

# 3次元音空間を自在に操る

## —音の波動と聴覚特性を活かした 新しい音の記録・再生・操作技術—

小山 翔一

国立情報学研究所 コンテンツ科学研究系

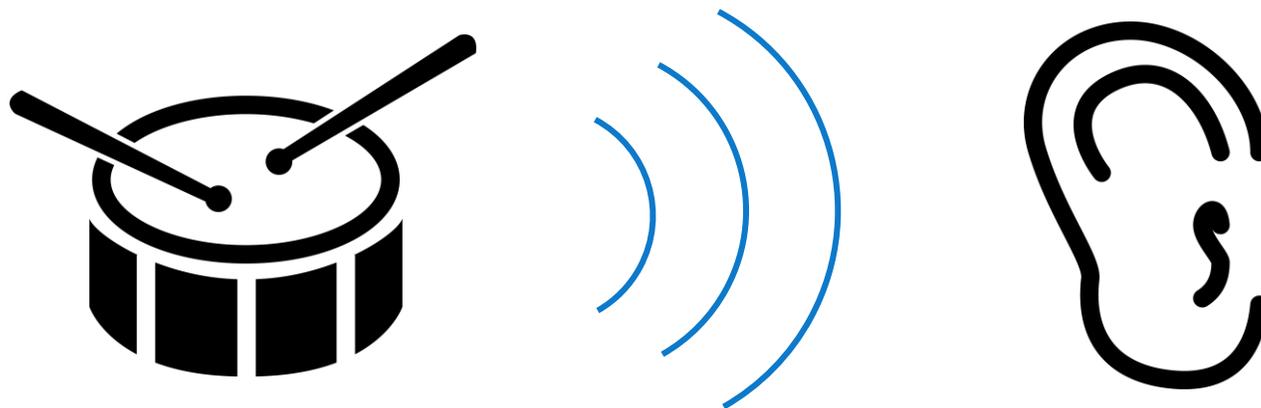


NII S.Koyama's Lab  
Audio Processing Research Group

# 音と情報学との関係

## 物理学としての音，情報学としての音

➤ 音とは，物体が振動することで生じる空気の振動が波として伝わるもの

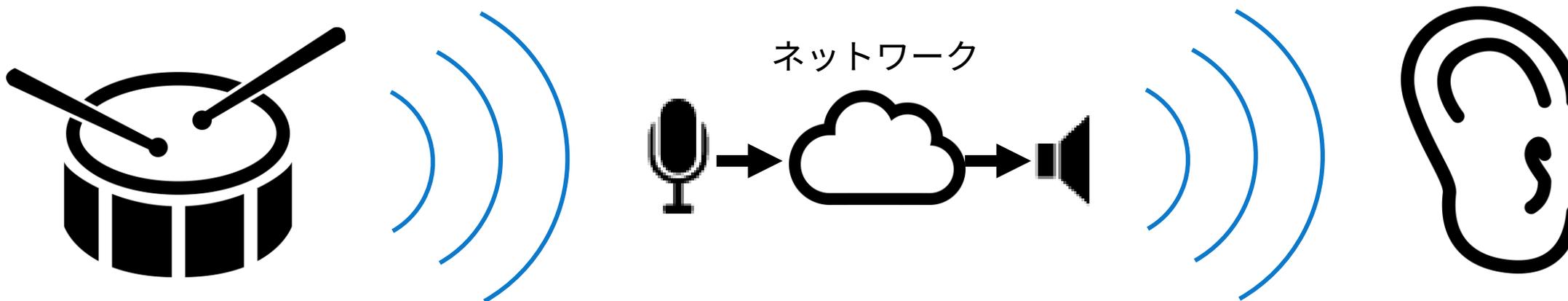


➡ **音の物理的な性質を扱う分野は，19世紀頃に音響学として確立された**

# 音と情報学との関係

## 物理学としての音，情報学としての音

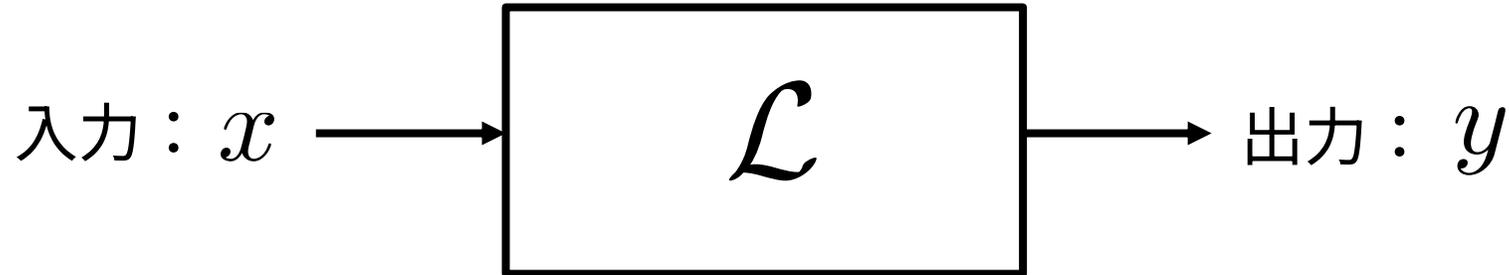
➤ 音を記録，伝送，再生するための情報処理技術



- 音をデジタルデータとして記録，伝送，再生するための技術は，情報学の黎明期から存在する分野の一つ
- 現在でも信号処理／機械学習技術として発展を続けている

# 音の信号処理／機械学習

- 信号処理とは，音や画像などの信号を分析，編集，合成する技術

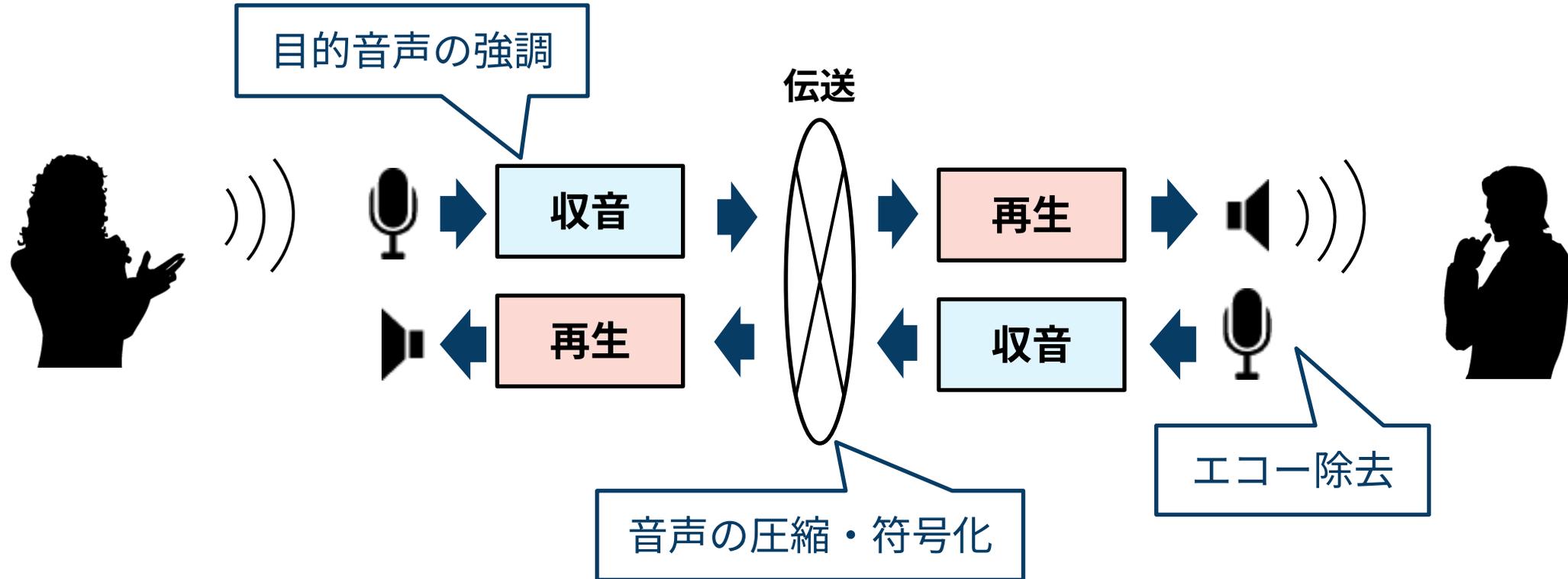


入力  $x$  に対して所望の出力  $y$  を得るために，システム（写像）  $\mathcal{L}$  をデザインする

➡ 連続信号は関数，離散信号はベクトルと捉えられ，信号処理／機械学習技術は広い数理的基盤の上に成り立っている

# 音の信号処理／機械学習

## ➤ 遠隔コミュニケーションシステムにおける音の信号処理技術

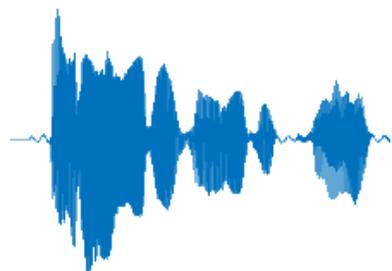


➡ 身近なところで多くの音の信号処理／機械学習技術が使われている！

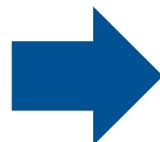
# 音の時間信号から時空間信号へ

4次元関数である時空間信号としての音を扱うには？

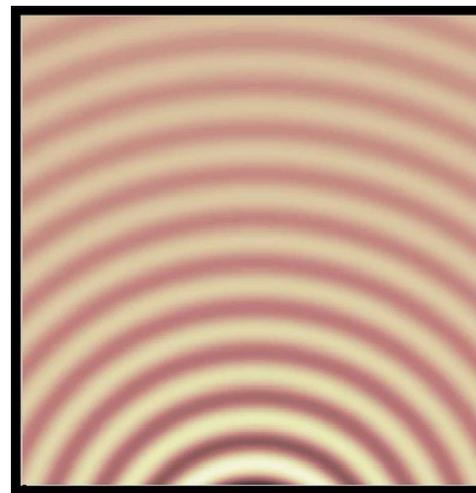
時間信号



$U(t)$



時空間信号



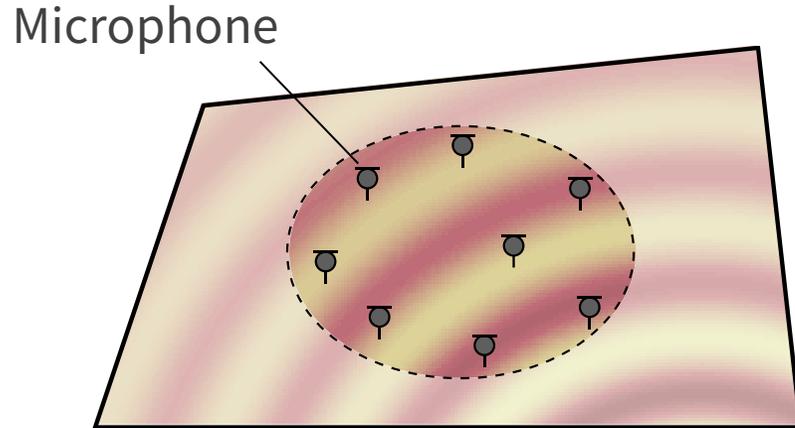
$U(x, y, z, t)$



- デバイスの小型化・低コスト化，計算機の高性能化などによるシステムとしての実現可能性の高まり
- 3次元の音空間 (=音場) の分析・合成に基づく様々な技術に応用できる可能性

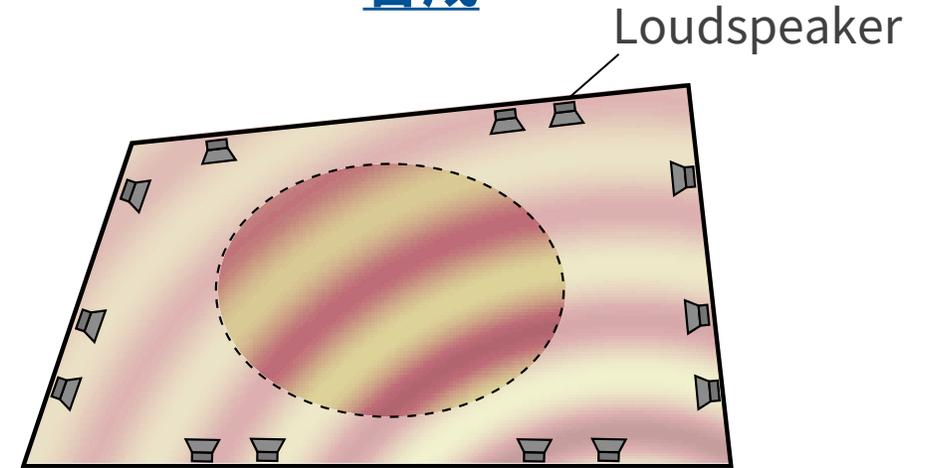
# 音場の分析と合成とは？

## 分析



複数のマイクを用いて対象領域内の音場を推定

## 合成

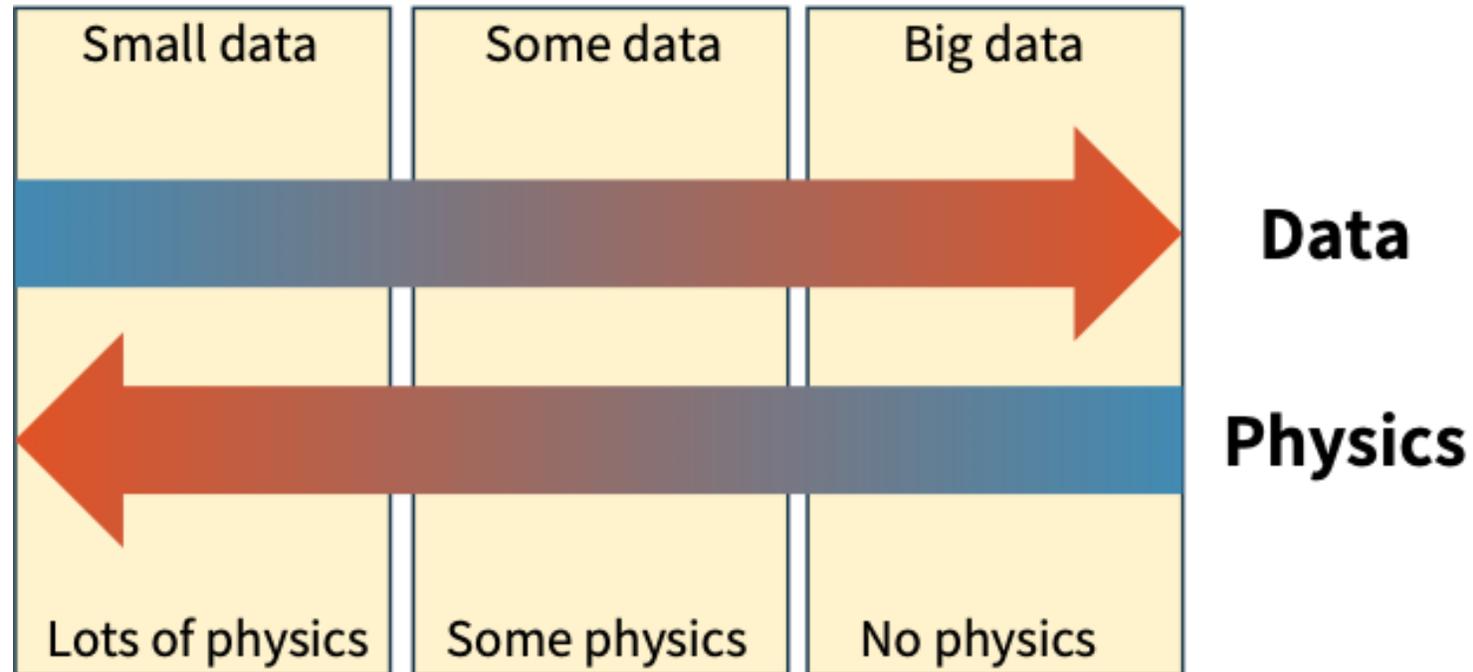


複数のスピーカを用いて対象領域内に所望の音場を合成

音場の物理的性質を取り入れた信号処理／機械学習による  
音場の分析・合成技術

# 物理的性質に基づく信号処理／機械学習

一般的な信号処理／機械学習も適用できるが、大量のデータを要する

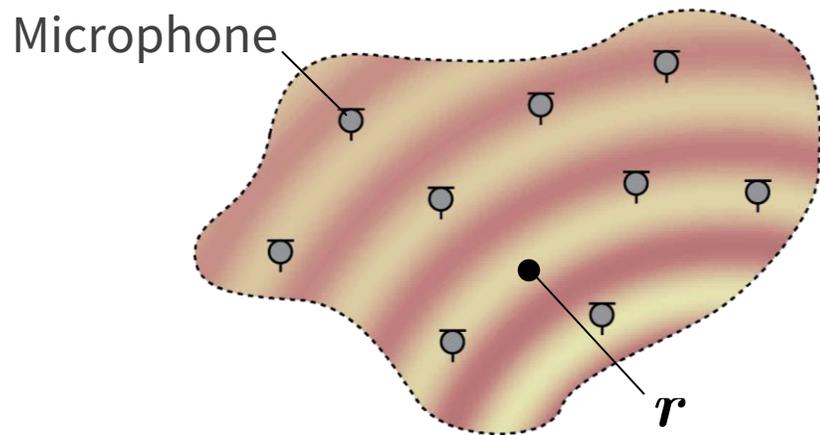


➡ 物理的性質を機械学習に組み入れる技術は、最近ではPhysics-Informed Machine Learning (PIML)と呼ばれる

# 音場の分析・合成におけるPIML

どのような音場の物理的性質を取り入れられる？

➡ 最も効果的な物理的性質の一つが音場の支配方程式



➤ 時間領域：波動方程式

$$\left( \Delta_{\mathbf{r}} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) U(\mathbf{r}, t) = 0$$

➤ 周波数領域：ヘルムホルツ方程式

$$(\Delta_{\mathbf{r}} + k^2) u(\mathbf{r}, \omega) = 0$$

推定対象の関数は、波動方程式／ヘルムホルツ方程式を満たすはず

# 音場のカーネル補間法

## ヘルムホルツ方程式制約下でのカーネルリッジ回帰

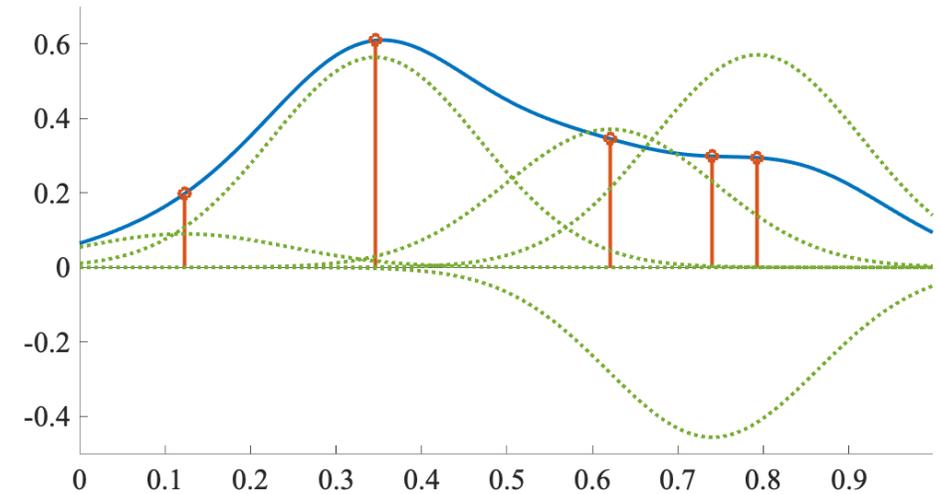
### ➤ カーネルリッジ回帰

$$\begin{aligned}\hat{u}(\mathbf{r}, \omega) &= \sum_{m=1}^M \alpha_m \kappa(\mathbf{r}, \mathbf{r}_m) \\ &= \sum_{m=1}^M [(\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{s}]_m \kappa(\mathbf{r}, \mathbf{r}_m)\end{aligned}$$

カーネル関数

観測値

カーネル関数の行列



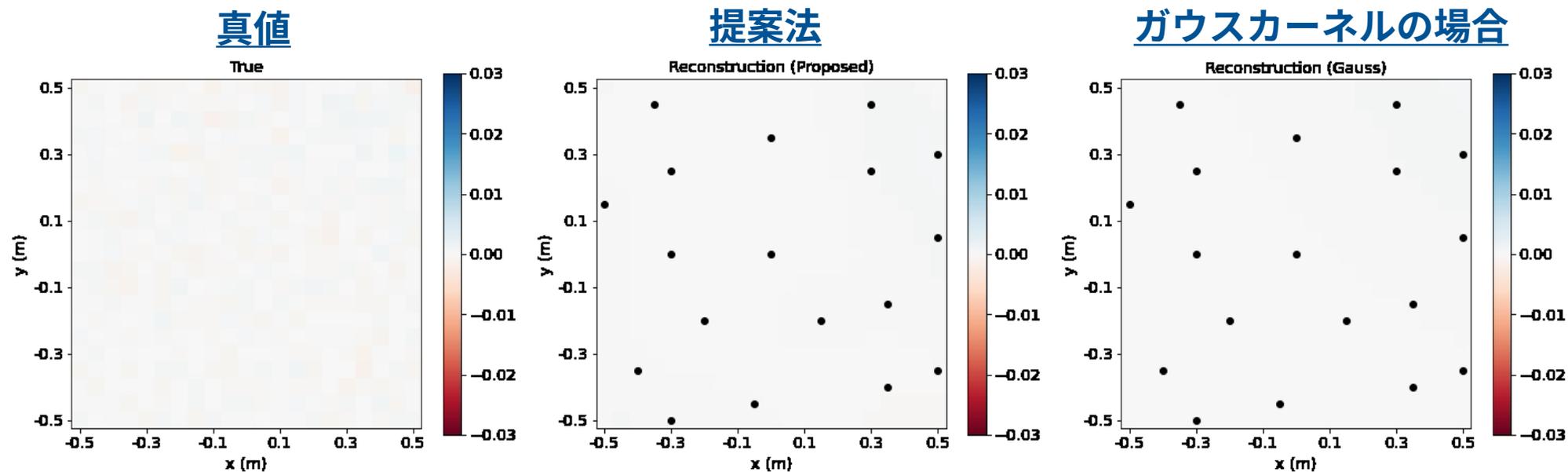
### ➤ ヘルムホルツ方程式の解に制約するカーネル関数

$$\kappa(\mathbf{r}_1, \mathbf{r}_2) = j_0 \left( \sqrt{(j\beta\xi - k\mathbf{r}_{12})^\top (j\beta\xi - k\mathbf{r}_{12})} \right) \quad \left[ \mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1 \right]$$

音源方向に関する事前情報

# 音場のカーネル補間法

- 実環境での測定データを利用した実験結果
  - 18個のマイク信号から単スピーカからのパルス信号を再構成



インパルス応答の測定システム

# 音場分析・合成の応用技術

音空間を対象とする領域での幅広い応用範囲



VR/AR音響

室内音響解析



信号強調

音場分析・合成の  
基礎技術



エリア收音・再生

アクティブ騒音抑圧



可視化／可聴化



# 音空間知覚とVR/AR音響

## ➤ VR/ARにおける空間音響再現の目的

- 3次元の音環境を提示することによる没入感と臨場感の向上
- 特定の方向からの音の提示によるインタラクションや状況認識の改善
- バーチャルな物体やイベントの聴覚的手がかりの提示による視覚的負荷の軽減



音響心理学，物理音響学，信号処理，機械学習などの分野における長年の研究



# 音空間知覚とVR/AR音響

## ➤ 人間はどのように空間的な音を知覚する？

➡ いくつかの聴覚的な手がかりを組み合わせることで、人間の心理空間における音像が作られる

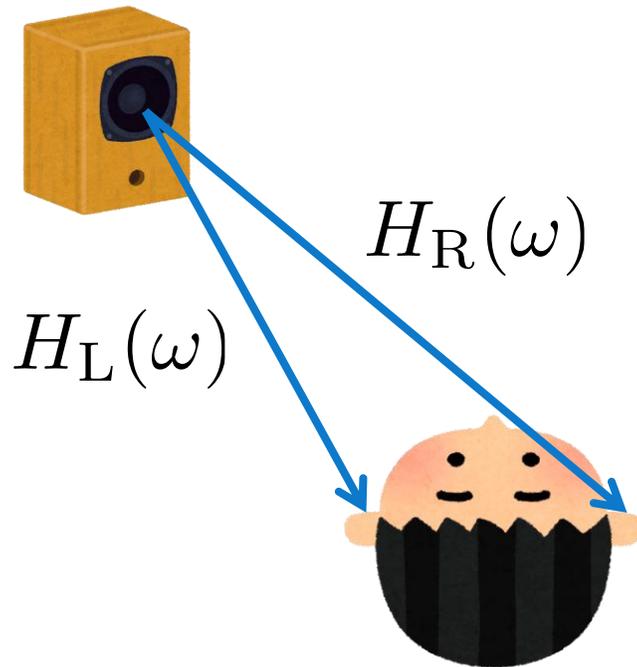
- 両耳間時間差 (Interaural Time Difference: ITD) : 両耳に到達する音の時間差
- 両耳間レベル差 (Interaural Level Difference: ILD) : 両耳に到達する音のレベル差
- スペクトルキュー : 耳介での反射によって生じる音のスペクトル (音色)

音の方向知覚がこれらの手がかりによって可能となる

# 音空間知覚とVR/AR音響

- **頭部伝達関数 (Head-Related Transfer Function: HRTF)** が音空間知覚の聴覚的手がかりを特徴付けるのに重要な役割を持つ
  - **HRTF**: 音源から両耳までの音の伝達特性

音源



音源から右耳までの  
HRTF

外耳道入口/鼓膜  
までの伝達特性

$$\text{HRTF}_R(\omega) = \frac{H_R(\omega)}{H_0(\omega)}$$

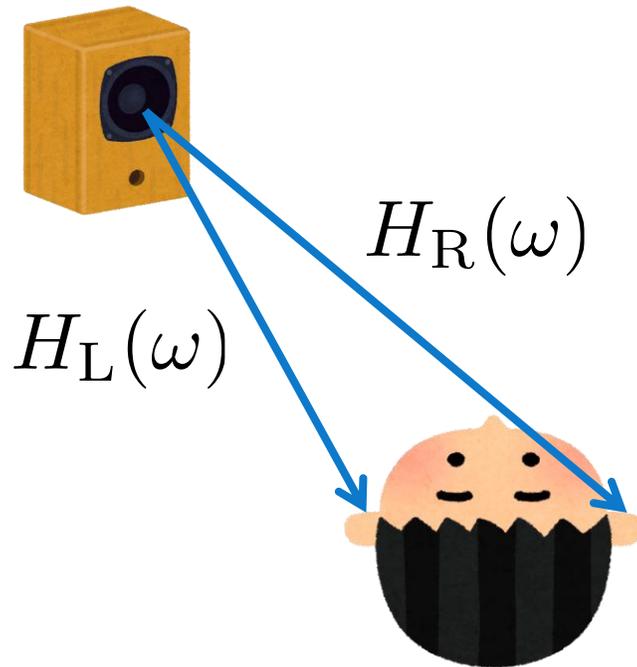
頭部がない状態での頭部  
中心までの伝達特性

J. Blauert, "Spatial Hearing: The Psychophysics of Humans Sound Localization," *The MIT Press*, 1996.

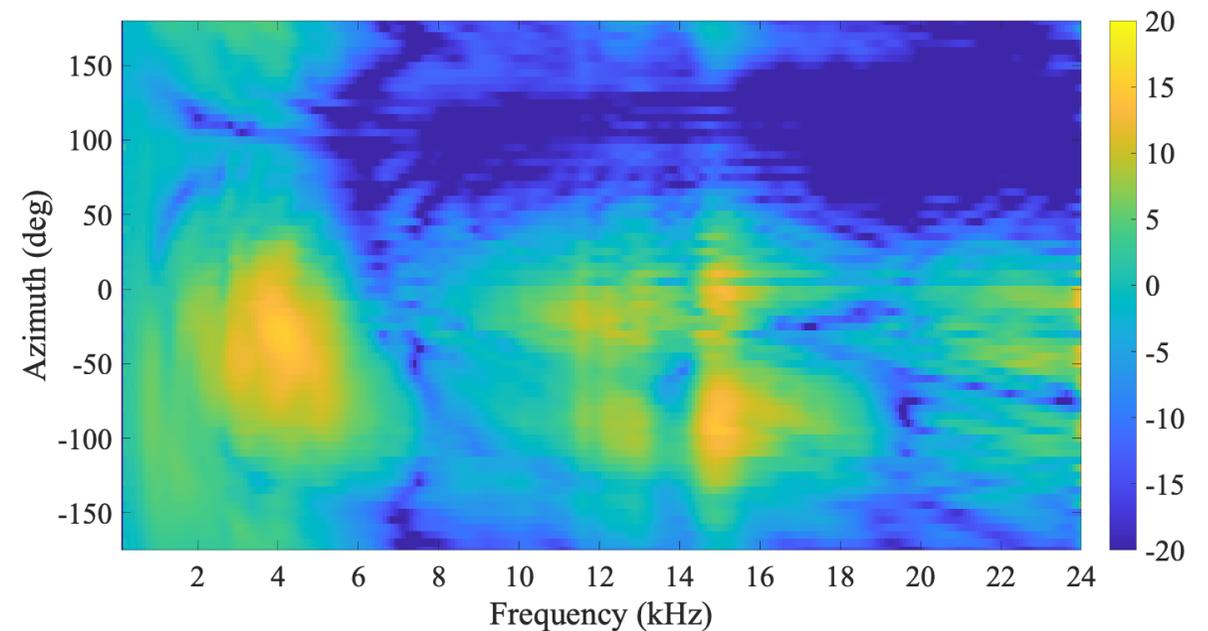
# 音空間知覚とVR/AR音響

- **頭部伝達関数 (Head-Related Transfer Function: HRTF)** が音空間知覚の聴覚的手がかりを特徴付けるのに重要な役割を持つ
  - **HRTF**: 音源から両耳までの音の伝達特性

音源



HRTFの振幅特性の例



J. Blauert, "Spatial Hearing: The Psychophysics of Humans Sound Localization," *The MIT Press*, 1996.

# 空間音響再現技術

## 空間音響再現技術の分類

### ➤ 聴覚特性に基づく技術

#### – ステレオフィニック／サラウンド方式

- Summing localizationと呼ばれる聴覚的な効果を利用して空間的な音像を提示

### ➤ 物理特性に基づく技術

#### – バイノーラル再現

- HRTFやその近似を用いて、両耳の音（バイノーラル音）を合成

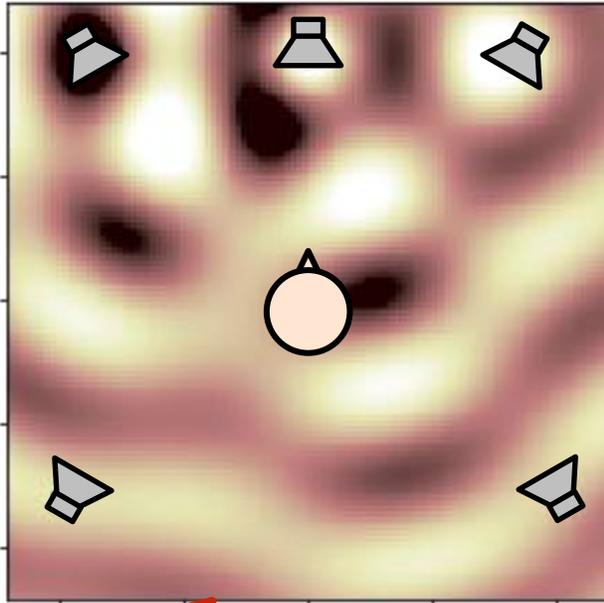
#### – 音場再現

- 複数のスピーカを用いて物理的な音場を合成

# 空間音響再現技術

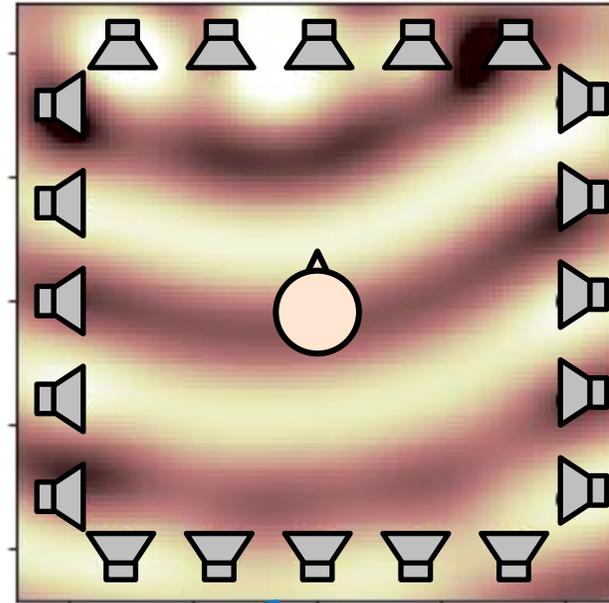
## 空間音響再現技術の長所・短所

### 聴覚特性に基づく技術 (サラウンド方式)



- 小規模なシステムで実現可能
- 受聴領域が小さい（スイートスポット）

### 物理特性に基づく技術 (音場再現)



- 広い受聴領域を実現可能
- 大規模なシステムが必要になる

# 音の自由聴点

## VR空間のバイノーラル再現から実環境のバイノーラル再現へ

### VR空間



音場のシミュレーション  
によって実現可能



### バイノーラル再現

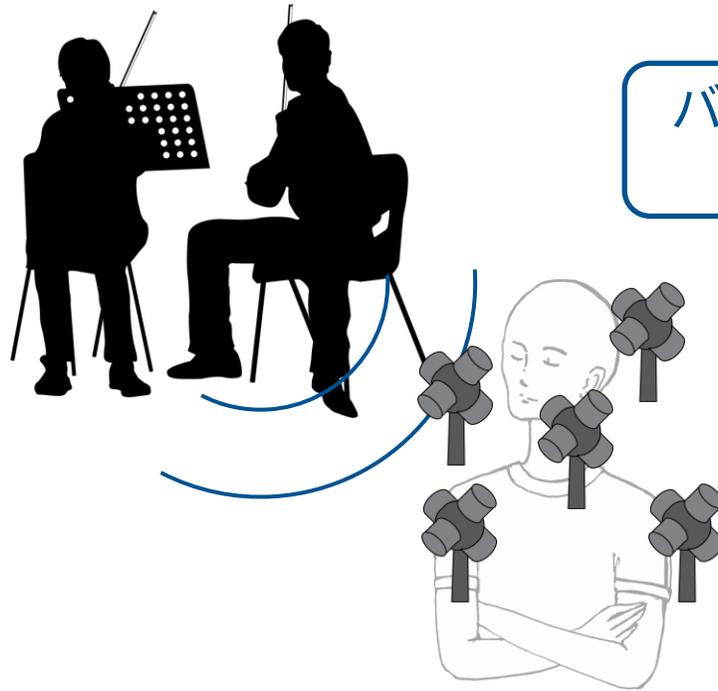


実環境の音を収録し、任意の位置でのバイノーラル音を再現する、  
音の自由聴点を実現するには？

# 音の自由聴点

## VR空間のバイノーラル再現から実環境のバイノーラル再現へ

### 複数マイクでの收音



バイノーラル音  
への変換



### バイノーラル再現

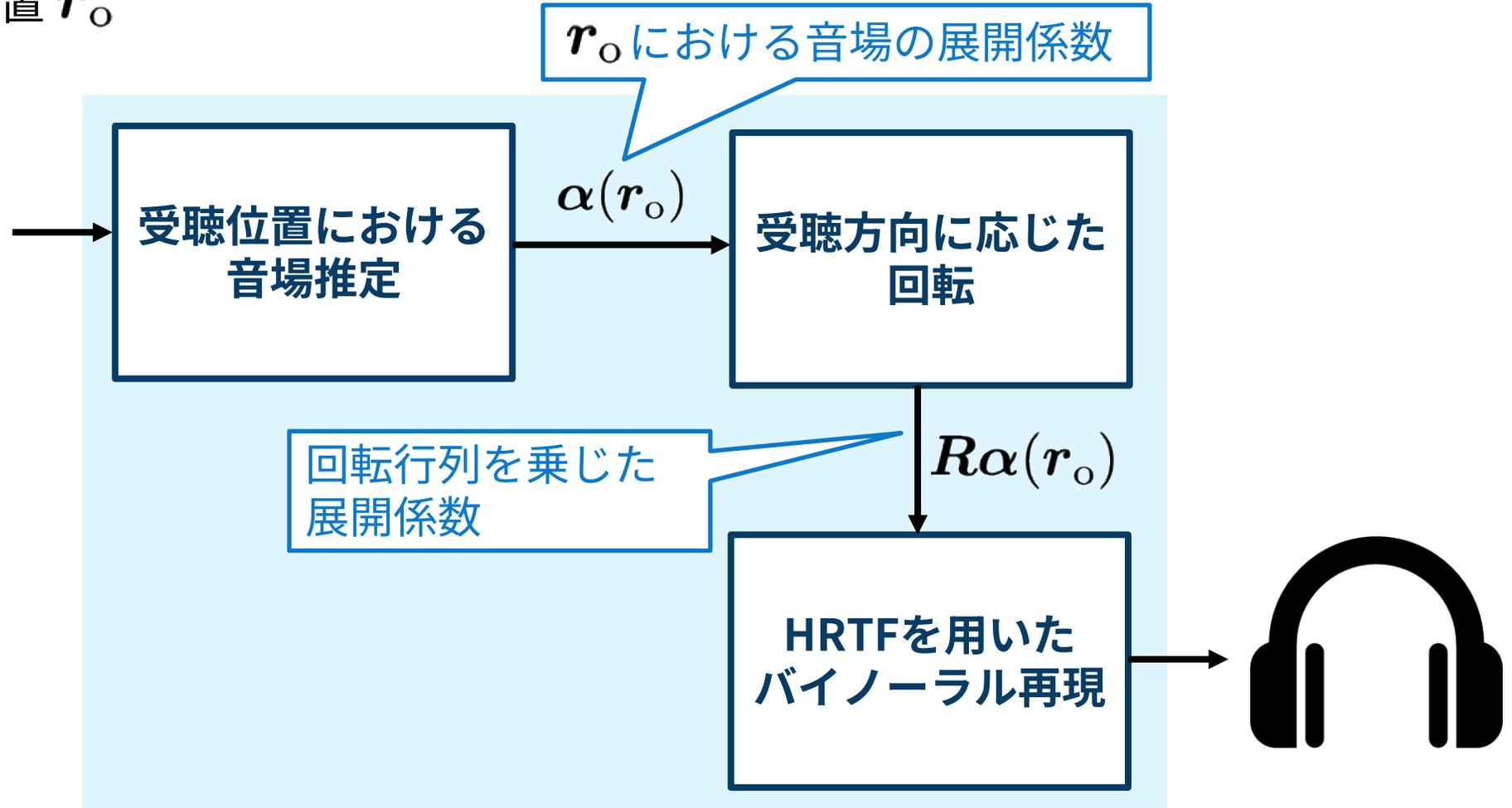
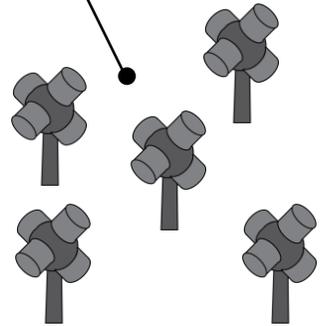


複数マイクで收音した信号から音場を推定することで、任意の受聴位置／方向におけるバイノーラル音を再現

N. Iijima, et al. "Binaural Rendering from Microphone Array Signals of Arbitrary Geometry," *J. Acoustical Society of America*, 2021.

# 音の自由聴点

選択した受聴位置  $r_0$



マイク收音信号に対するフィルタとして実装

# 音の自由聴点

- 複数の小規模マイクアレイと360度カメラを用いた収録システム



- 8つの小規模マイクアレイ
- 各アレイは8つの指向性マイクで構成
- 2つの360度カメラによる映像収録



Listening position: A (0 m, 0 m, 0 m) ▾

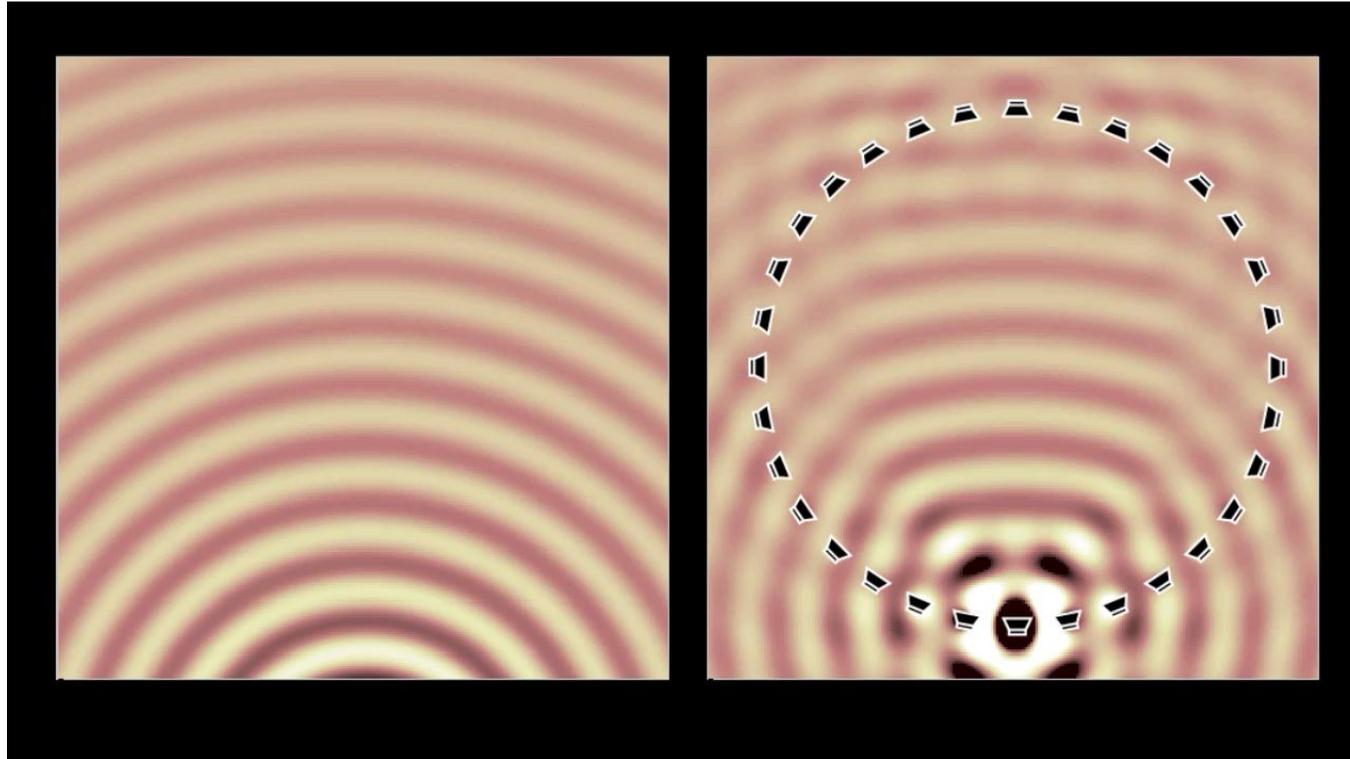
Pause Resume Restart

# Headphones Recommended



# 音場再現

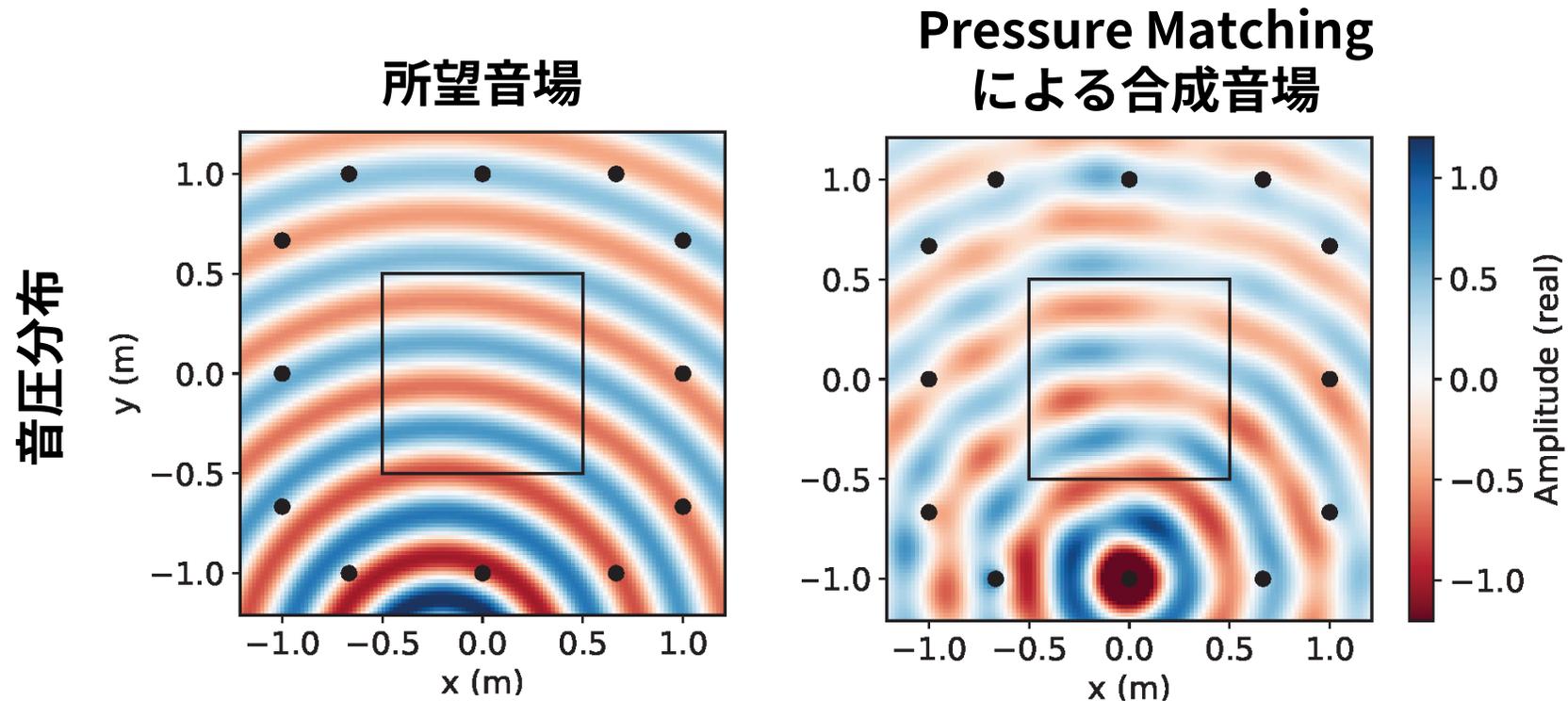
- 複数のスピーカを用いて物理的な音場を合成



**➡ 単純にスピーカからその位置の音圧を出力しても音場を再現することはできず、所望の音場を合成するための一種の逆問題を解くことが必要**

# 音場再現における空間エイリアシングと知覚的精度向上

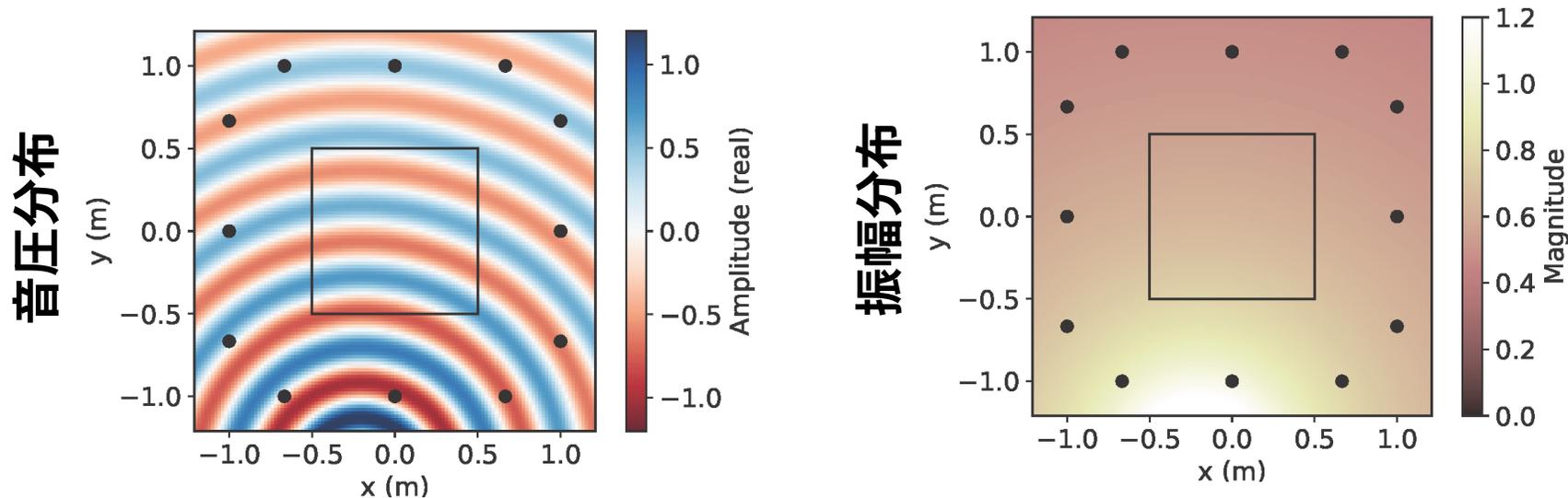
- 音場再現では高い周波数で精度が低下する空間エイリアシングが問題となる
  - 例：12個のスピーカで800Hzの音場を合成した場合



➡ 音像定位精度の低下と音源信号の音質劣化を引き起こす

# 音場再現における空間エイリアシングと知覚的精度向上

高い周波数において位相分布が音像定位に大きく影響しないことを利用し、  
振幅分布のみを合成することによる知覚的精度向上

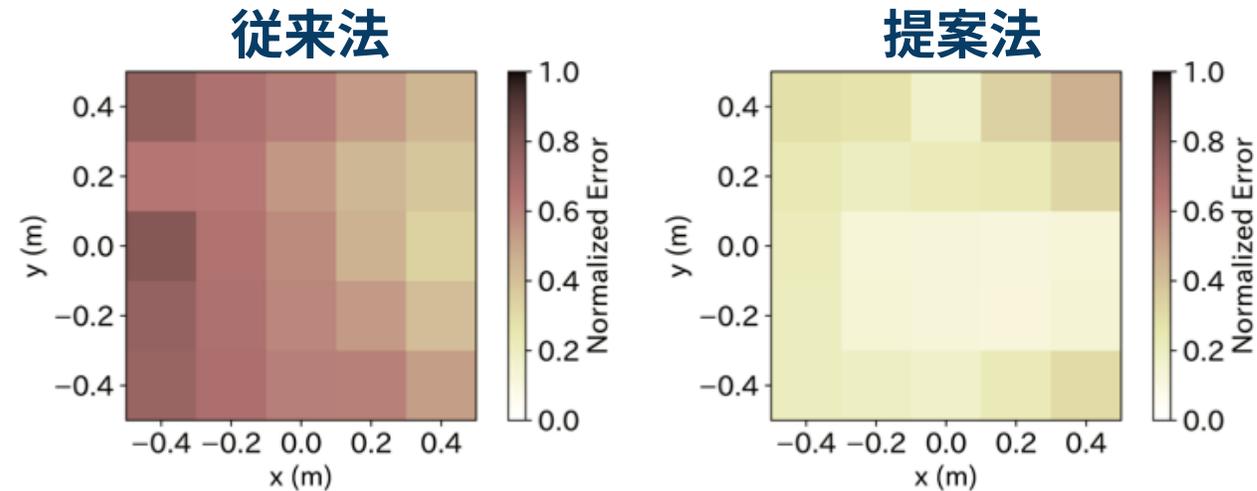


- 1500Hz以上ではITDに比べてILDが音像定位の支配的な手がかり
- 振幅応答をできるだけフラットにすることで音質劣化を改善

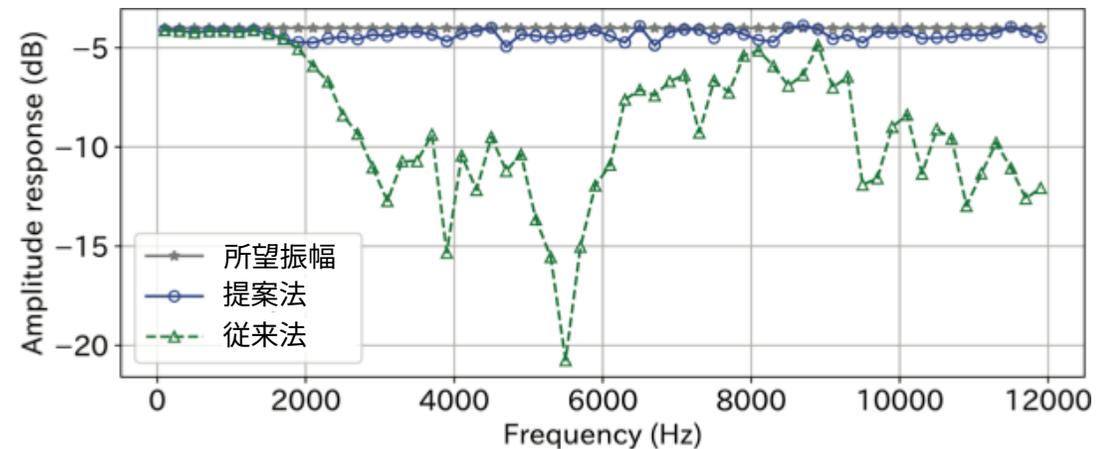
➡ **高い周波数帯域では位相分布を任意として、振幅分布のみを合成する**

# 音場再現における空間エイリアシングと知覚的精度向上

- 各受聴位置でのILDの誤差平均



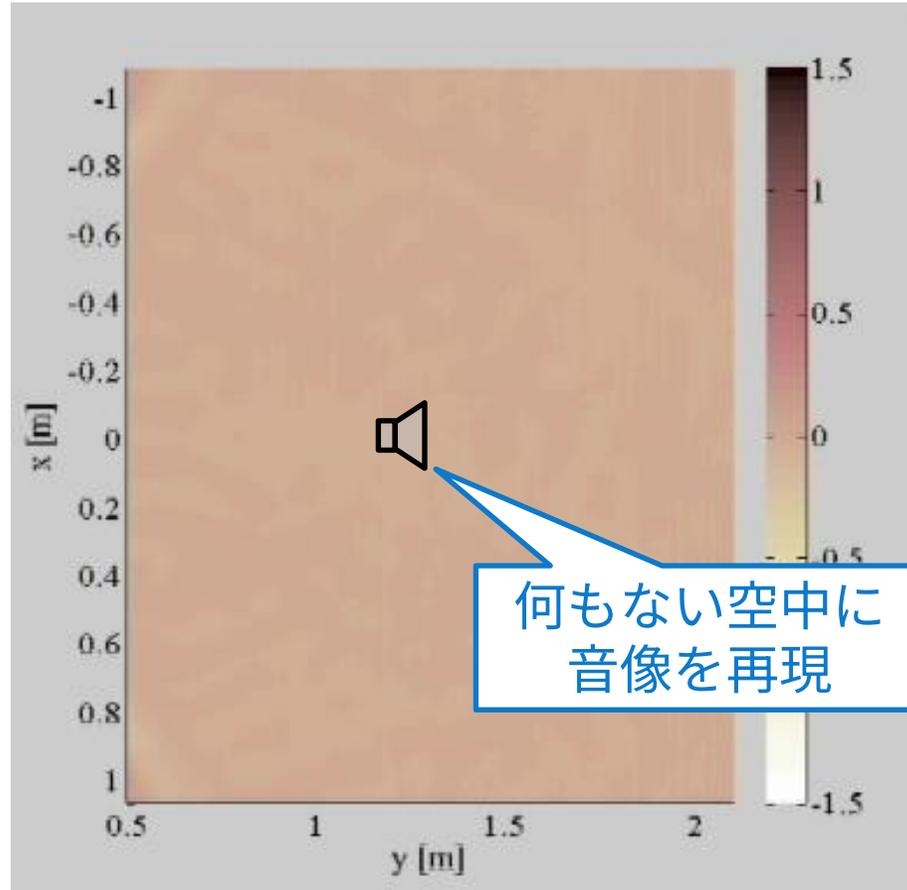
- 対象領域中心での振幅応答



K.Kimura, et al. "Perceptual Quality Enhancement of Sound Field Synthesis Based on Combination of Pressure and Amplitude Matching," *IEEE WASPAA*, 2023.

# 音のホログラフィ

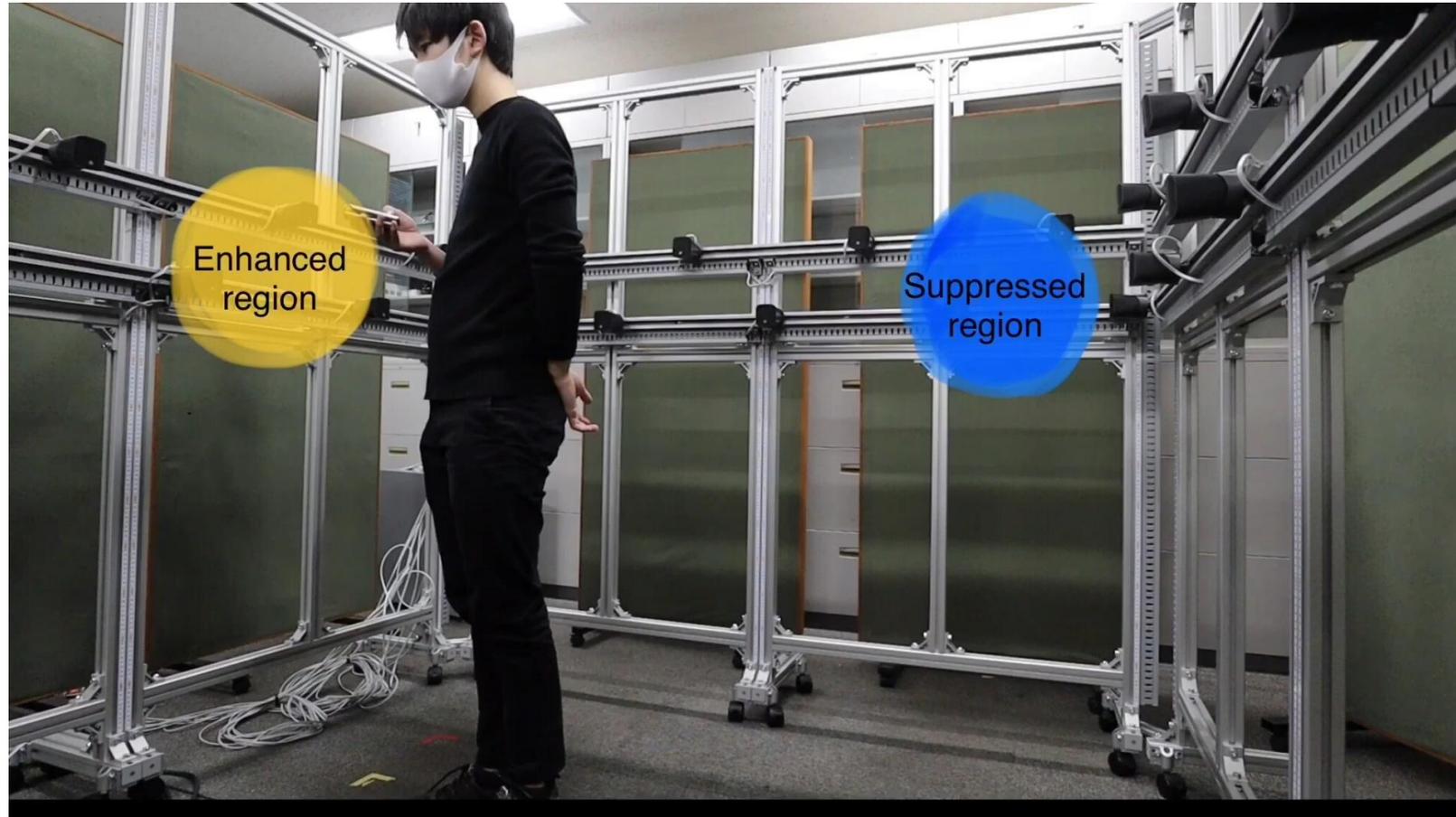
## ➤ 複数スピーカでの空中音像の再現



S.Koyama, et al. "Reproducing Virtual Sound Sources in Front of a Loudspeaker Array Using Inverse Wave Propagation," *IEEE Trans. ASLP*, 2012.

# エリア再生

## ➤ 複数スピーカでの音のゾーニング



T. Abe, et al. "Amplitude Matching for Multizone Sound Field Control," *IEEE/ACM Trans. ASLP*, 2023.

# 空間でのアクティブ騒音抑圧

## ➤ アクティブ騒音抑圧 (Active Noise Control: ANC)

- 外部から到来する騒音を誤差マイクでモニタリングし、スピーカ信号で打ち消す

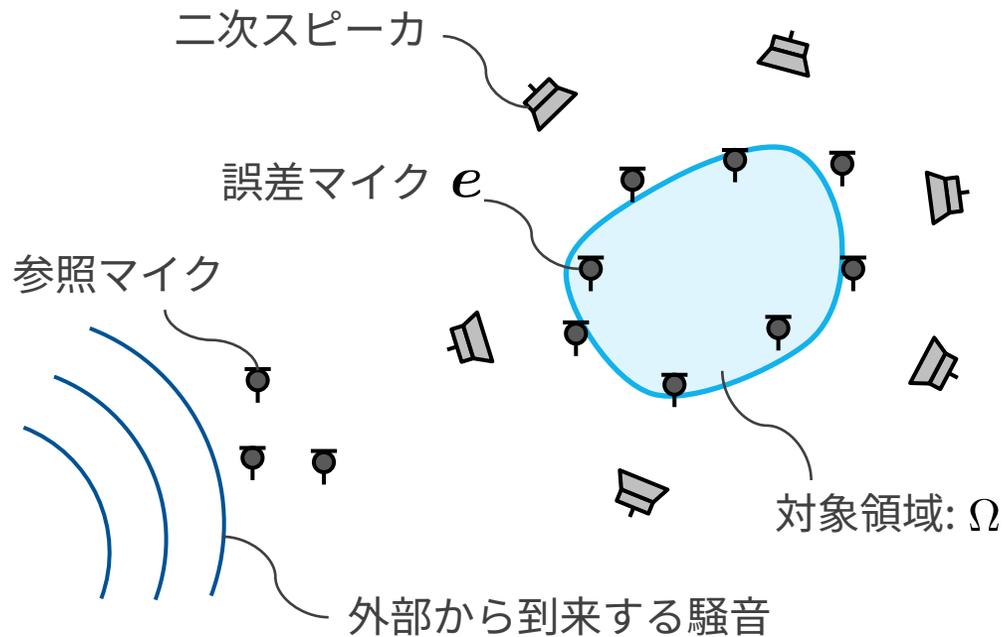


## ➡ 広い領域を静音化する空間でのアクティブ騒音抑圧を実現するには？



# 空間でのアクティブ騒音抑圧

## 音場の推定に基づく3次元空間でのANC



### ➤ 従来法のコスト関数

$$\mathcal{L} = \|e\|^2 \quad : \text{誤差マイクのパワー}$$

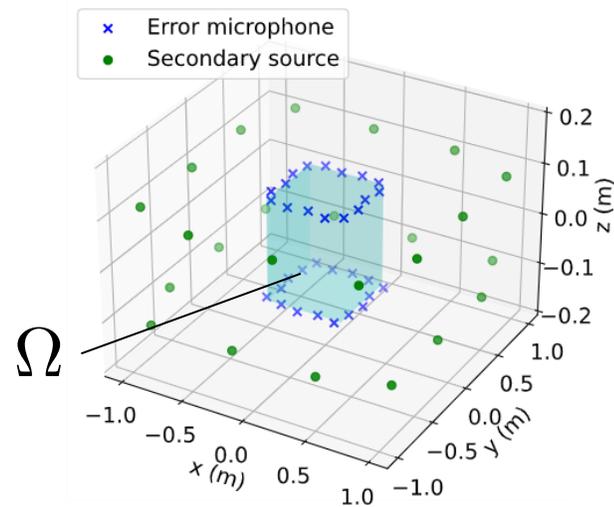
### ➤ 提案法のコスト関数

$$\mathcal{L} = \int_{\Omega} |u(\mathbf{r})|^2 d\mathbf{r} \quad : \text{領域的な騒音パワー}$$

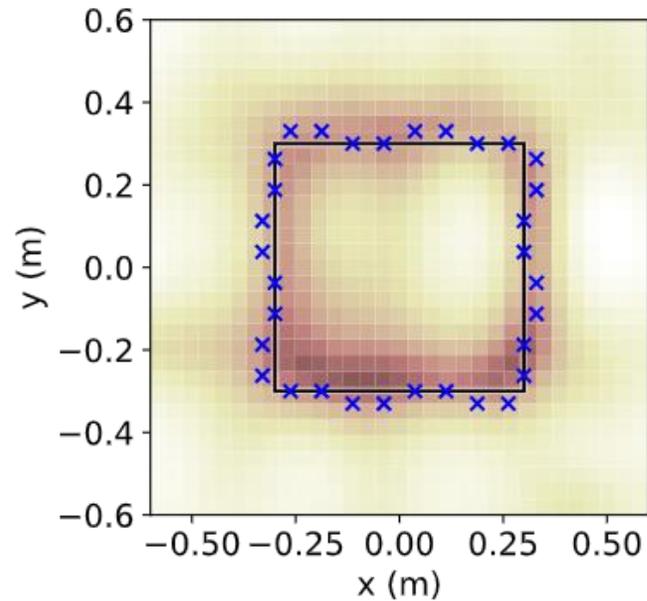
➡ 誤差マイク信号から対象領域内の音場を予測することによる空間ANC

# 空間でのアクティブ騒音抑圧

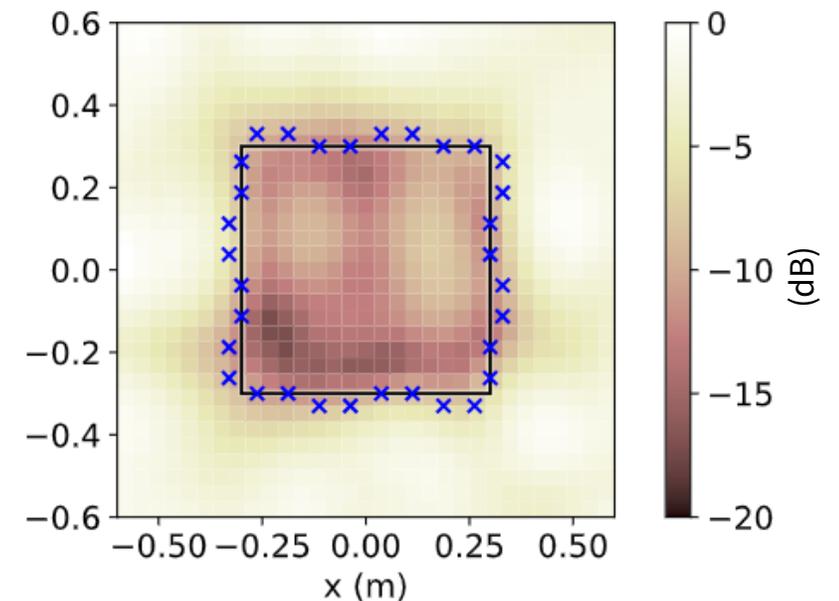
- 帯域制限ノイズ (500-800Hz), 残響時間 ( $T_{60}$ ) 240msの場合の実験結果



従来法: -6.2 dB



提案法: -10.5 dB



S.Koyama, et al. "Spatial Active Noise Control Based on Kernel Interpolation of Sound Field," *IEEE/ACM Trans. ASLP*, 2021.

音場の予測に基づく提案法によって、領域的な騒音抑圧を実現

# 音場分析・合成とその応用技術

## 3次元音空間を自在に操る基礎技術と応用技術



VR/AR音響

室内音響解析

信号強調



音場分析・合成の  
基礎技術



エリア收音・再生

アクティブ騒音抑圧

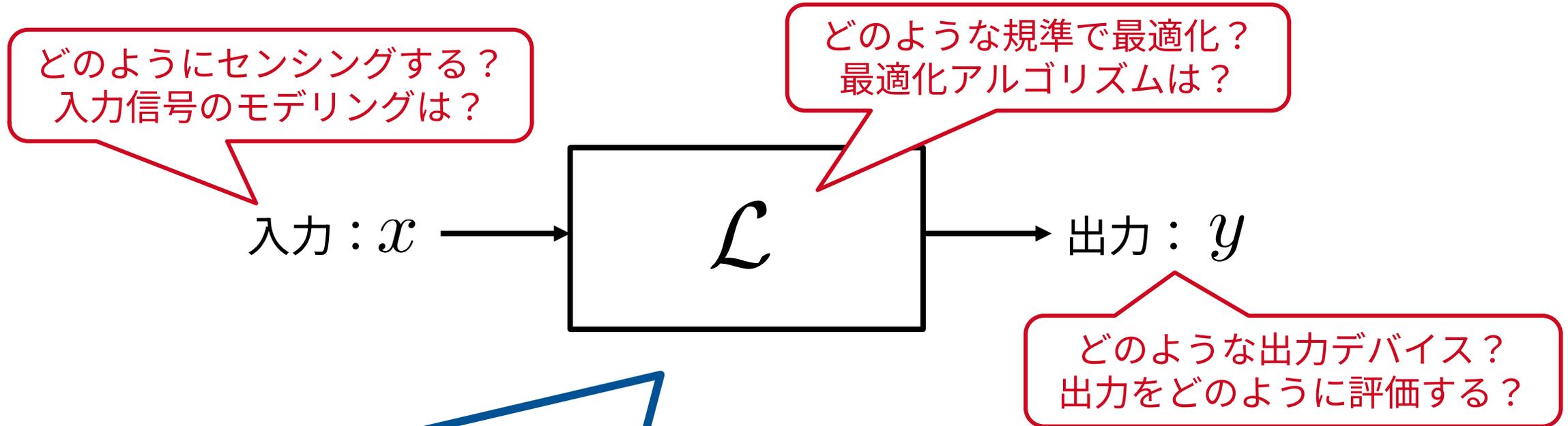


可視化／可聴化



# 音メディア処理研究への誘い

- 音メディア処理は身近な社会課題や文化・芸術に関わる技術分野であり，広範囲な分野の上に成り立っている



物理音響学，統計モデリング，機械学習，聴覚などを総合的に用いて，音声・音響・音楽信号の性質をうまく取り入れることで所望のシステムを実現