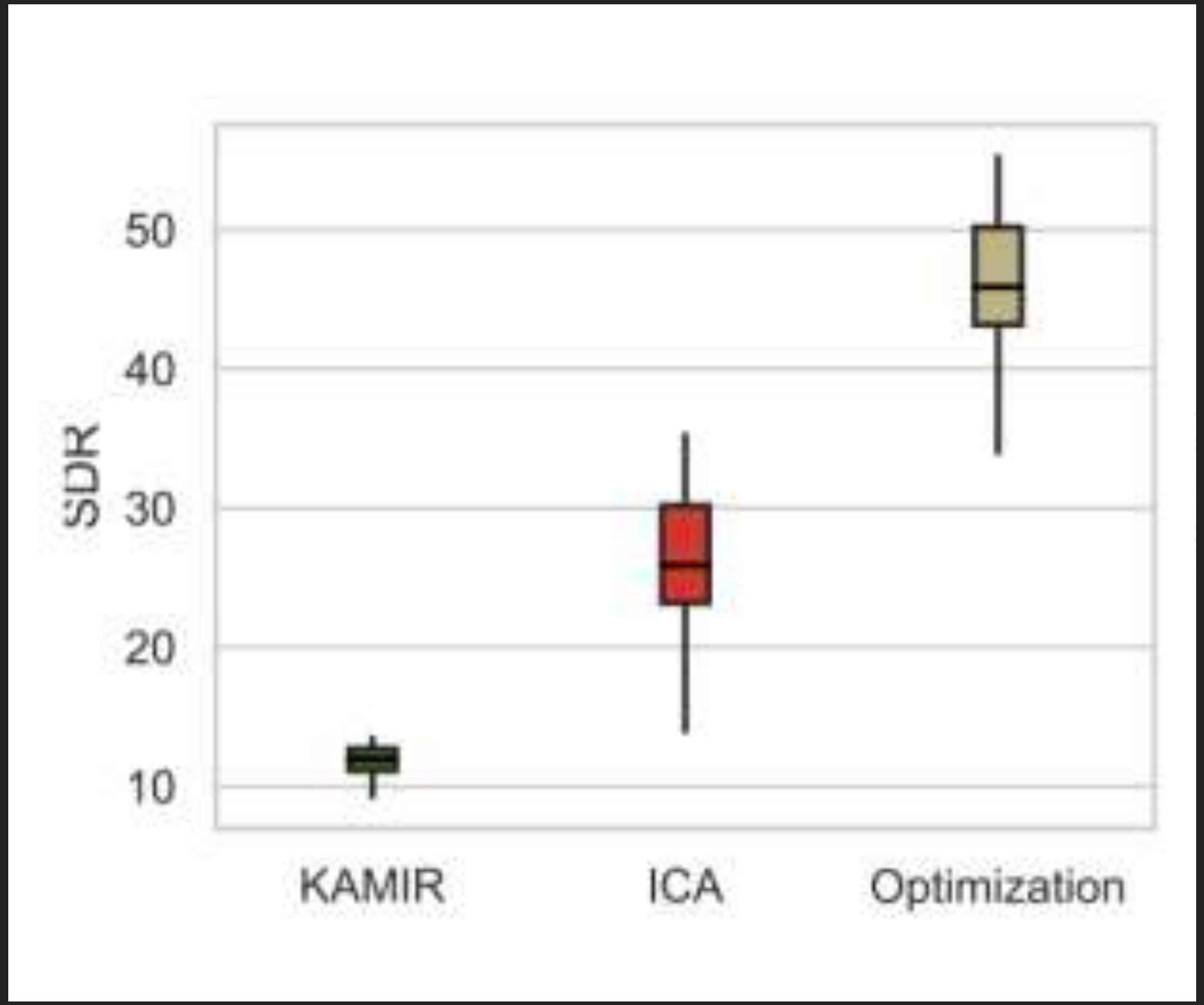
BLOCKS

DATASET 1: FOR TESTING (LINEAR MIXTURES - LM)

- Linear mixtures as per $X = \Lambda S$
- MUSDB18HQ training set: Artificially bleeded with randomly generated Λ
- Diagonals of Λ are in range 0.6 to 1
- Off diagonals of Λ are in range 0 to 0.4

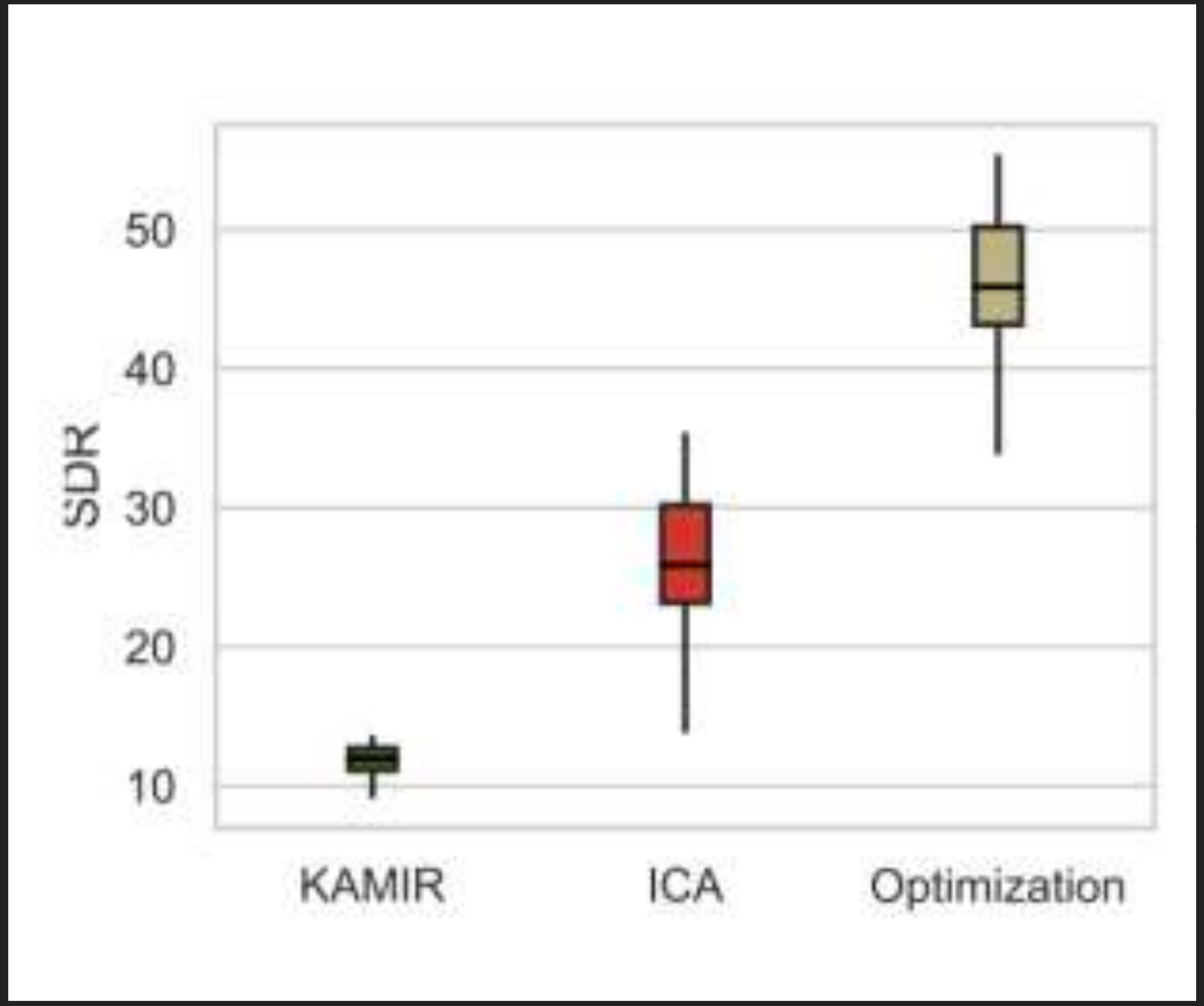
RESULTS

RESULTS

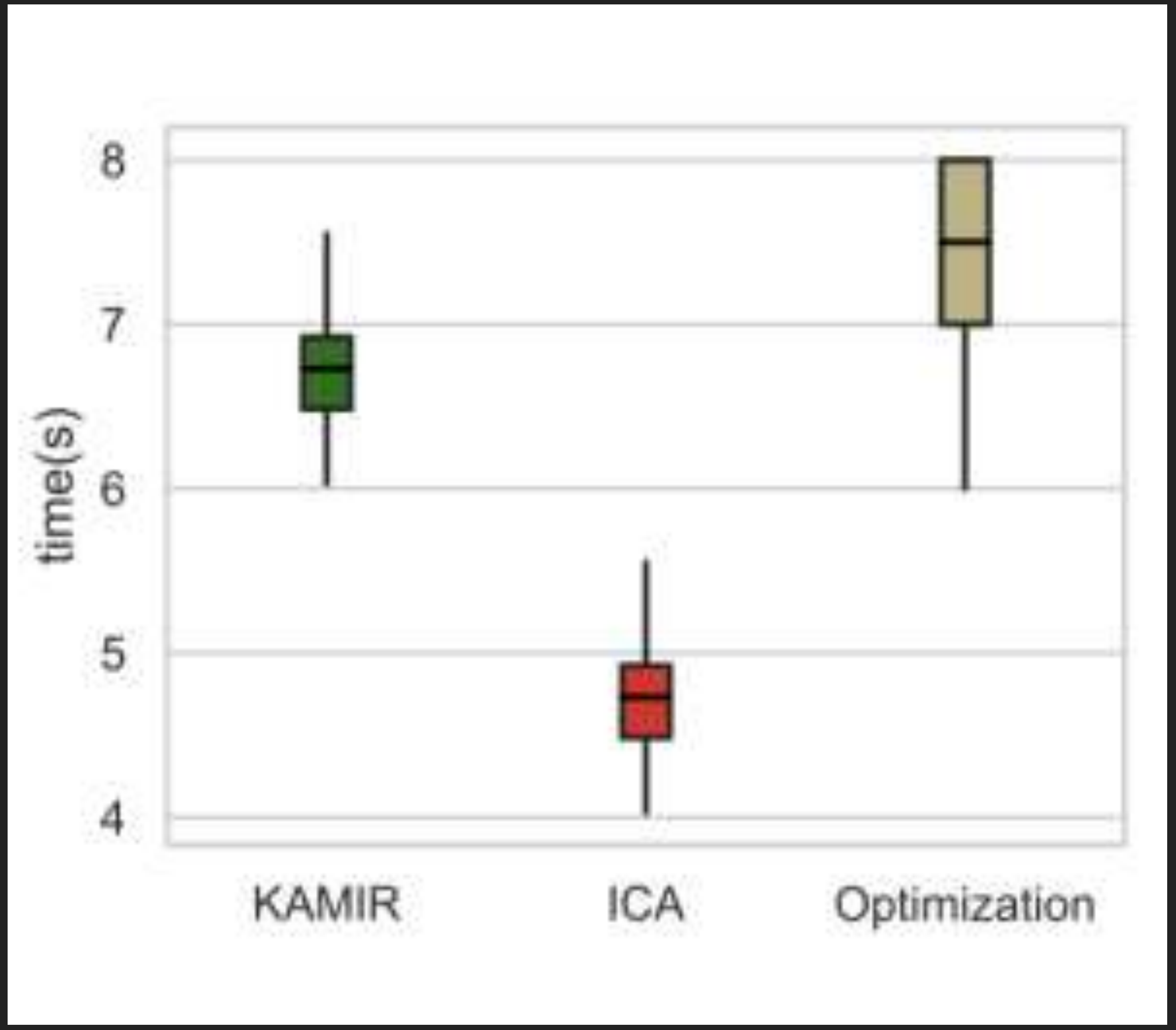


Average SDR across sources

RESULTS



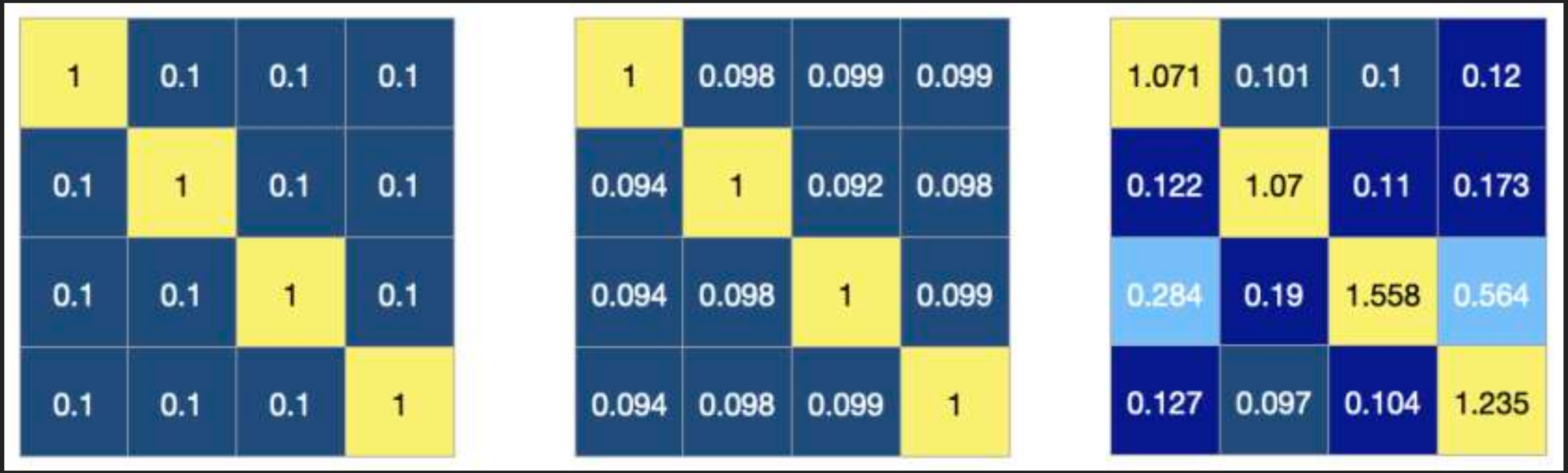
Average SDR across sources



Time taken in seconds

RESULTS

$$X = \Lambda S$$



True Λ

1	0.1	0.1	0.1
0.1	1	0.1	0.1
0.1	0.1	1	0.1
0.1	0.1	0.1	1

Predicted Λ

1	0.098	0.099	0.099
0.094	1	0.092	0.098
0.094	0.098	1	0.099
0.094	0.098	0.099	1

KAMIR Λ

1.071	0.101	0.1	0.12
0.122	1.07	0.11	0.173
0.284	0.19	1.558	0.564
0.127	0.097	0.104	1.235

Interference Matrix Λ

TEST ON LIVE RECORDINGS AND LIMITATIONS

- **Linearity**: Mixtures in real world follows non-linear mixing.
- High computation time.
- Basic model.

REPLACING WITH NEURAL NETWORK

REPLACING WITH NEURAL NETWORK

- Why?

REPLACING WITH NEURAL NETWORK

- Why?
- Datasets?

REPLACING WITH NEURAL NETWORK

- Why?
- Datasets?
- Generalisability?

REPLACING WITH NEURAL NETWORK

$$X = \Lambda S$$

- Why?
- Datasets?
- Generalisability?

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N} \\ \vdots & & \vdots & \\ \lambda_{K1} & \lambda_{K2} & \dots & \lambda_{KN} \end{pmatrix} \quad X = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_K(t) \end{pmatrix} \quad S = \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{pmatrix}$$

REPLACING WITH NEURAL NETWORK

$$X = \Lambda S$$

- Why?
- Datasets?
- Generalisability?

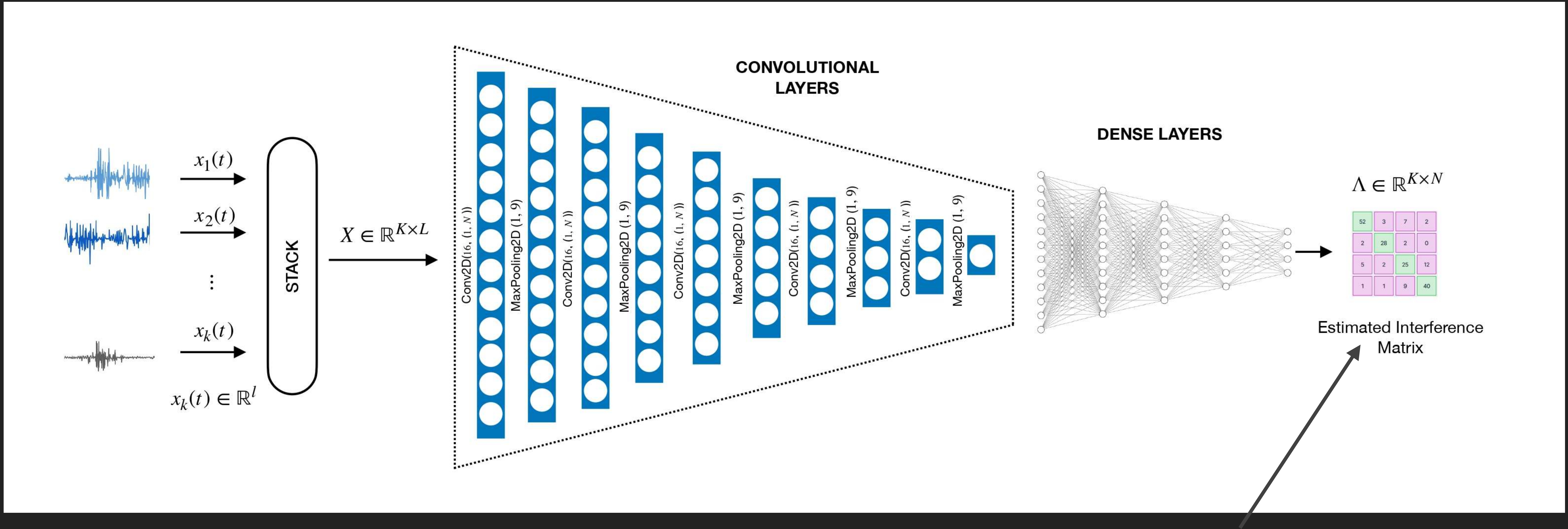
$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N} \\ \vdots & & \vdots & \\ \lambda_{K1} & \lambda_{K2} & \dots & \lambda_{KN} \end{pmatrix} \quad X = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_K(t) \end{pmatrix} \quad S = \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{pmatrix}$$

The interference reduced sources can be estimated by,

$$\hat{S} = \Lambda^{\dagger} X$$

Where † is the pseudo inverse of Λ.

TRUNCATED UNET ARCHITECTURE



52	3	7	2
2	28	2	0
5	2	25	12
1	1	9	40

$$X = \Lambda S$$

DATASET 2: FOR TESTING (REAL MIXTURES – CM)

ACOUSTICALLY TREATED

https://images.app.goo.gl/oMMMJN7VJ4inwNnq8

RANDOM ROOM

https://images.app.goo.gl/65HCSCiKP55FfWVMA

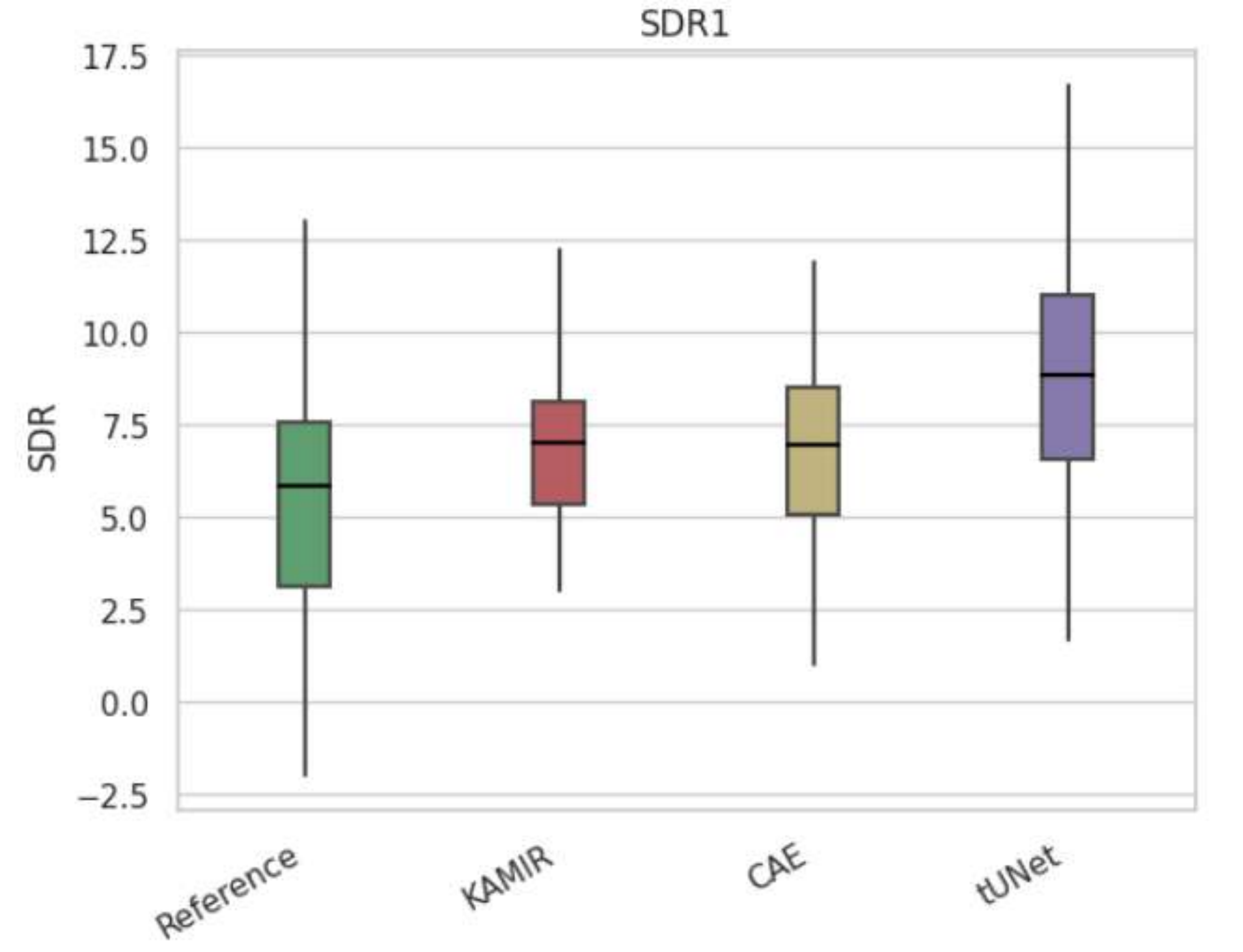
DATASET 2: FOR TESTING (REAL MIXTURES – CM)

- Stimulated artificial room using *pyroomacoustics*[3]
- Dataset created with room impulse response, time delays, reverberations.
- Resembles more natural with live recordings. Same set of LM source set utilised

---

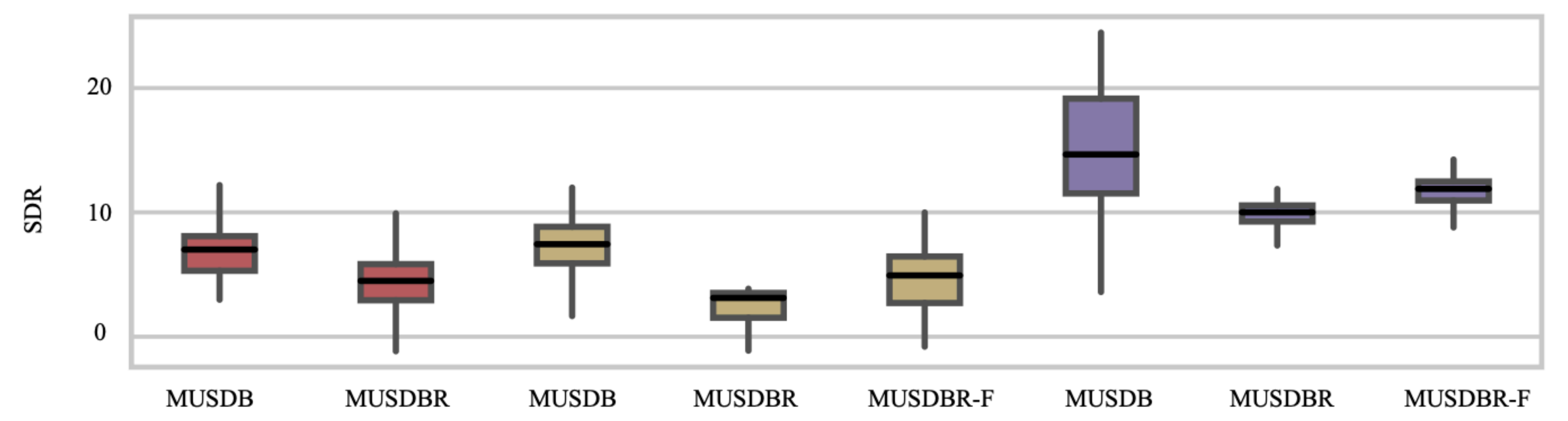
[3] *Scheibler, Robin, Eric Bezzam, and Ivan Dokmanić. "Pyroomacoustics: A python package for audio room simulation and array processing algorithms." 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2018.*

RESULTS



Linear Mixtures

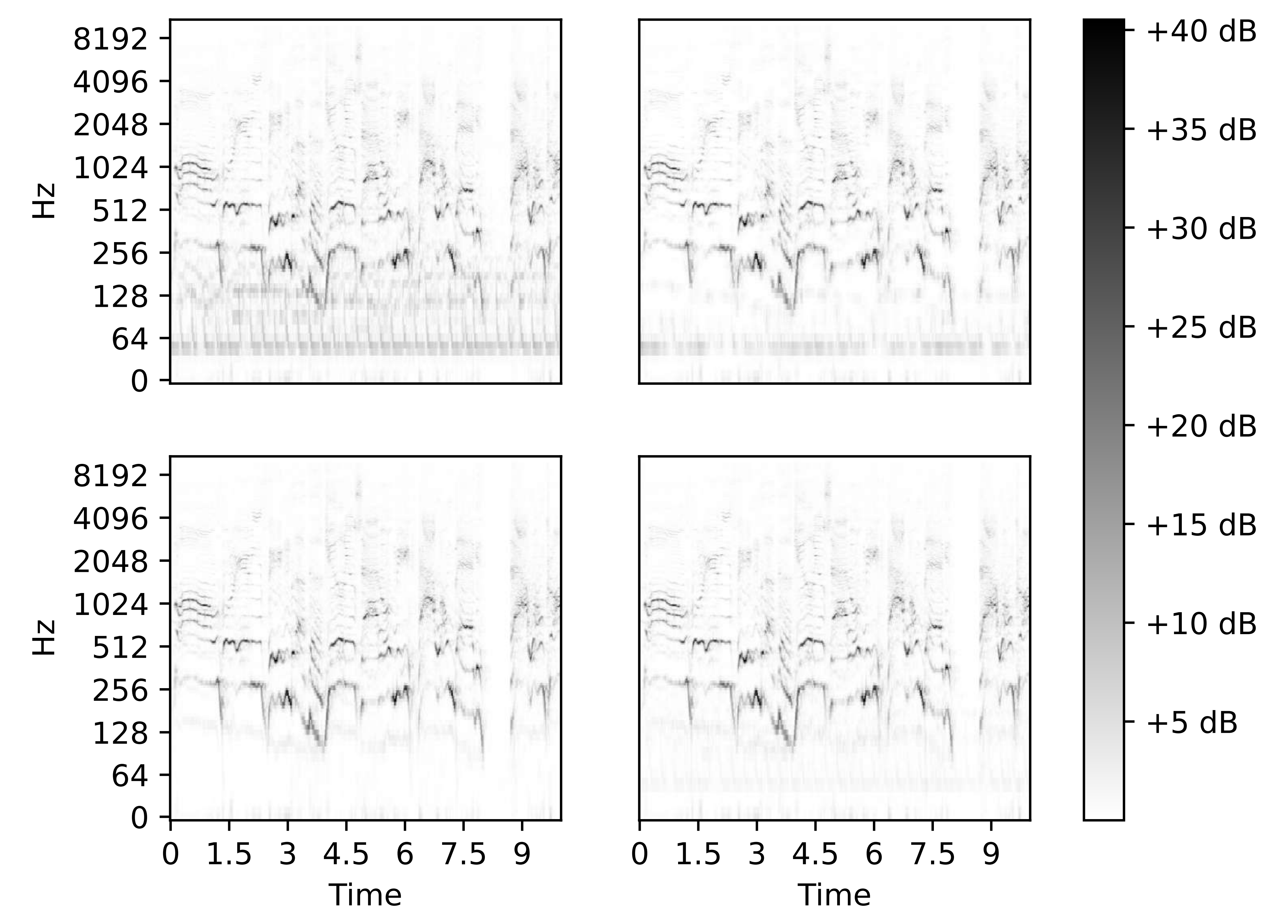
RESULTS



- 🟥 KAMIR
- 🟨 CAEs
- 🟪 t-UNet

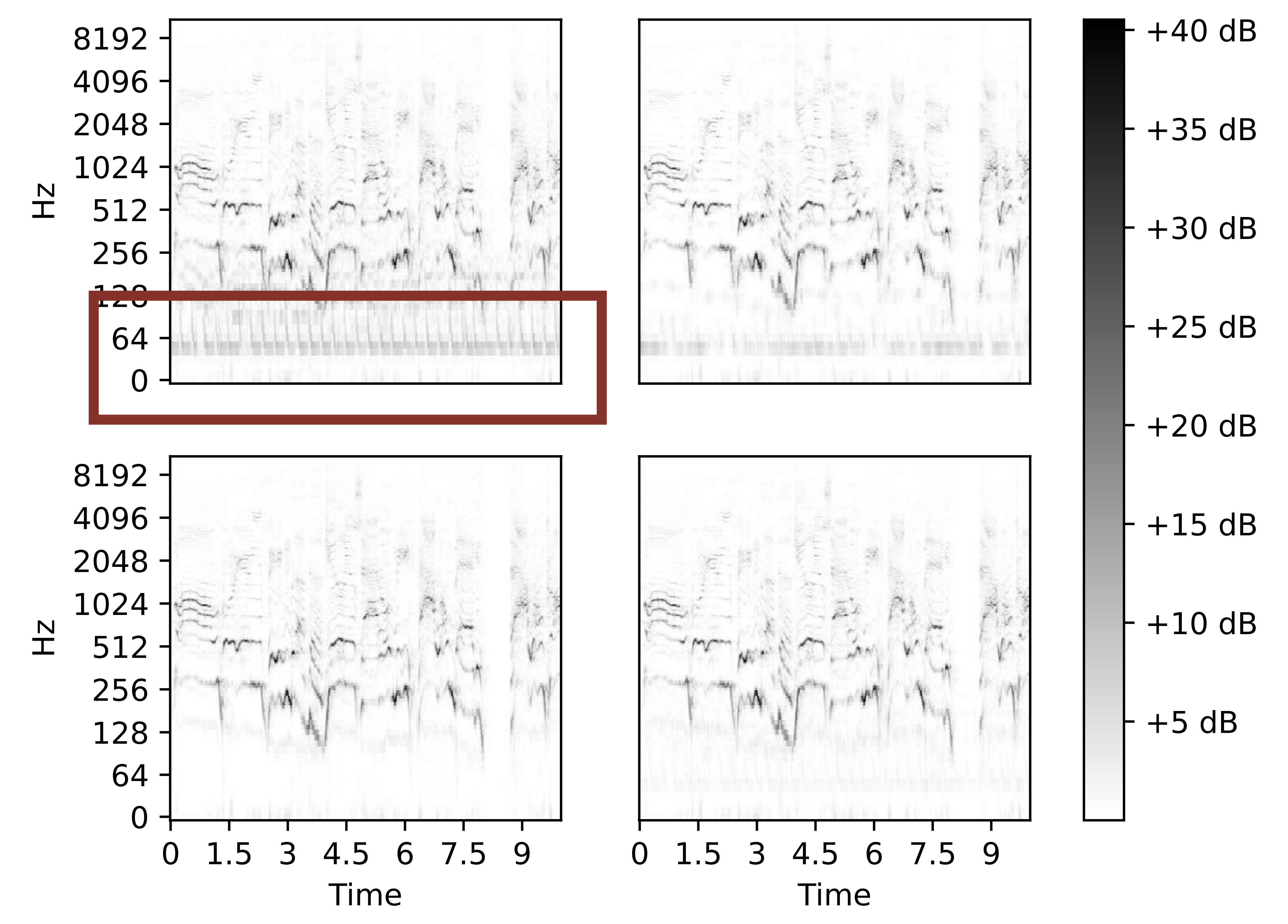
MUSDB	Linear Mixtures (LM)
MUSDBR	Realistic Mixtures (CM)
MUSDBR-F	LM finetuned with CM

RESULTS



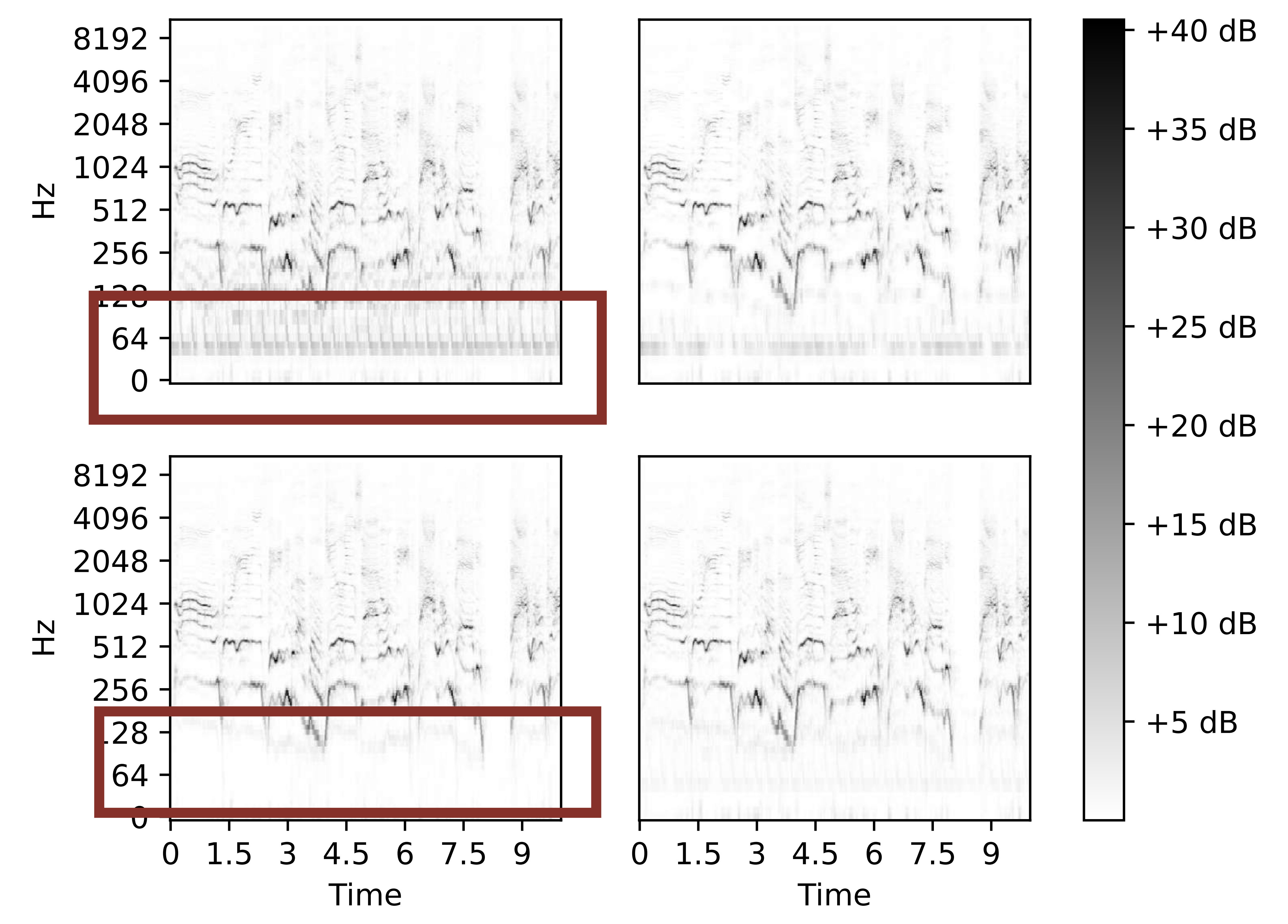
Spectrograms of Vocal Source

RESULTS



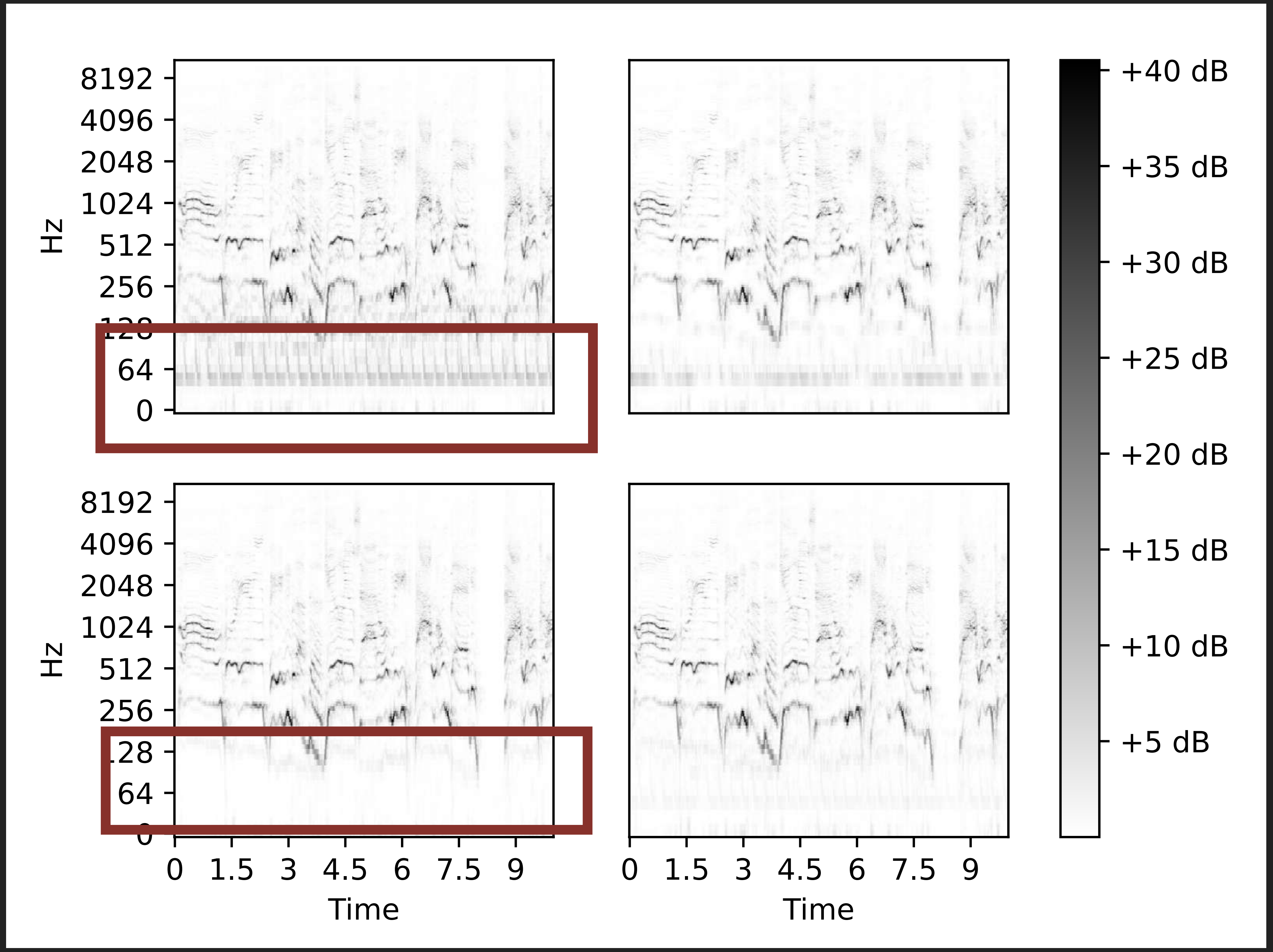
Spectrograms of Vocal Source

RESULTS

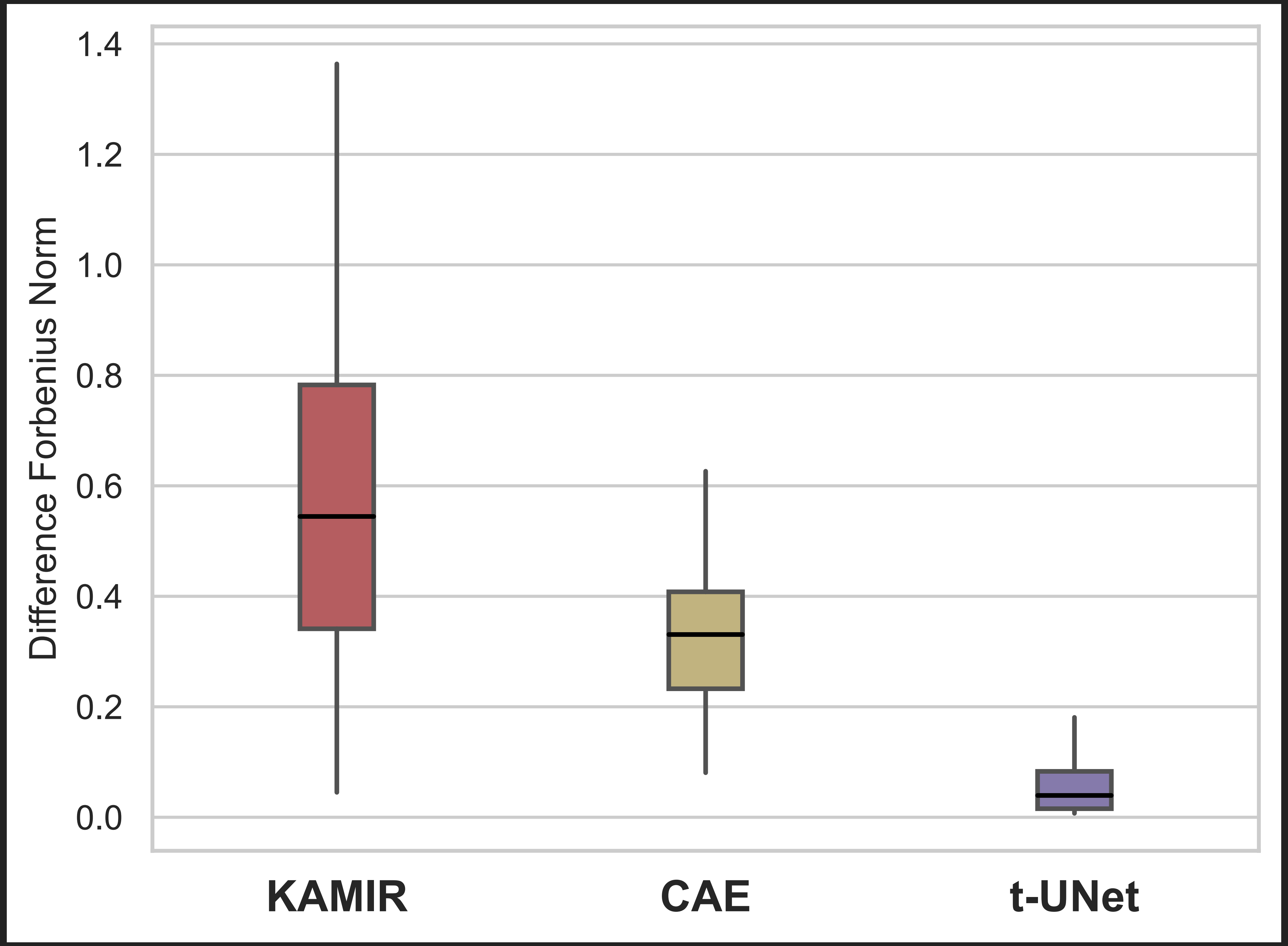


Spectrograms of Vocal Source

RESULTS



Spectrograms of Vocal Source



Difference of Frobenius norm of the true Λ with the predicted $\hat{\Lambda}$.

CONCLUSION OF THE APPROACHES

CONCLUSION OF THE APPROACHES

- Proposed two neural networks for interference reduction: CAEs and t-UNet, both performing better than KAMIR

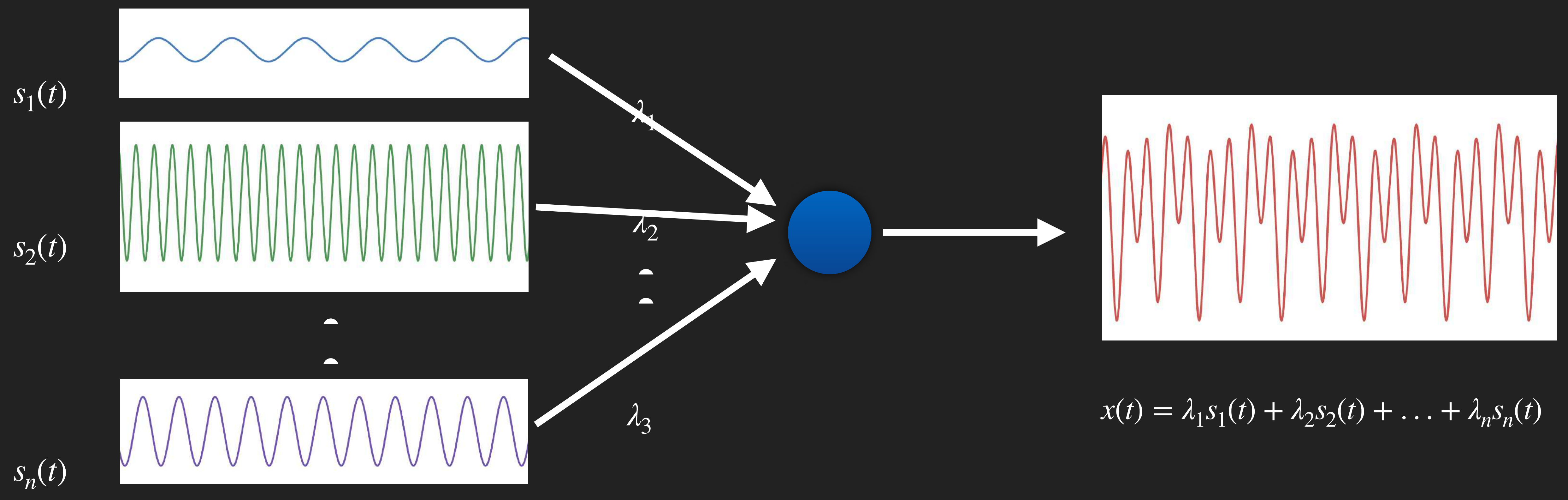
CONCLUSION OF THE APPROACHES

- Proposed two neural networks for interference reduction: CAEs and t-UNet, both performing better than KAMIR
- CAEs has difficulties in generalising and works in TF domain where t-UNet reduces interference directly by learning interference matrix.

CONCLUSION OF THE APPROACHES

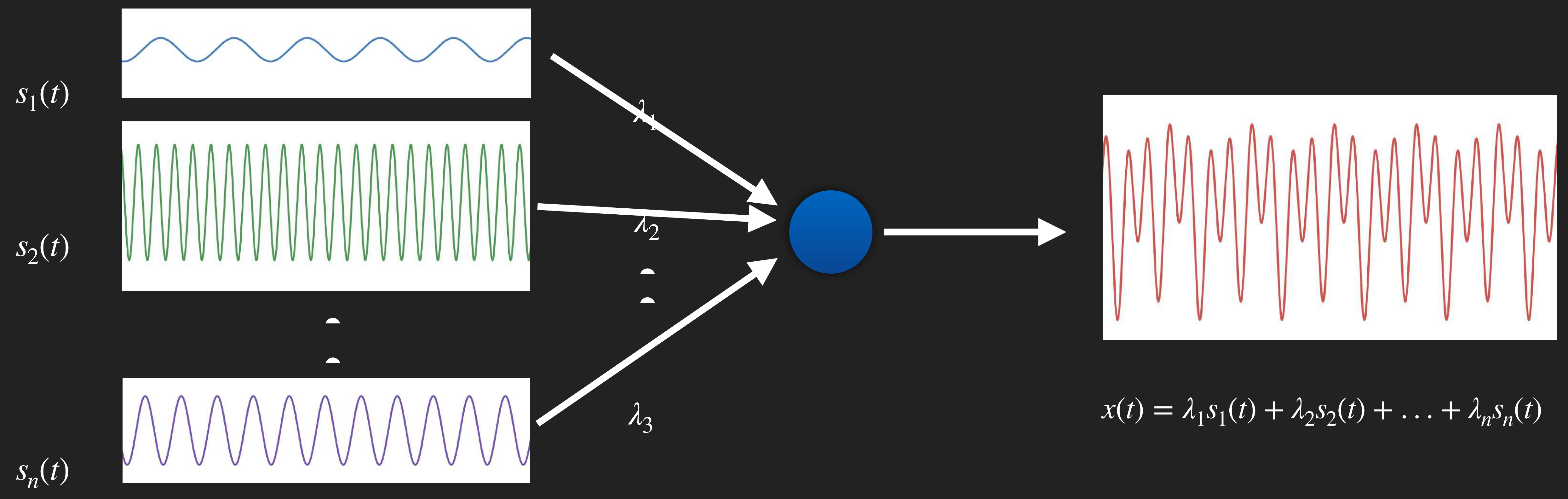
- Proposed two neural networks for interference reduction: CAEs and t-UNet, both performing better than KAMIR
- CAEs has difficulties in generalising and works in TF domain where t-UNet reduces interference directly by learning interference matrix.
- t-UNet outperforms all the models in-terms of SDR and computationally faster

RE-DEFINING PROBLEM



$$x(t) = \lambda_1 s_1(t) + \lambda_2 s_2(t) + \ldots + \lambda_n s_n(t)$$

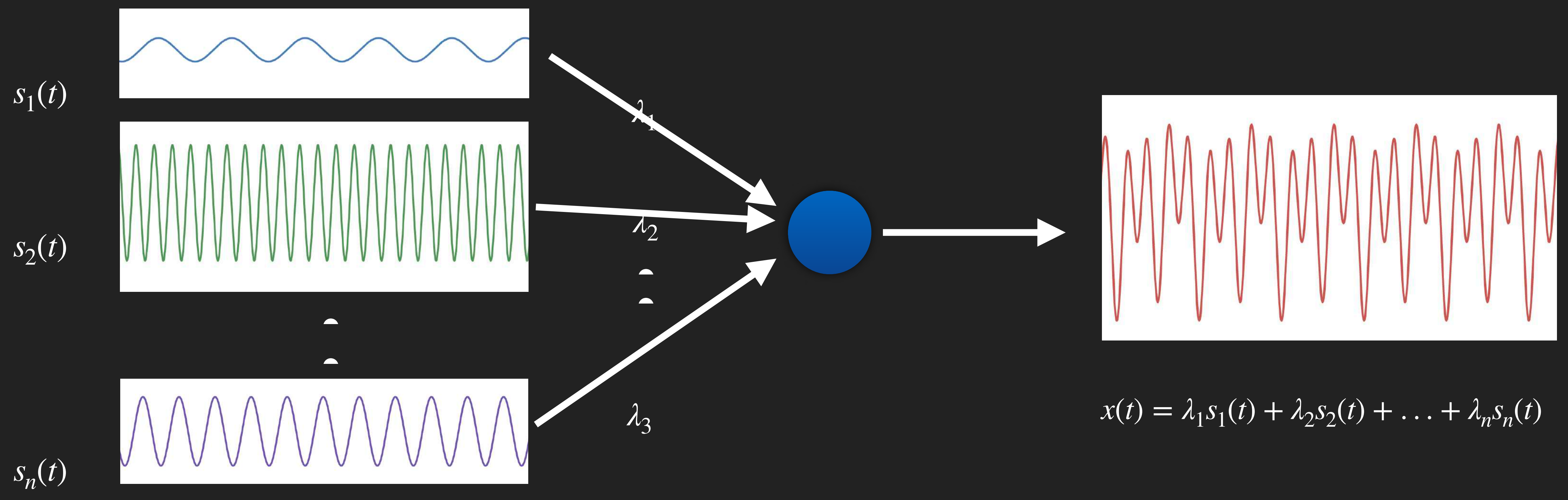
RE-DEFINING PROBLEM



$$x(t) = \lambda_1 s_1(t) + \lambda_2 s_2(t) + \ldots + \lambda_n s_n(t)$$

- tUNet built with the mathematical approximation of the problem as $X = \Lambda S$ which is still **linear!**

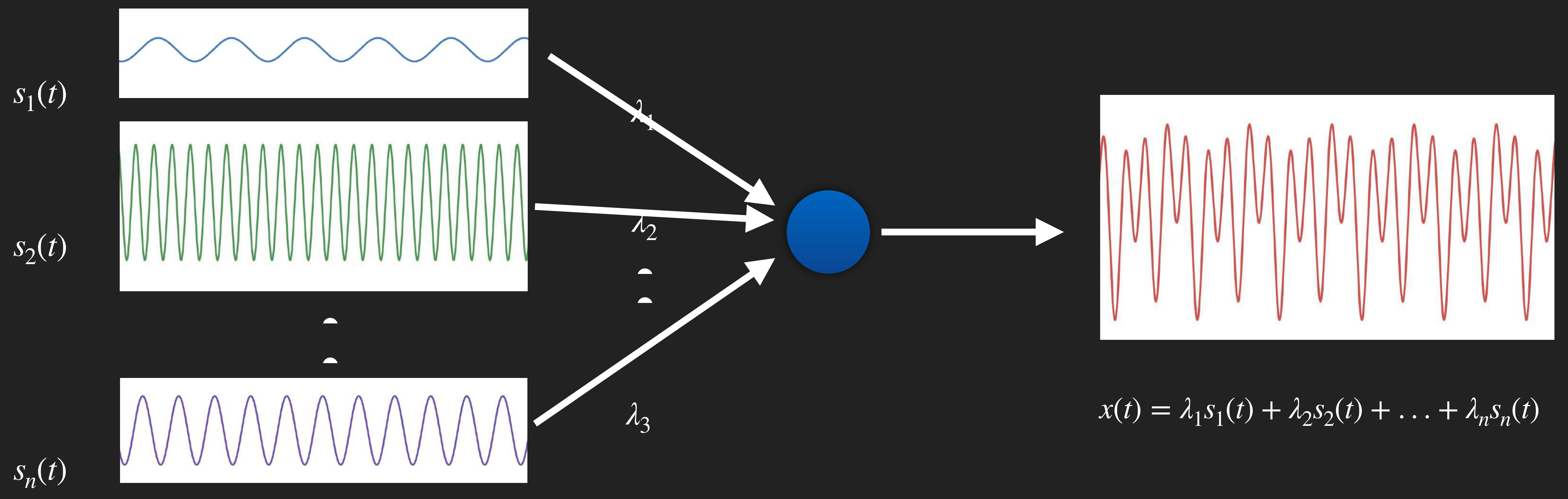
RE-DEFINING PROBLEM



$$x(t) = \lambda_1 s_1(t) + \lambda_2 s_2(t) + \ldots + \lambda_n s_n(t)$$

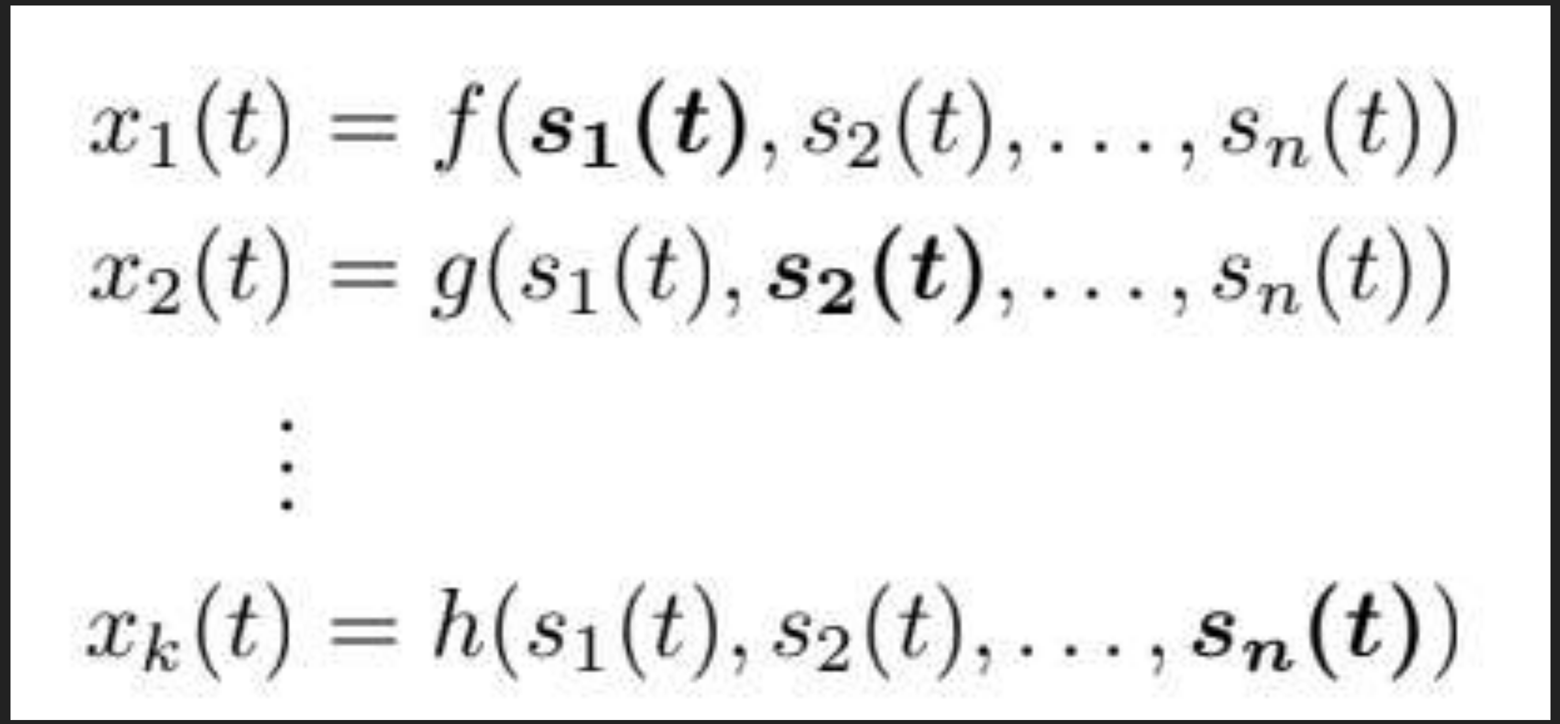
- tUNet built with the mathematical approximation of the problem as $X = \Lambda S$ which is still **linear!**
- Initial evaluations of the live recordings reveals the t-UNet is not effective.

RE-DEFINING PROBLEM



$x(t) = \lambda_1 s_1(t) + \lambda_2 s_2(t) + \ldots + \lambda_n s_n(t)$

For k microphones and n sources,

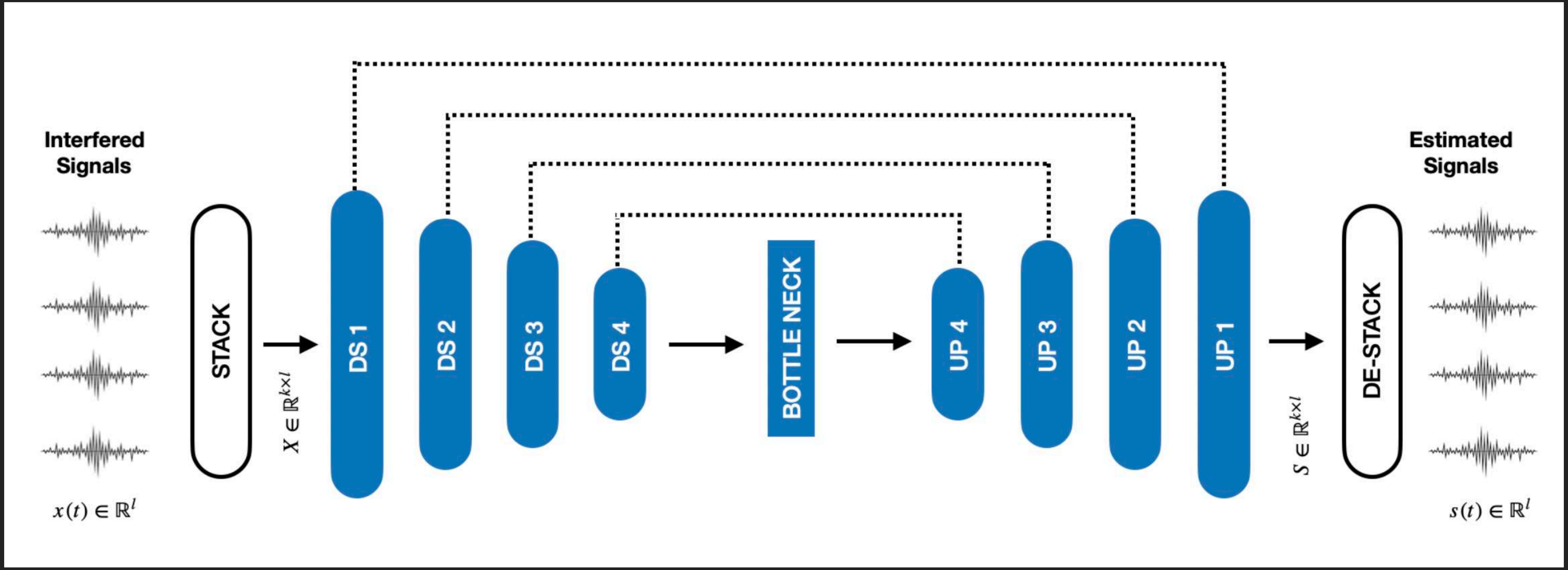


$$\begin{aligned} x_1(t) &= f(\mathbf{s_1(t)}, s_2(t), \ldots, s_n(t)) \\ x_2(t) &= g(s_1(t), \mathbf{s_2(t)}, \ldots, s_n(t)) \\ &\vdots \\ x_k(t) &= h(s_1(t), s_2(t), \ldots, \mathbf{s_n(t)}) \end{aligned}$$

Where $f(\,.\,)$, $g(\,.\,)$, and $h(\,.\,)$ are unknown functions

- tUNet built with the mathematical approximation of the problem as $X = \Lambda S$ which is still **linear!**
- Initial evaluations of the live recordings reveals the t-UNet is not effective.

DILATED FULL WAVE U NET ARCHITECTURE

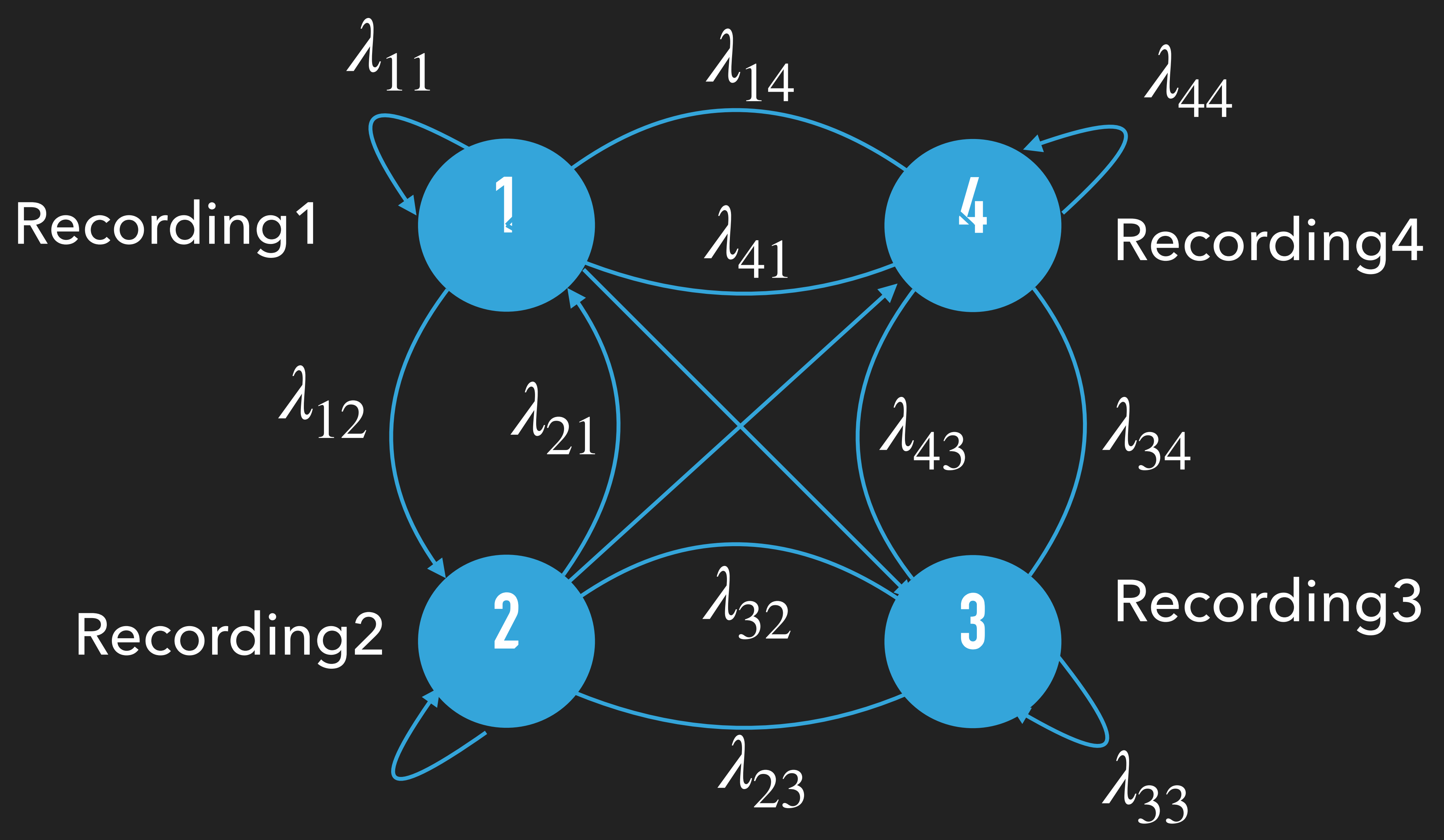


GRAPH ATTENTIONS IN AUDIO DOMAIN

- Treating each audio as a node
- Each vertices strength corresponds to the interference strength among recordings

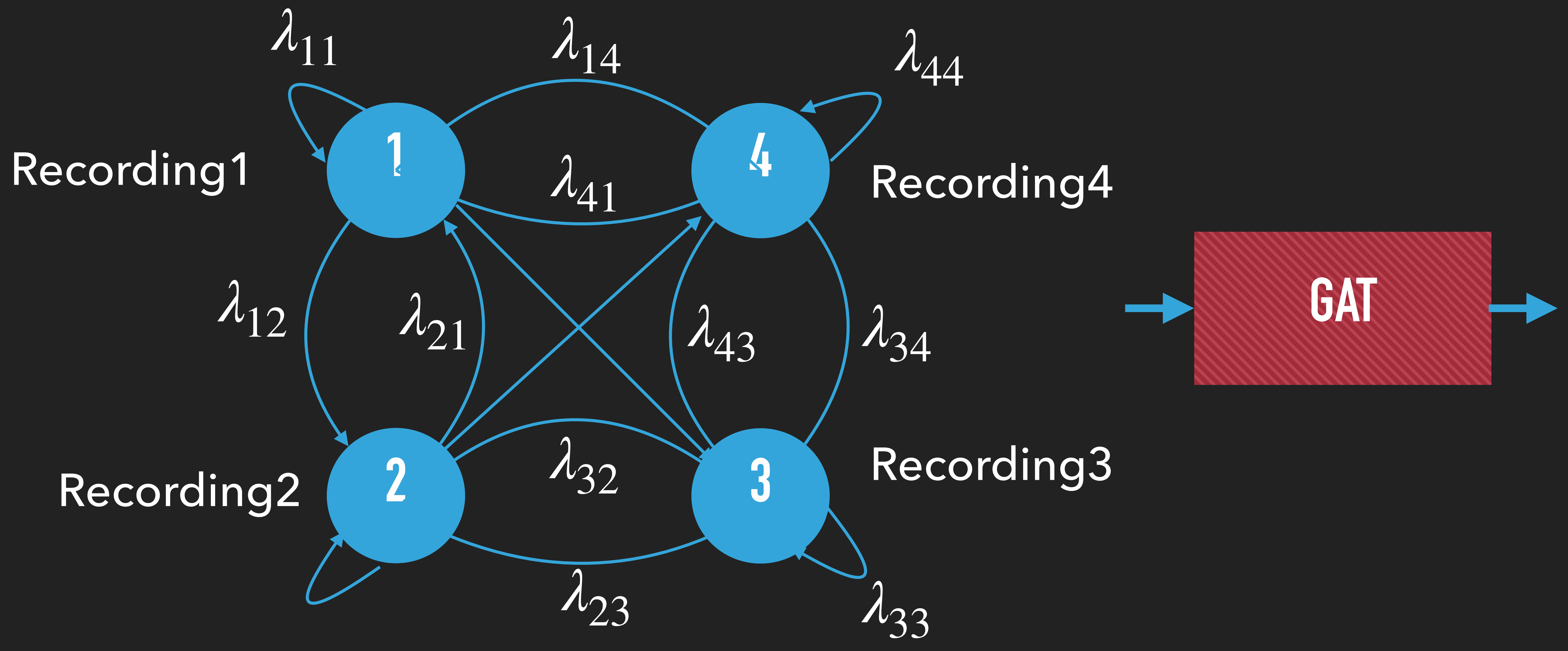
GRAPH ATTENTIONS IN AUDIO DOMAIN

- Treating each audio as a node
- Each vertices strength corresponds to the interference strength among recordings



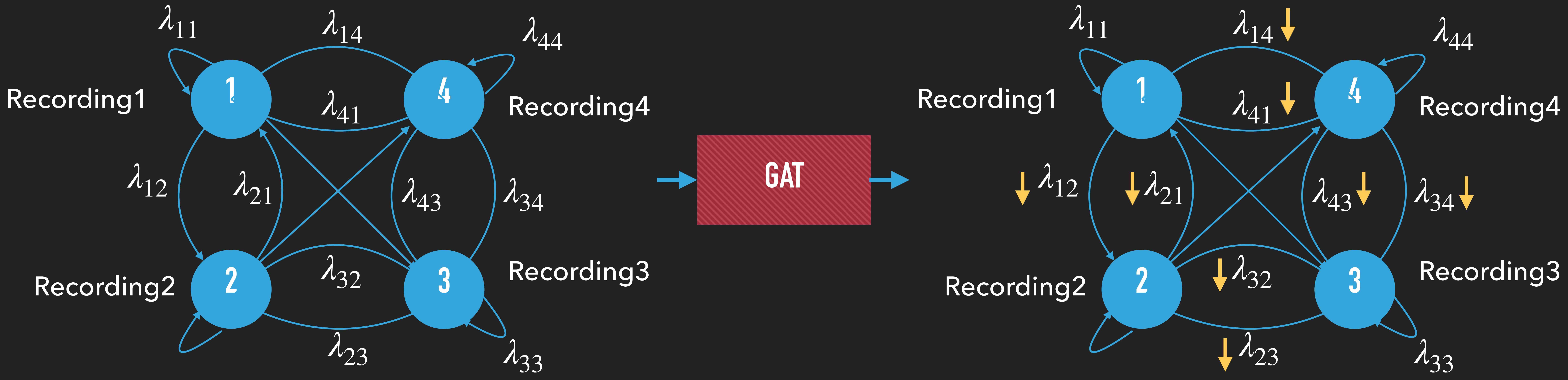
GRAPH ATTENTIONS IN AUDIO DOMAIN

- Treating each audio as a node
- Each vertices strength corresponds to the interference strength among recordings

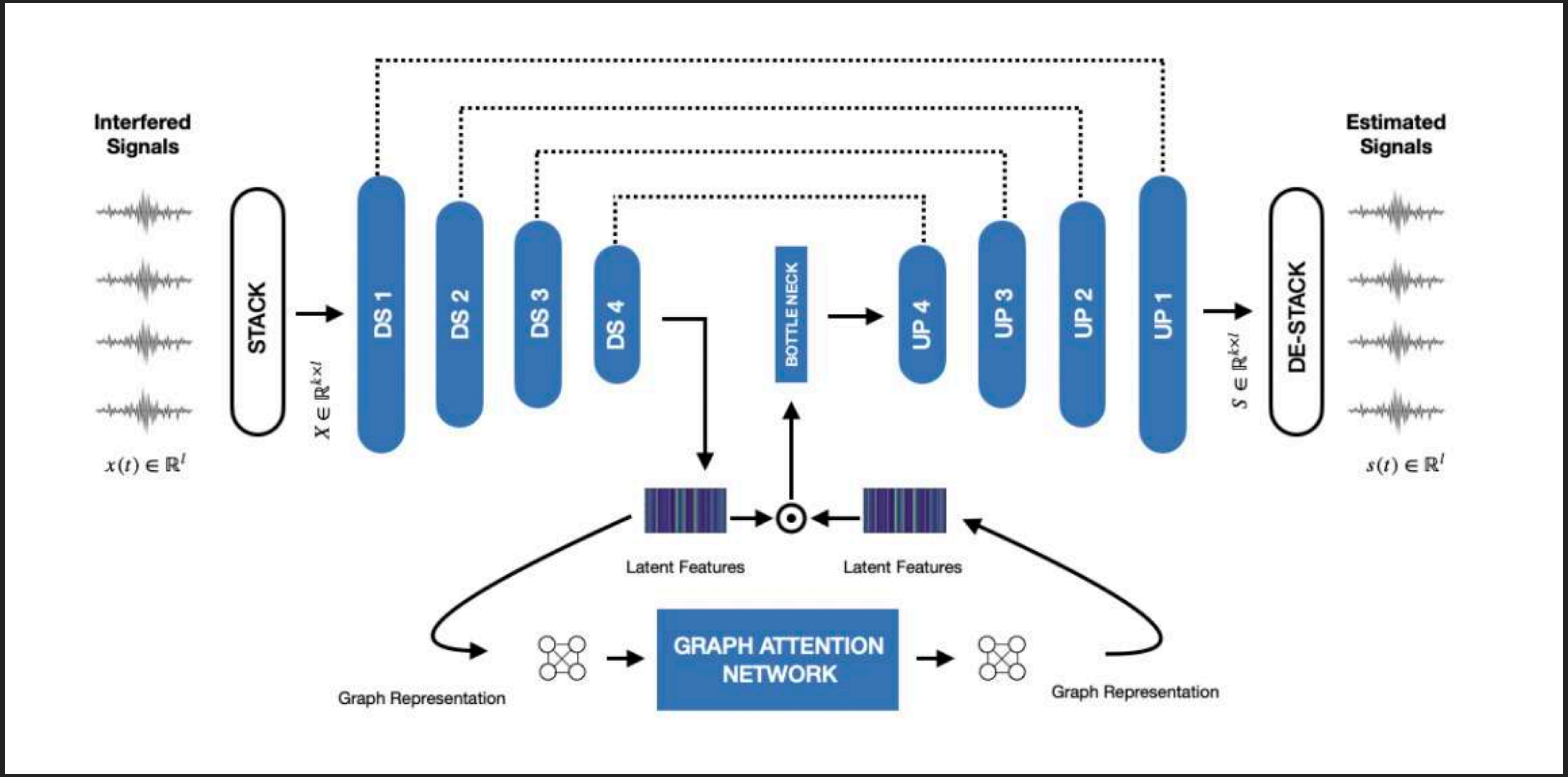


GRAPH ATTENTIONS IN AUDIO DOMAIN

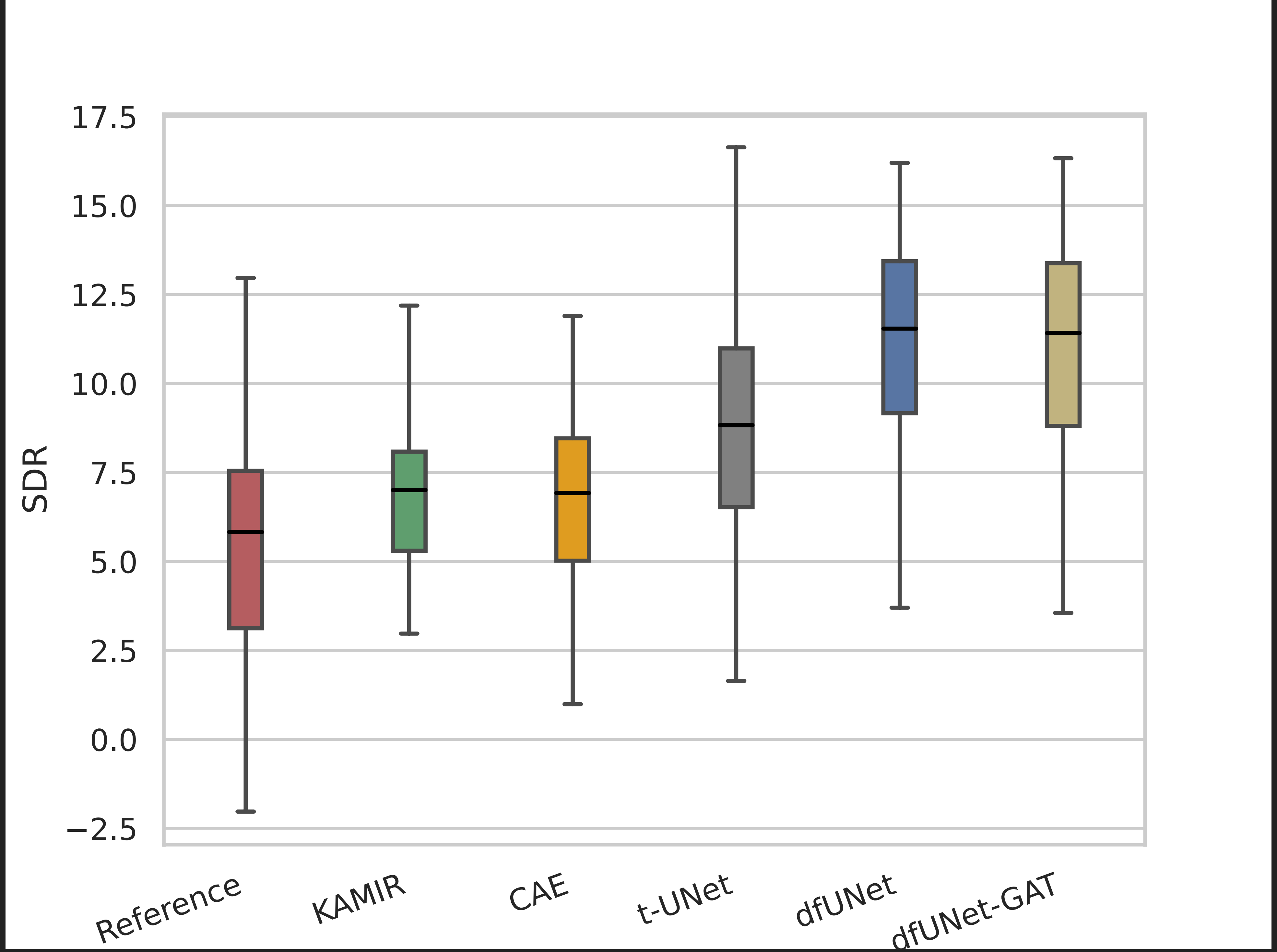
- Treating each audio as a node
- Each vertices strength corresponds to the interference strength among recordings



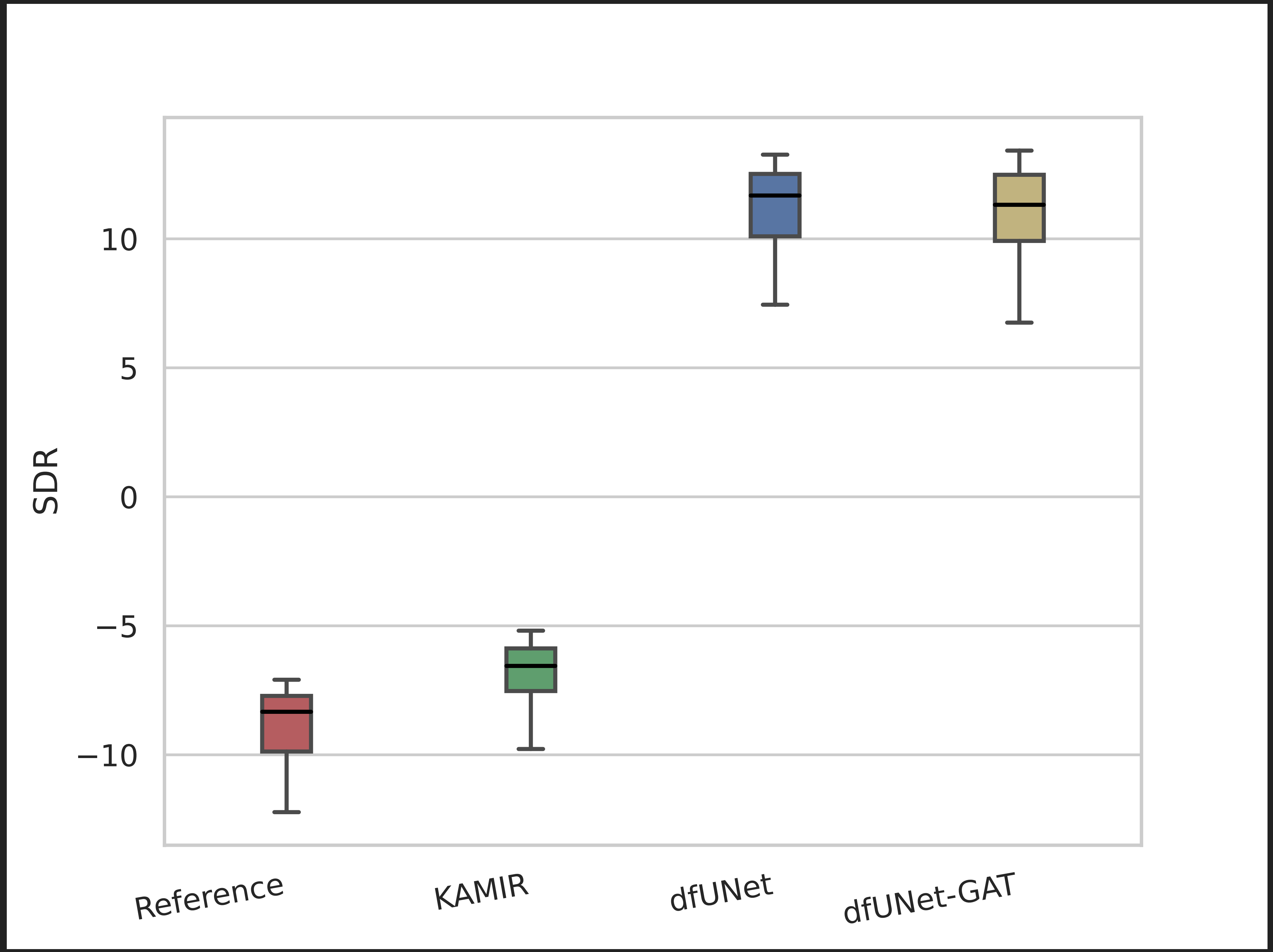
DILATED FULL WAVE U NET WITH GRAPH ATTENTIONS



RESULTS



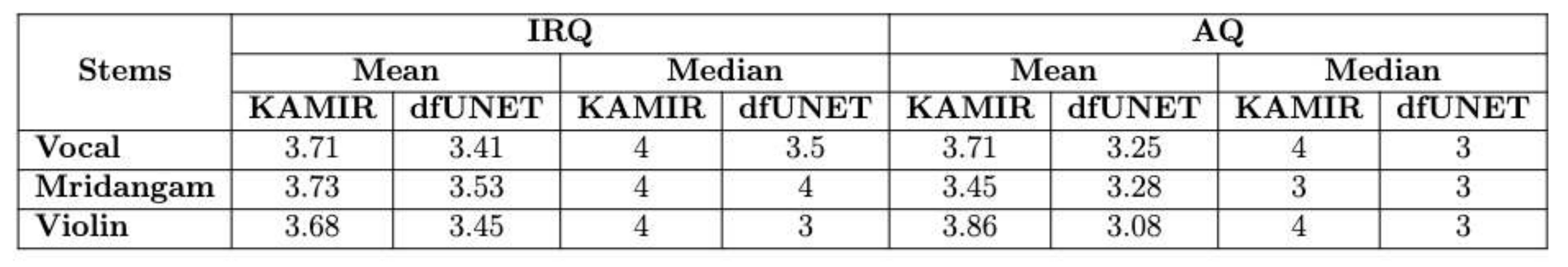
Linear Mixtures



Realistic Mixtures

TEST ON LIVE RECORDINGS (OUT OF DOMAIN SAMPLES)

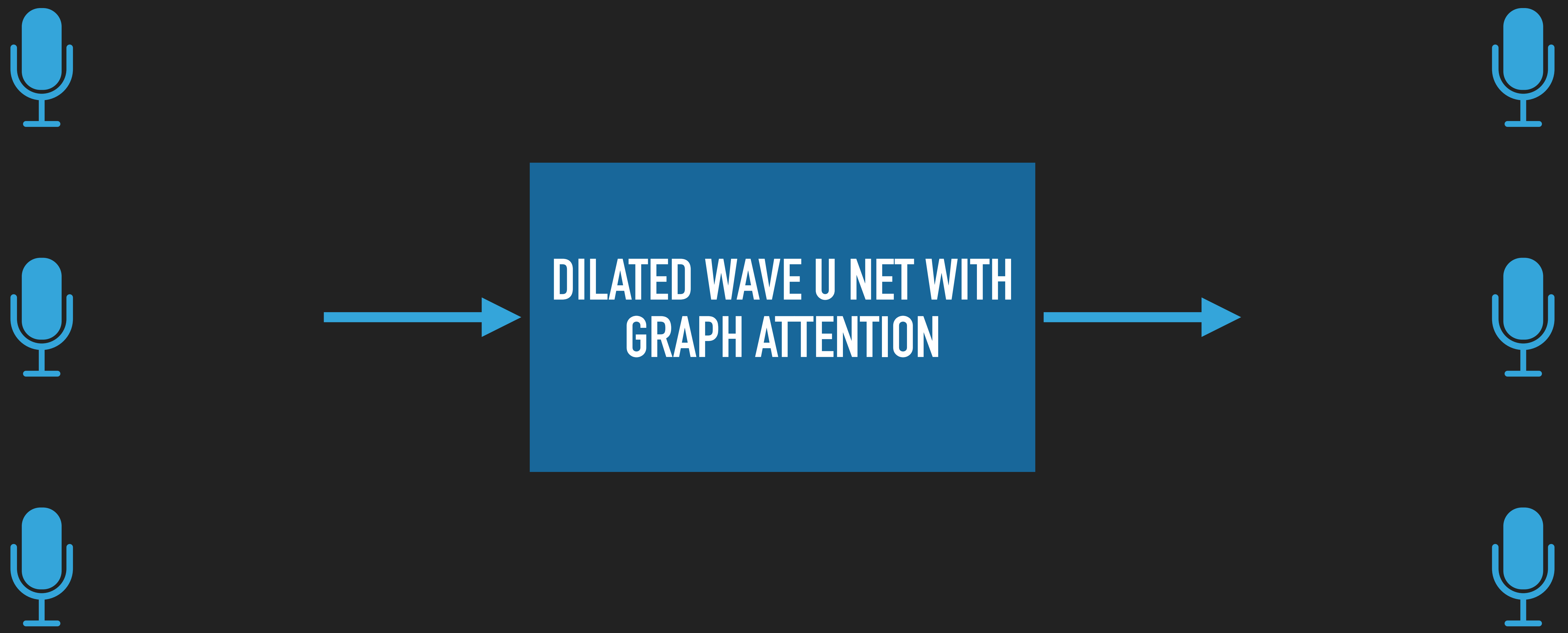
Interference Reduction Quality　　　　Audio Quality



Stems	IRQ				AQ			
	Mean		Median		Mean		Median	
	KAMIR	dfUNET	KAMIR	dfUNET	KAMIR	dfUNET	KAMIR	dfUNET
Vocal	3.71	3.41	4	3.5	3.71	3.25	4	3
Mridangam	3.73	3.53	4	4	3.45	3.28	3	3
Violin	3.68	3.45	4	3	3.86	3.08	4	3

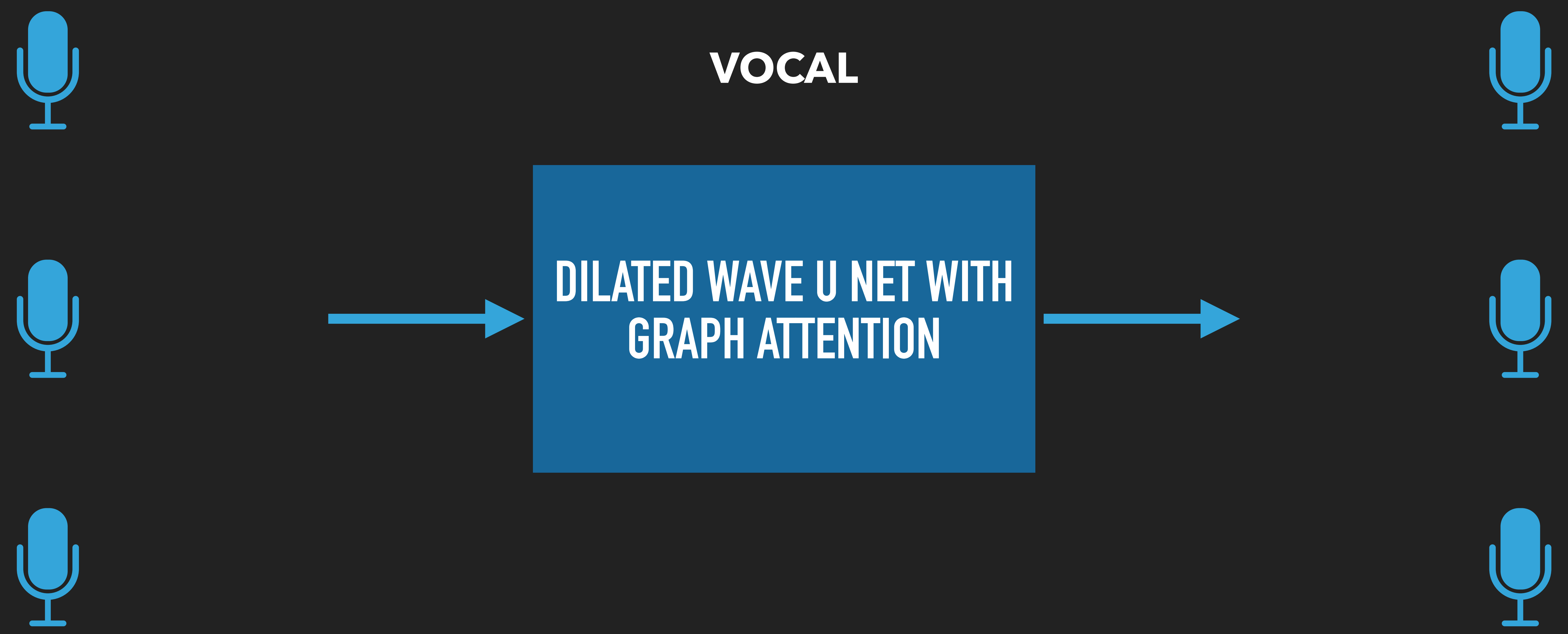
Listening Test Results: **44 Participants**

HOW DOES IT SOUND?



DILATED WAVE U NET WITH GRAPH ATTENTION

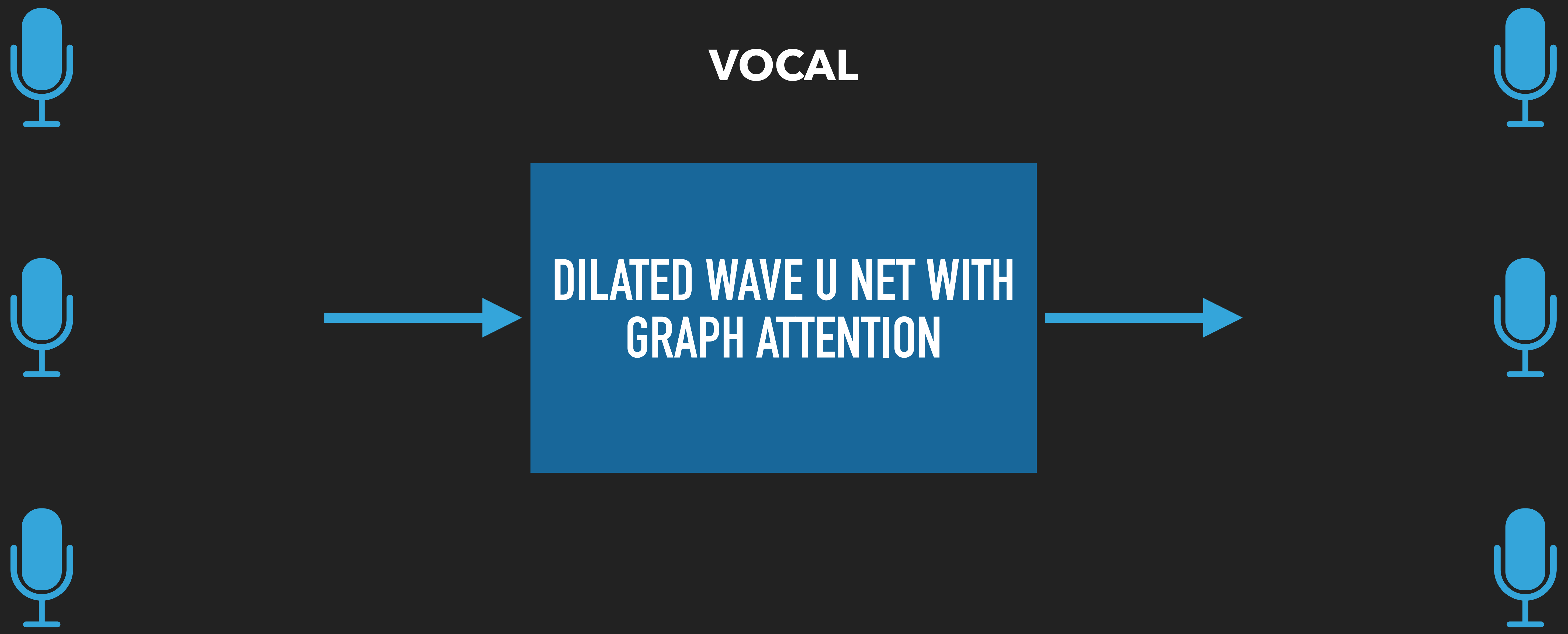
HOW DOES IT SOUND?



VOCAL

DILATED WAVE U NET WITH GRAPH ATTENTION

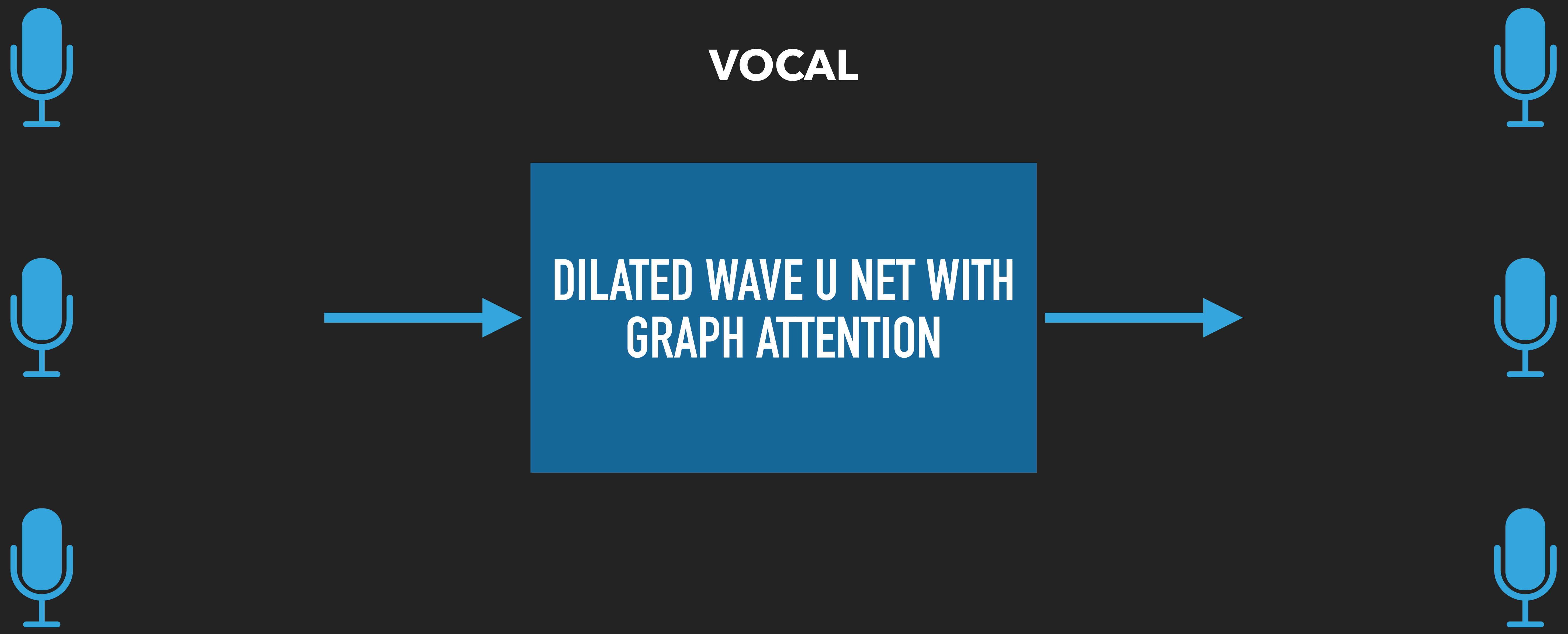
HOW DOES IT SOUND?



VOCAL

DILATED WAVE U NET WITH GRAPH ATTENTION

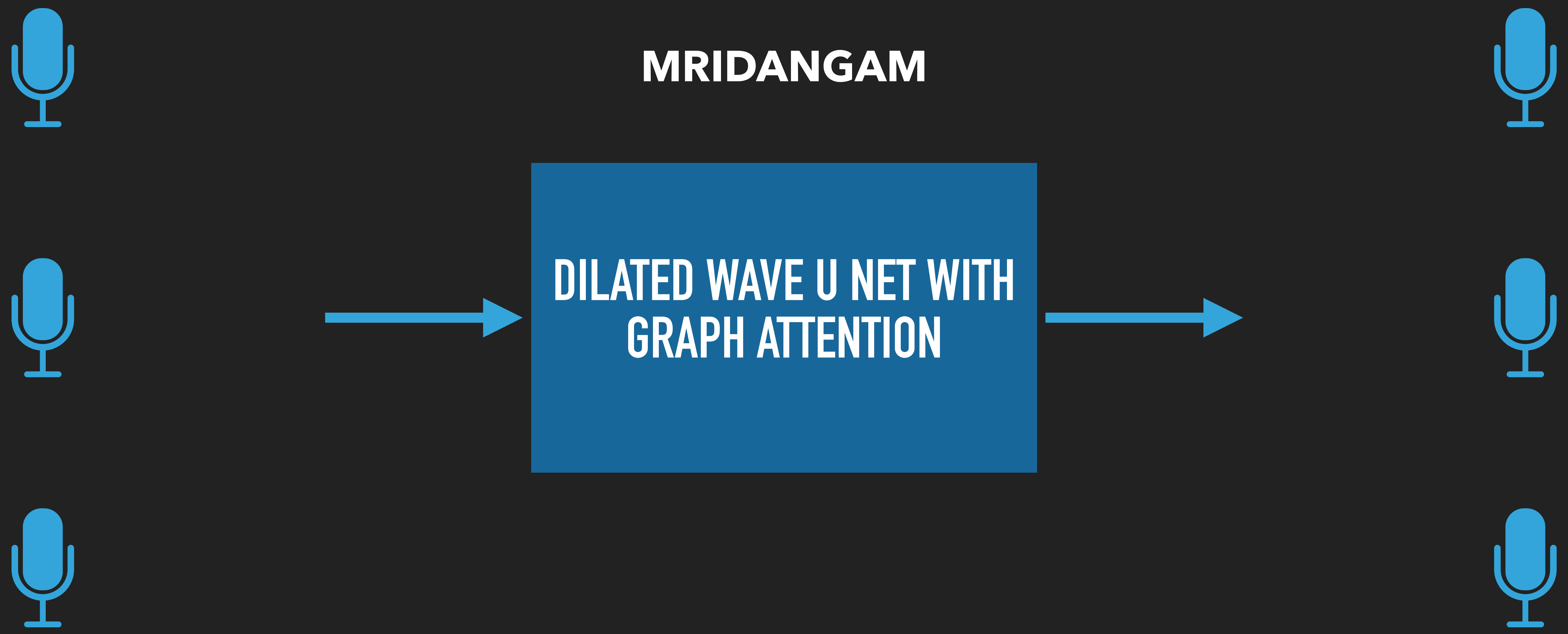
HOW DOES IT SOUND?



VOCAL

DILATED WAVE U NET WITH GRAPH ATTENTION

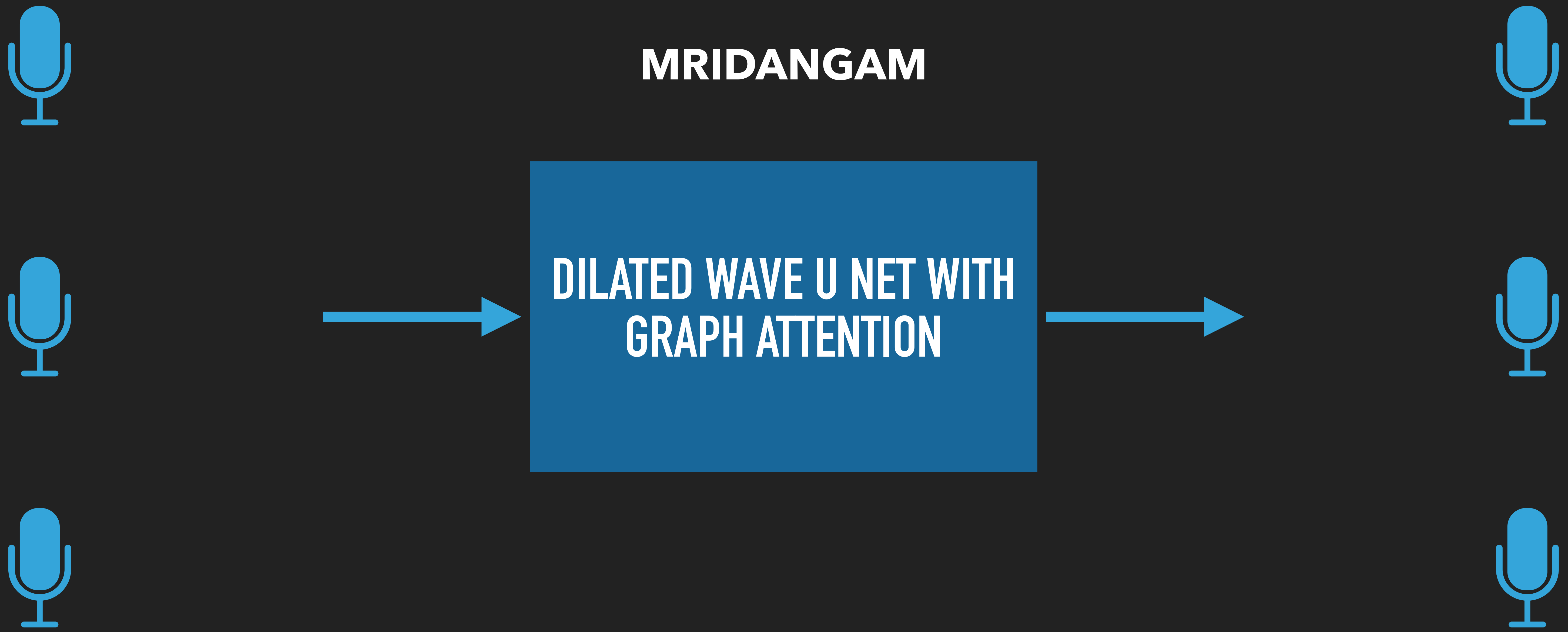
HOW DOES IT SOUND?



MRIDANGAM

DILATED WAVE U NET WITH GRAPH ATTENTION

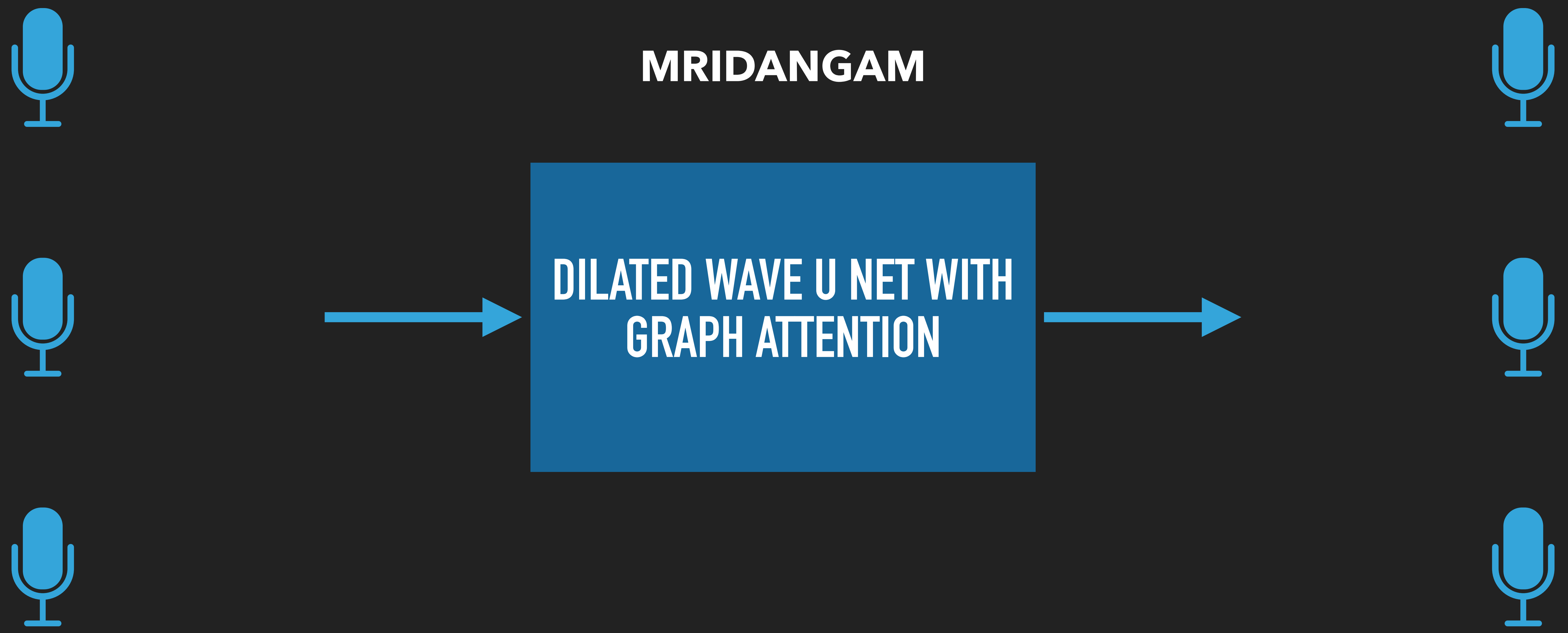
HOW DOES IT SOUND?



MRIDANGAM

DILATED WAVE U NET WITH GRAPH ATTENTION

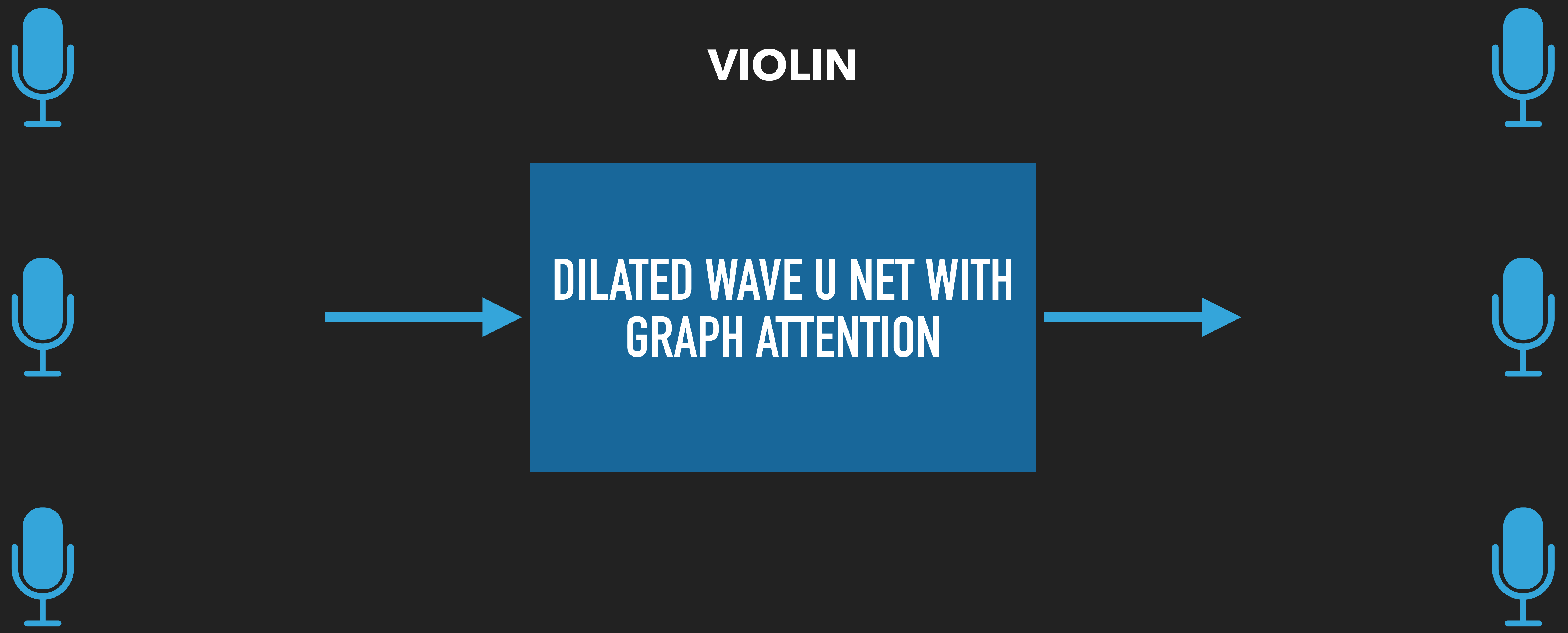
HOW DOES IT SOUND?



MRIDANGAM

DILATED WAVE U NET WITH GRAPH ATTENTION

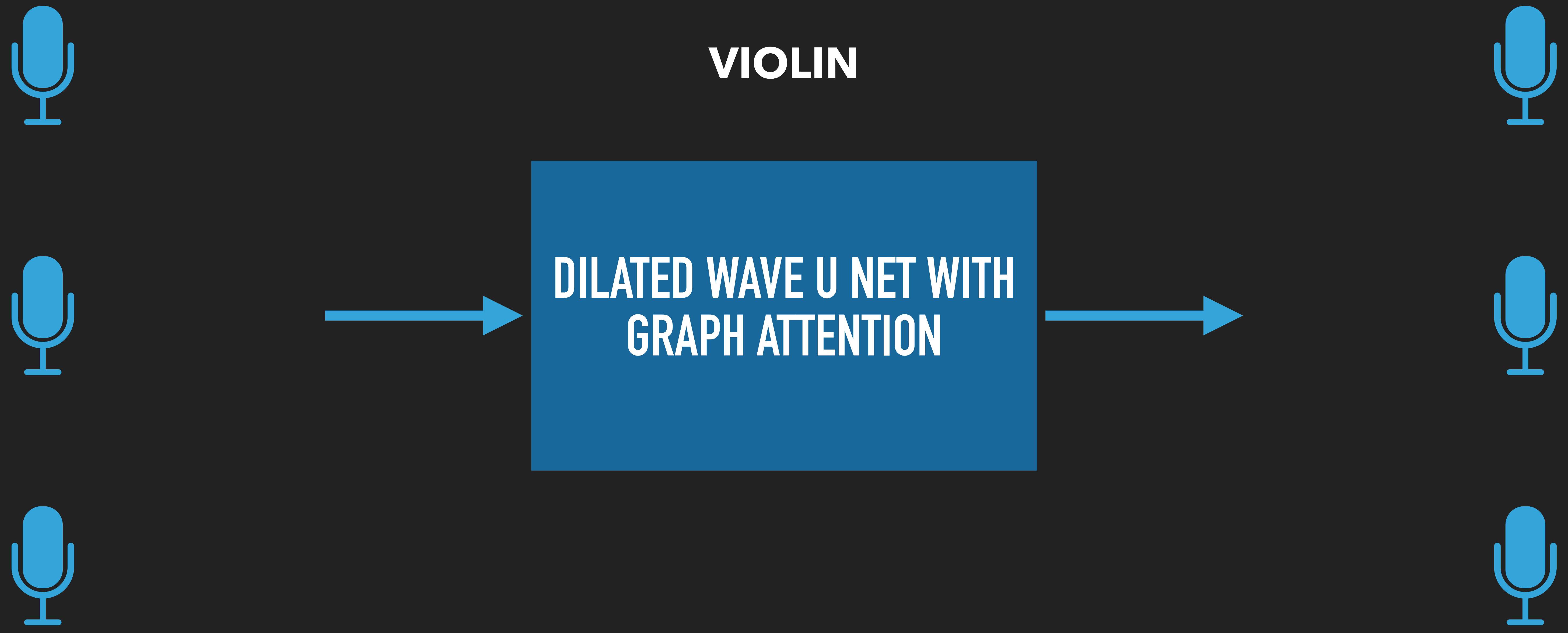
HOW DOES IT SOUND?



VIOLIN

DILATED WAVE U NET WITH GRAPH ATTENTION

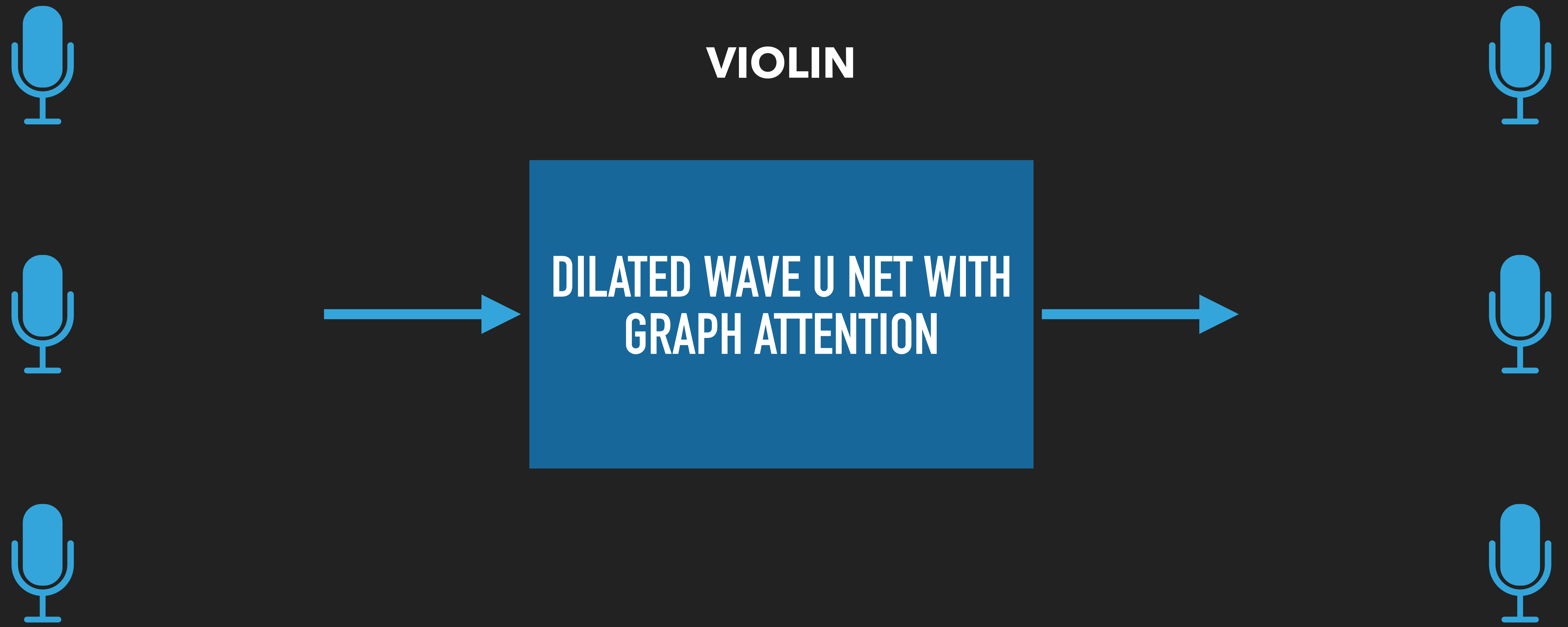
HOW DOES IT SOUND?



VIOLIN

DILATED WAVE U NET WITH GRAPH ATTENTION

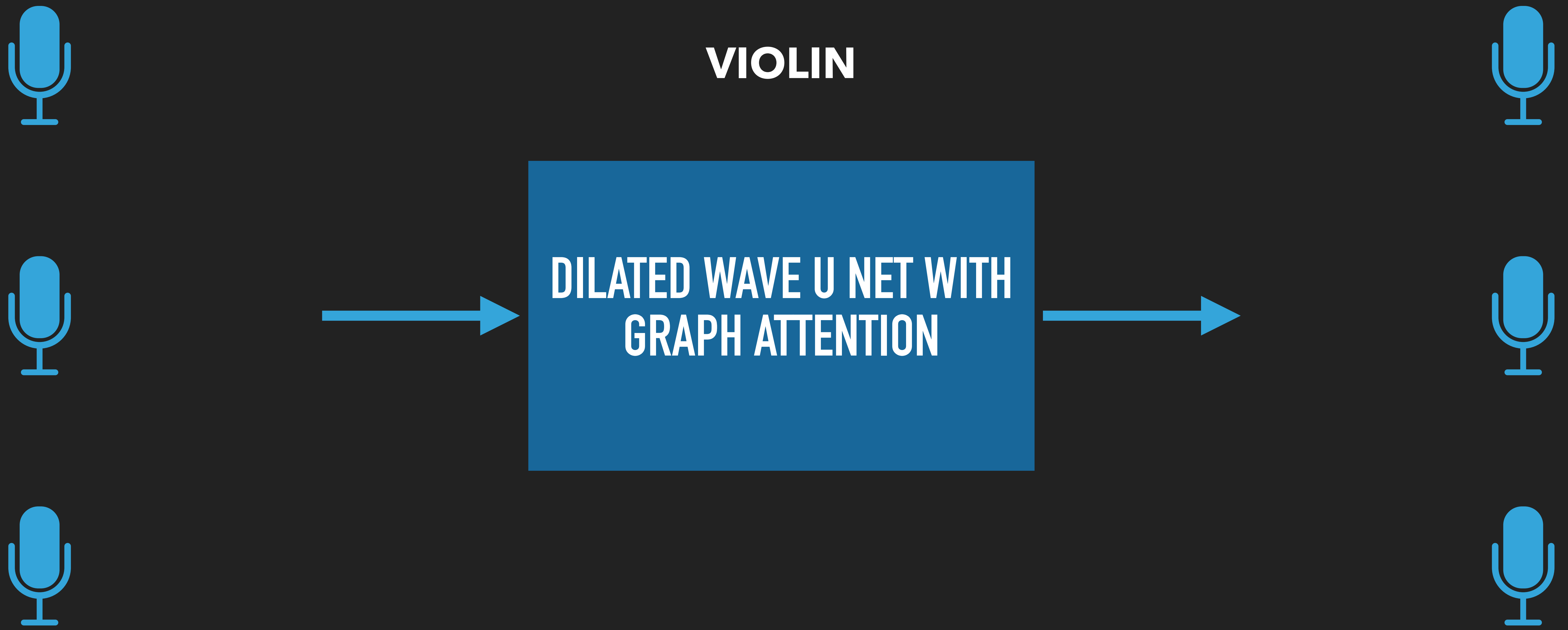
HOW DOES IT SOUND?



VIOLIN

DILATED WAVE U NET WITH GRAPH ATTENTION

HOW DOES IT SOUND?

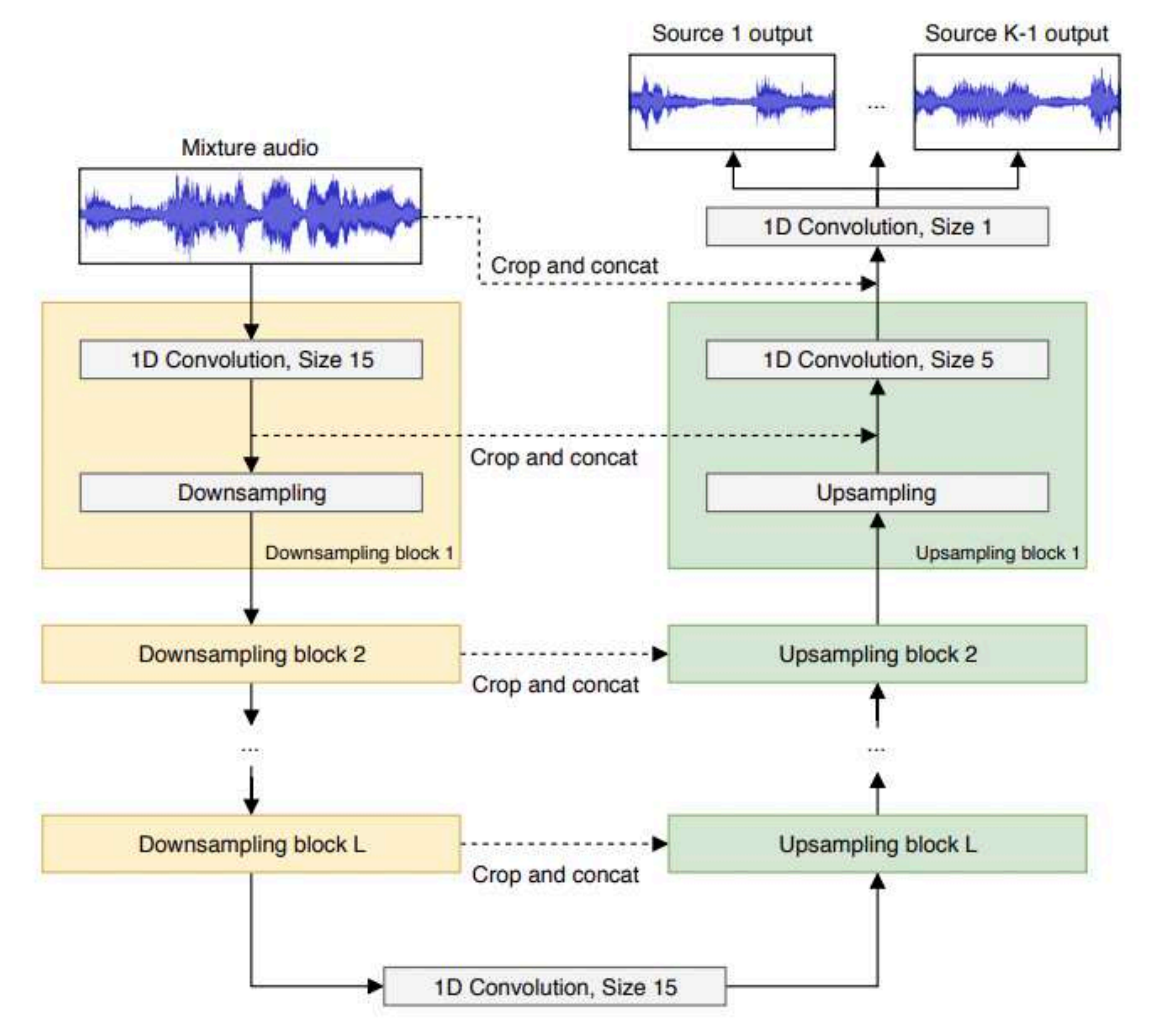


VIOLIN

DILATED WAVE U NET WITH GRAPH ATTENTION

MUSIC SOURCE SEPARATION FOR THE LIVE CARNATIC DATASET

TRAINING MSS: WAVE-U-NET MODEL



Two models:

1. Trained with MUSDB18 dataset
2. Trained with Live recorded Saraga dataset

Stoller, Daniel, Sebastian Ewert, and Simon Dixon. "Wave-u-net: A multi-scale neural network for end-to-end audio source separation." *arXiv preprint arXiv:1806.03185* (2018).

RESULTS FOR MUSDB18HQ & LIVE RECORDINGS

Wave-U-Net with MUSDB18HQ dataset,

	Clean	Interference	CAE Cleaned	t-UNet Cleaned
SDR	2.32	0.96	1.72	2.03

Wave-U-Net with Sagraga dataset,

	Clean	Interference (4 source)	Interference (4 source)	dfUNet Cleaned
SDR	NA	-0.19	1.16	To be filled

CONCLUSION

- Proposed IR models improves MSS performance
- Proposed IR models better than SOTA KAMIR in terms of SDR and Faster

	KAMIR	CAEs	tUNet	dfUNet
Average	1320.8	4.8	2.19	4.2

Table: Time taken in seconds for 200 test tracks of 10 seconds

FUTURE DIRECTIONS

- **Informed Source Separation**: Build end-to-end IR-MSS systems.
- DSP Techniques for IR: Beamformers, Direction of Arrival Estimation, etc.

PUBLICATIONS

- Rajesh R and Padmanabhan Rajan, "Neural Networks for Interference Reduction in Multi-Track Recordings," *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2023, pp. 1-5.
- Rajesh R and Padmanabhan Rajan, “Interference reduction in live recordings” communicating to *Transactions in Audio, Speech, and Language Processing (TASLP)* 2024 (under preparation)

THANKS FOR YOUR TIME AND ATTENTION