

$$-\frac{\partial \text{KL}(\mathbf{y}_{(\text{ICAL})}(t))}{\partial \mathbf{w}_{(\text{ICAL})}(n)} \cdot \mathbf{W}_{(\text{ICAL})}(z^{-1})^T \mathbf{W}_{(\text{ICAL})}(z)$$

$$= \sum_{d=0}^{D-1} \left\{ \left(I\delta(n+d) - \left\langle \varphi\left(x(t-\frac{D}{2}) - \sum_{l=1}^{L-1} \mathbf{y}_{(\text{ICAL})}(t-l)\right) \cdot \left(x(t-n+d-\frac{D}{2}) - \sum_{l=1}^{L-1} \mathbf{y}_{(\text{ICAL})}(t-l-n+d)\right) \right\rangle_t \right) \cdot \left(I\delta(d-\frac{D}{2}) + \sum_{l=1}^{L-1} \mathbf{w}_{(\text{ICAL})}(d-l) \right) \right\}$$

システム情報 猿渡・小山研究室 (創造情報 猿渡研究室)の紹介

$$f(x) = x^{k-1} \frac{e^{-x/\theta}}{\Gamma(k)\theta^k}$$

東京大学大学院・情報理工学系研究科
システム情報学・創造情報学専攻 猿渡・小山研

(2021年5月)

猿渡・小山研(システム情報第一研究室)

教授
猿渡洋



講師
小山翔一



助教
高道慎之介



特任助教
中村友彦



特任助教
齋藤佑樹



講師(協力)
郡山知樹



専門分野

- ・教師無し最適化
- ・統計・機械学習
- 論的信号処理

専門分野

- ・音場再生・伝送
- (音ホログラフ)
- ・スパース最適化

専門分野

- ・統計的音声合成
- ・声質変換
- ・深層学習(DNN)

専門分野

- ・音楽情報処理
- ・多重解像度解析
- ・DNN音源分離

専門分野

- ・統計的音声合成
- ・音声知覚モデリング
- ・人参加型機械学習

専門分野

- ・統計的音声合成
- ・DNN
- ・深層ガウス過程

協力教員 北村大地先生
特任研究員 高宗さん
秘書 丹治さん

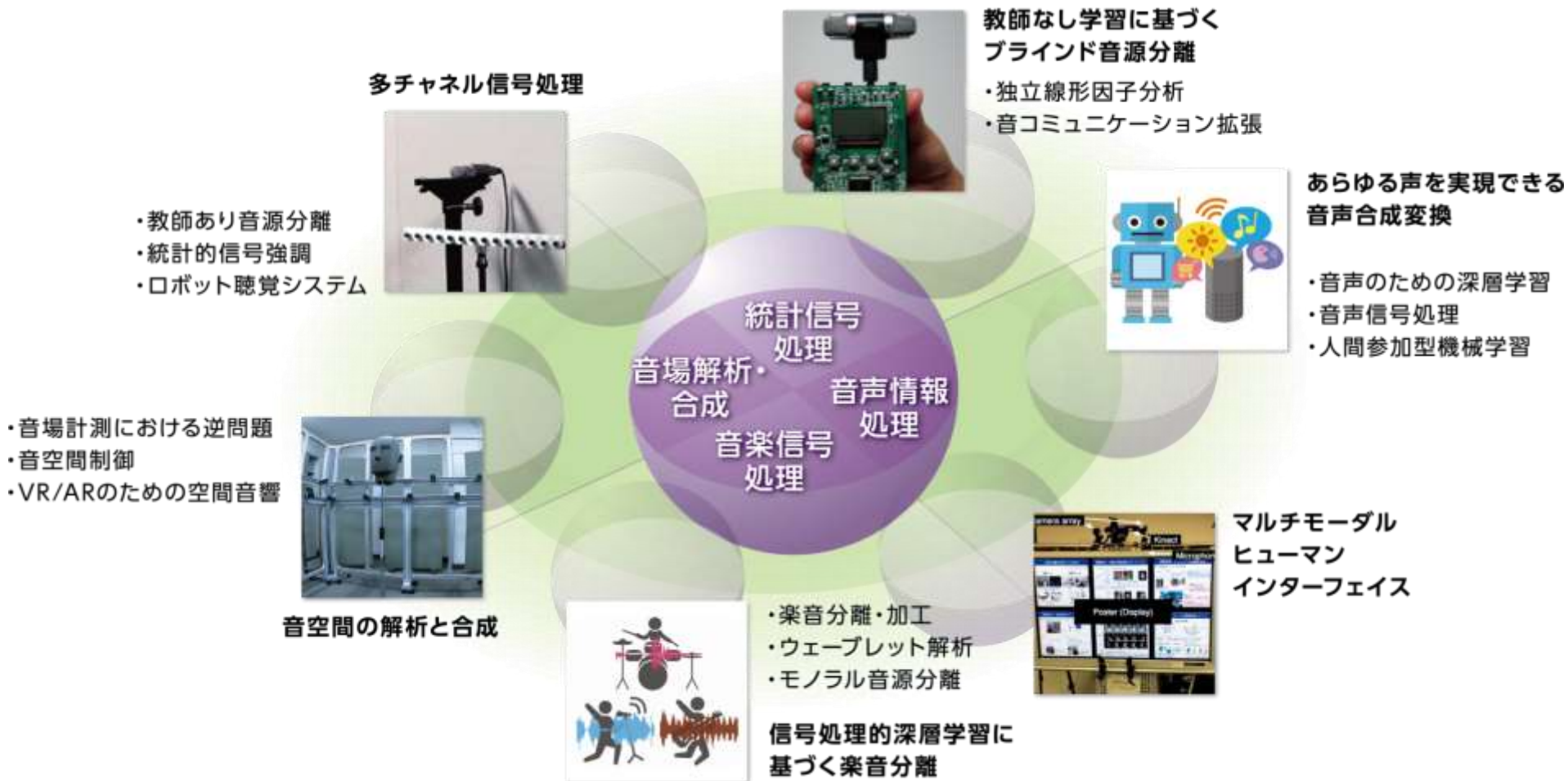
博士課程学生4名
修士課程学生10+10名

※郡山先生は郡山研として指導



研究俯瞰図

- 音声・音響・音楽メディアに関する信号処理・情報処理
- ヒューマンインターフェイス・コミュニケーションシステムの構築
- 統計的・機械学習論的信号処理、数理最適化問題等を研究



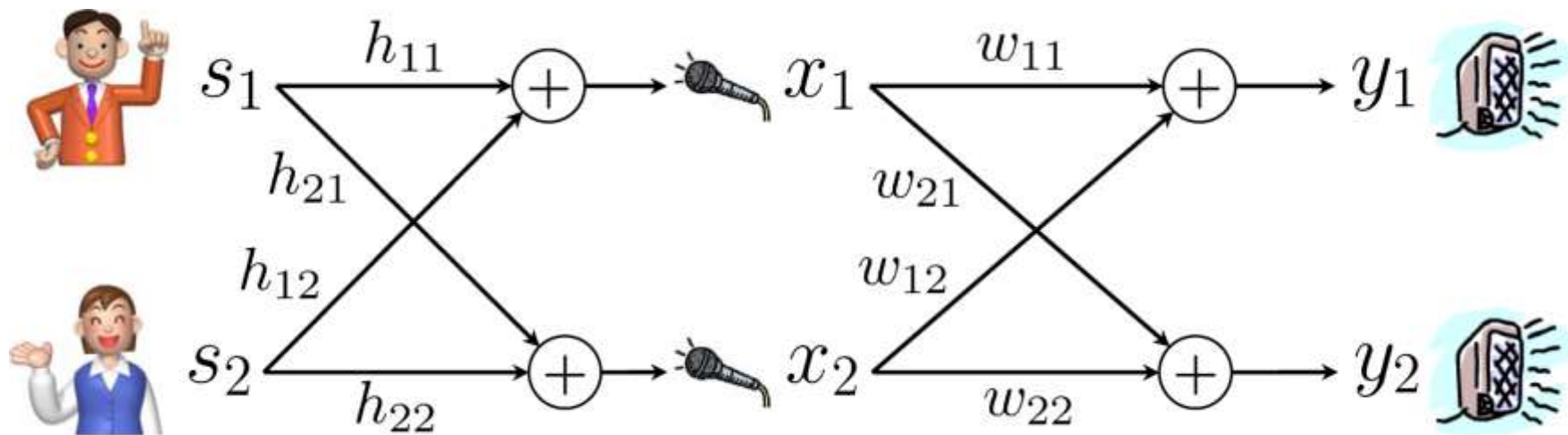
研究紹介1. ブラインド音源分離

- 音の方向・声質・音量など、事前に何も分かっていなくても、瞬時に音を「聞き分ける」ことの出来るシステムを目指す。
- 独自に開発した高速独立成分分析(ICA)、独立低ランク行列分析(ILRMA)という教師無し数理最適化アルゴリズムに基づいて、音を統計的に独立な成分に分解することにより、別々の音声信号を見つける。

スモールデータ
教師無し最適化
低ランクモデル

ブラインド音源分離(B_lindS_ourceS_eparation): 聞き分けるAI

- 混ざり合った信号 x_1, x_2 から元の信号を取り出す
- どのように混ざったかに関する情報 H は利用できない
- 事前トレーニング出来ない⇒ビッグデータではなく**スモールデータ**



実は上記は**2つのことを同時に推定**している

- [空間] 統計的に独立な音源の分類問題(分離行列 W の推定)
- [音源] 各音源が属する確率分布 $p(y)$ や構造の推定問題

上記を閉形式で解く方法は存在せず凸問題でもない⇒**大変困難!**

ILRMA: 音源の独立性と低ランク性に着目したBSS

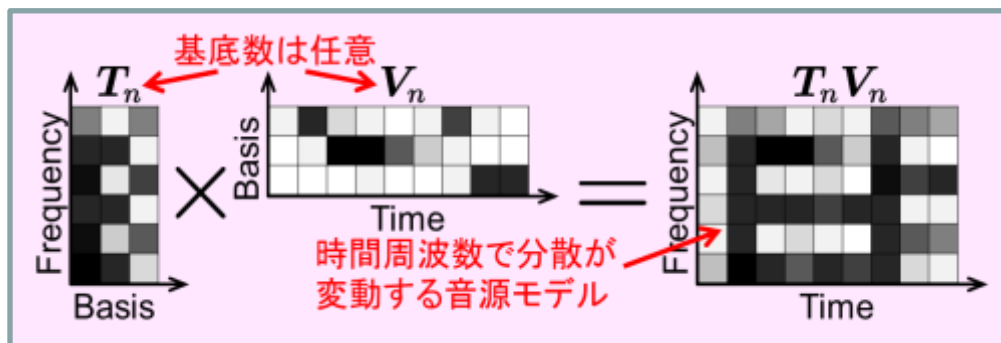
[IEEE Trans. ASLP 2016、IEEE SPS論文賞・ASJ粟屋賞・JSPS育志賞]

- ILRMAのコスト(対数尤度)関数→これを最小化

$$\mathcal{J} = \sum_{i,j} \left[\underbrace{\sum_m \log \sum_k z_{mk} t_{ik} v_{kj}}_{\text{音源の低ランク性コスト関数}} + \underbrace{\sum_m \frac{|y_{ij,m}|^2}{\sum_k z_{mk} t_{ik} v_{kj}}}_{\text{音源の独立性コスト関数}} - 2 \log |\det \mathbf{W}_i| \right]$$

音源の低ランク性コスト関数
(音源NMFモデルの推定に寄与)

音源の独立性コスト関数
(空間モデル $\mathbf{W}$ の推定に寄与)



$$p(\mathbf{y}) = p(y_1) p(y_2) \dots p(y_m)$$

となる $\mathbf{W}$ を推定

両者を交互にMajorization-Minimizationアルゴリズムで反復最小化

- ✓ コスト値の単調減少性を保証(勾配法には無い特徴)
- ✓ 高速かつ安定な求解法を実現(従来の多入力NMFと比較して2ケタ速い)

モデルの多様化・数理解法の開拓

● 音源生成モデルの多様化 IEEE SPS Tokyo Joint Chapter 学生賞!

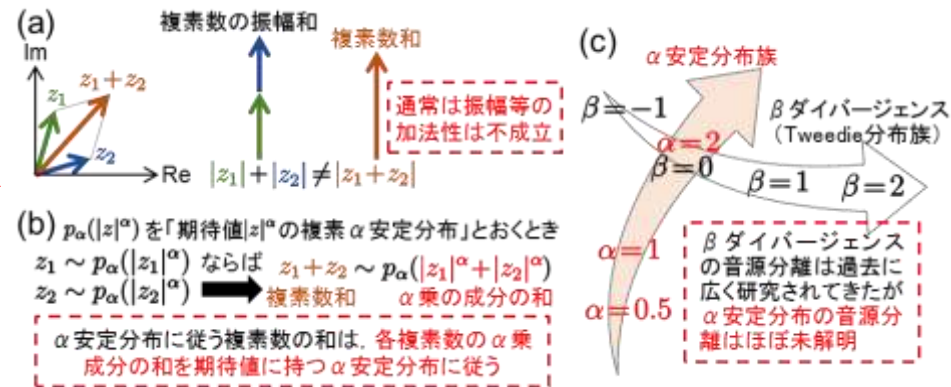
- 複素波形重ね合わせと整合する α 安定分布の導入

⇒ **t-ILRMA** [EURASIP-JASP2018]

- 複素球状ポアソン分布の導入

⇒ **$\beta=1$ -divergence最小化ILRMA**

[IEICE Trans. 2018]



● 座標降下法におけるバリエーション

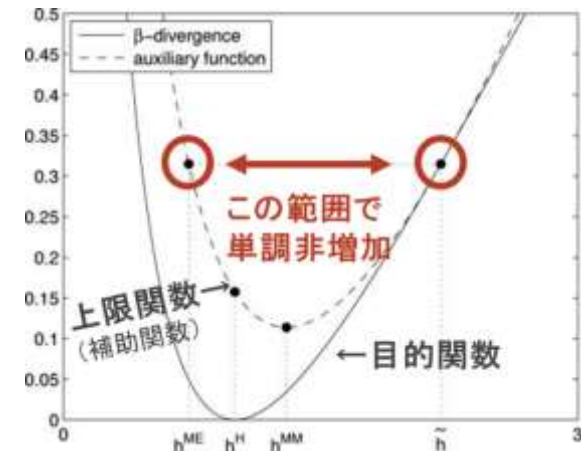
- **パラメトリックMajorization-Equalizationアルゴリズム**による音源・空間最適化の「バランス」化

[CAMSAP2017], [IEEE Trans. ASLP2019]

● 深層学習 (DNN) との融合

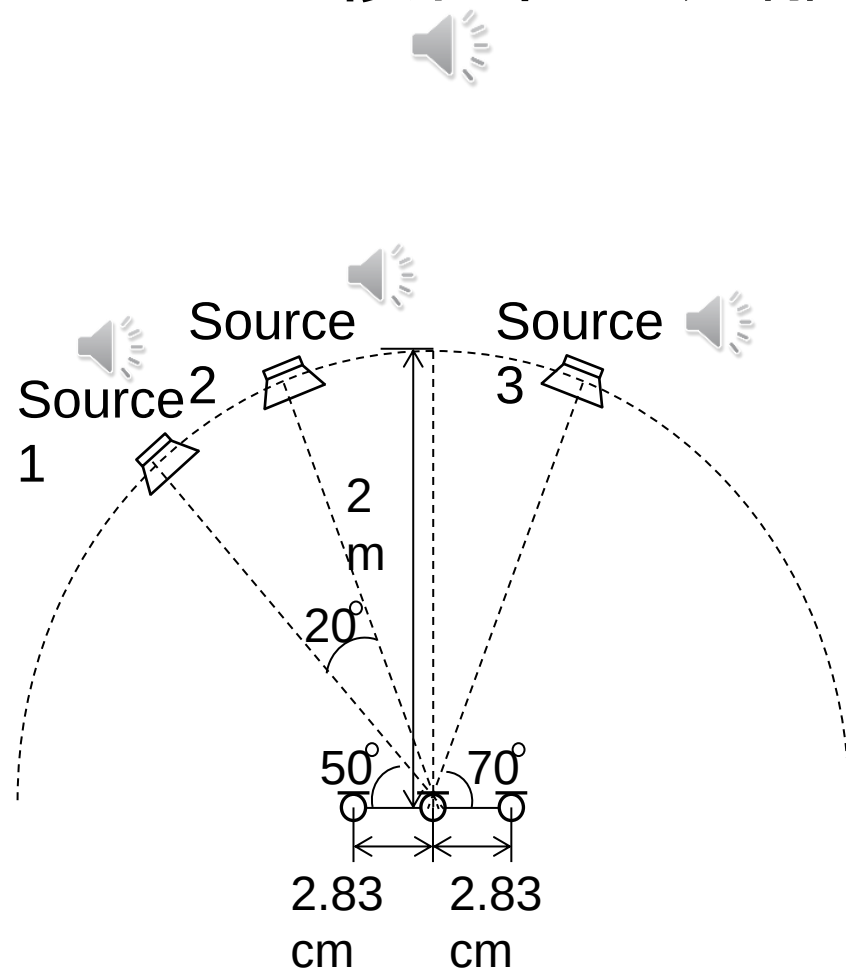
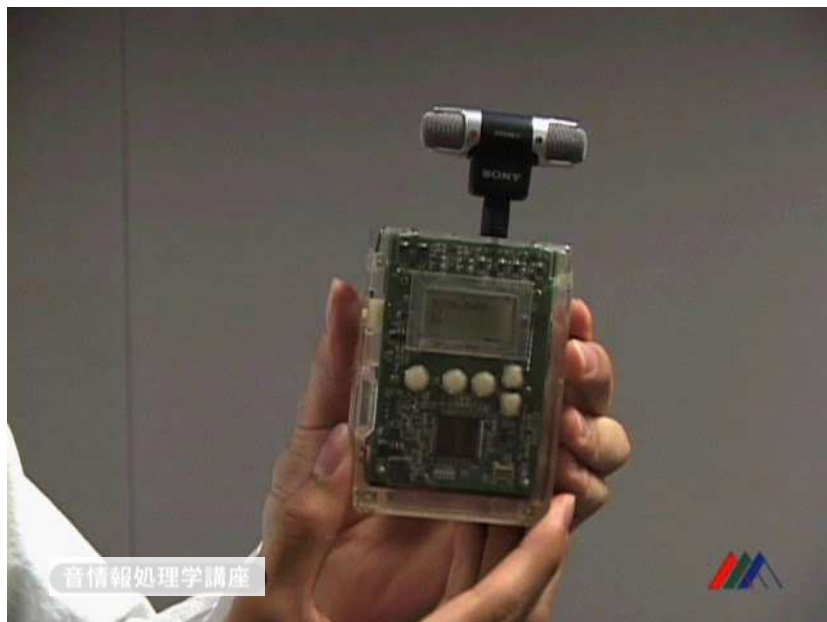
- 独立深層学習行列分析 (IDLMA) の提案

[IEEE Trans. ASLP2019]



高速ICA、独立低ランク行列分析によるデモ

- リアルタイム音声聞き分け(警察備品に採用)
- ドラム、弦楽器、音声からなる複合音の分離

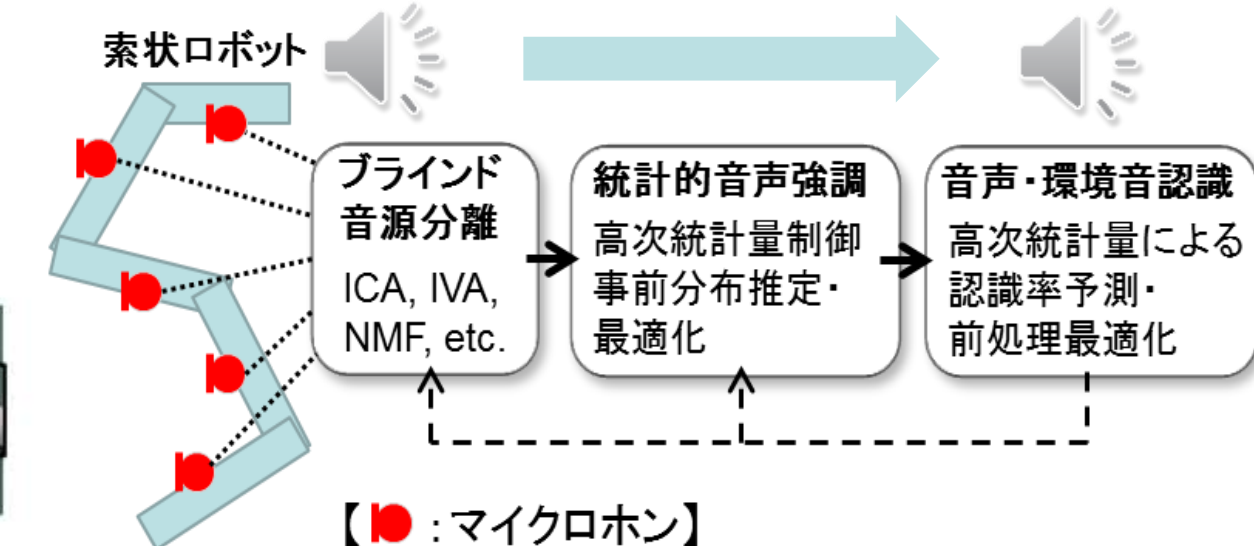


内閣府ImPACT災害対応タフロボット [2016年6月プレスリリース]

- 災害時の倒壊家屋に入り込んで被災者発見
- 環境音認識による状況把握・救助支援



被災者はいるか？



いかなる曲がりくねった形状においても
マイク同士が協調して騒音の中から被災者の声を見つけ出す

研究紹介2. 音楽信号解析

- 様々な楽器がまじりあった音楽信号の中から、自分の好きな楽器を見つけ出し、自分の好みのリミックス版を製作する。
- 非負値行列因子分解(NMF)や深層学習(DNN)という教師有リアルゴリズムを用い、音を事前に学習したパターンに基づいて分解することにより、信号を解析する。

スパース分解
低ランクモデル
教師有り学習

多重解像度深層分析(1/2): 動機

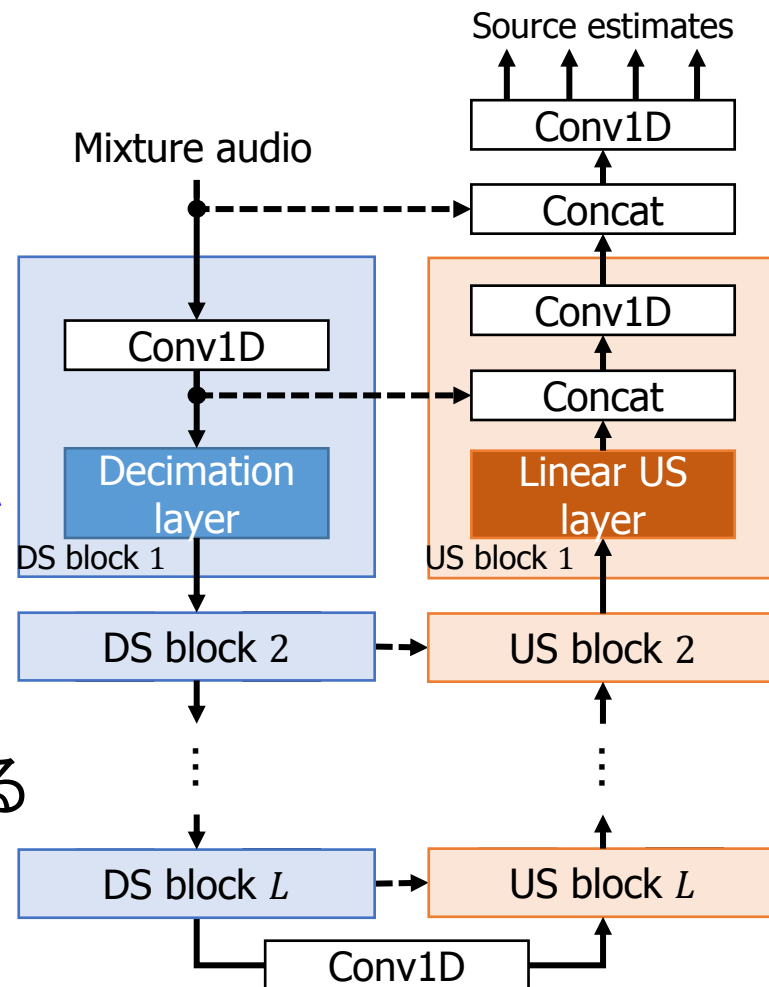
[Nakamura+ IEEE Trans. ASLP2021]

- Wave-U-Net [Stoller+2018]

- 時間信号を直接入力し分離音を得る
End-to-end DNNモデル
- 繰り返しダウン・アップサンプリングを行うU-Net構造をもつ

- しかし、信号処理の観点から見ると
ダウンサンプリングが問題...

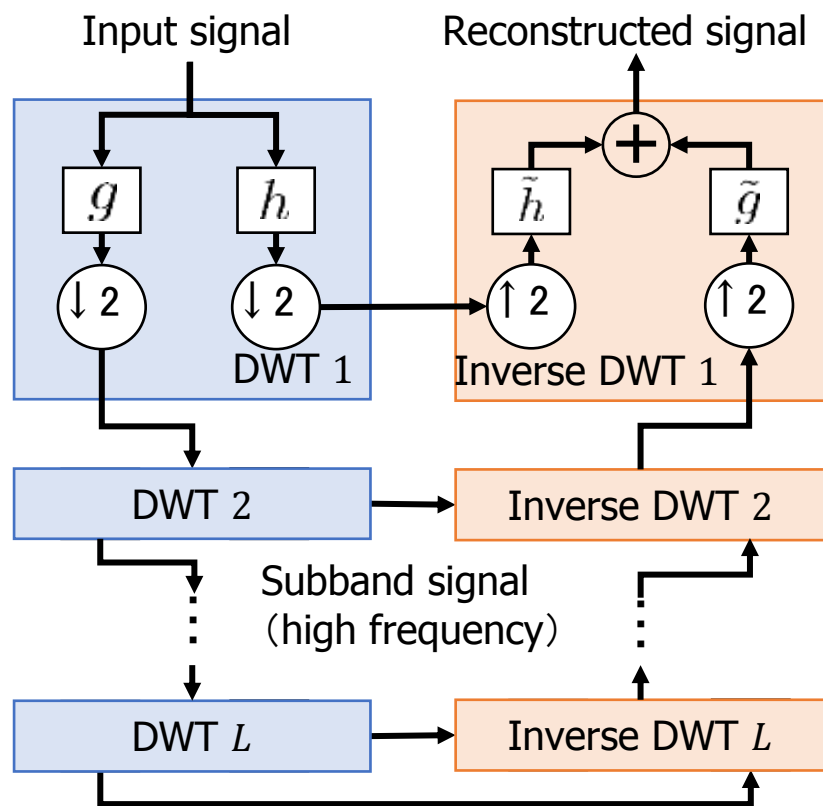
- 特徴量ドメインで**エリアシング**が発生
- ダウンサンプリングで**情報が欠落**する
⇒ **分離性能の低下**を招く 🤔



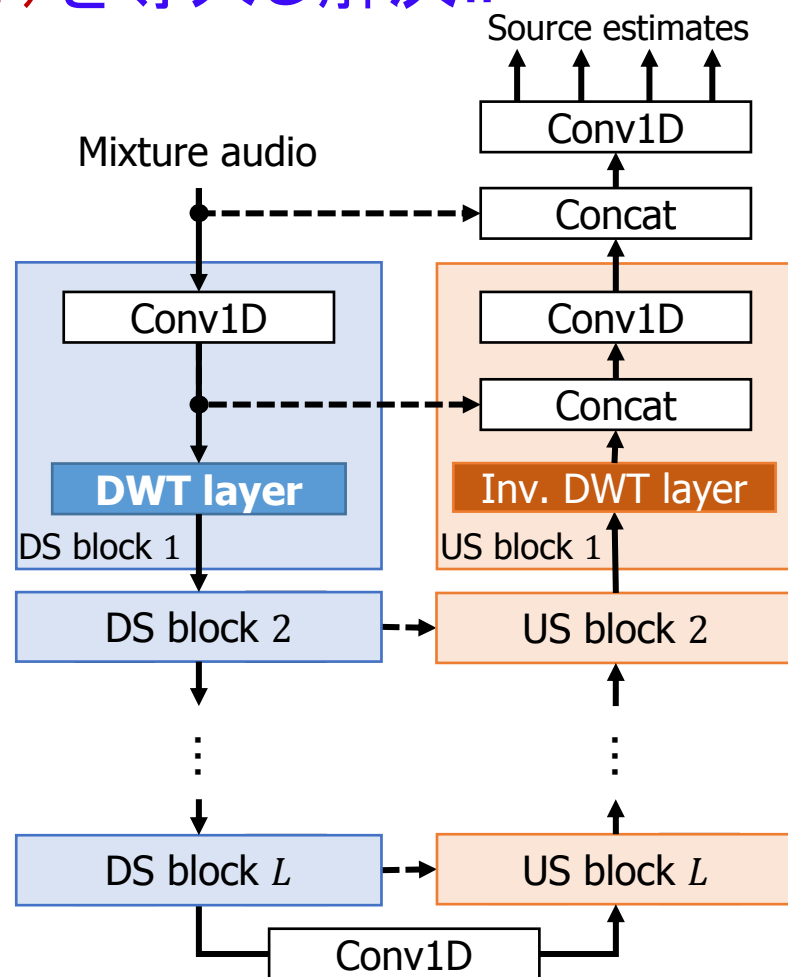
多重解像度深層分析(2/2): 解決方法

[Nakamura+ IEEE Trans. ASLP2021]

- Wave-U-Netと多重解像度解析の構造の類似性に着目
⇒ 離散ウェーブレット変換(DWT)を導入し解決!!



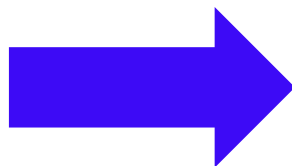
多重解像度分析



多重解像度深層分析(MRDLA)

多重解像度深層分析による楽音分離デモ

- Vocal, bass, drums, guitarの音に分離



分離音

正解音

Vocal:



Bass:



Drums:



Guitar:



研究紹介3. 音場再現

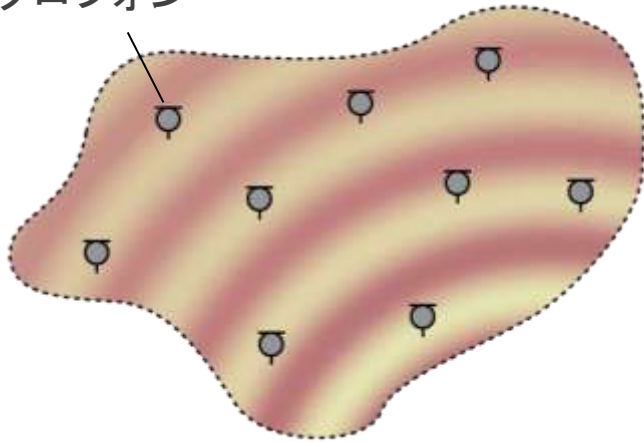
- ステレオや5.1chサラウンドなどの従来の音の再生技術と異なり、広い領域で音の空間そのものを再現する。
- 音のホログラフを実装し、音バーチャルリアリティや音拡張現実感システムを実現する。
- 音場を「スパース(疎)」な性質に基づいて分解する。

球面調和解析
スパース最適化
音VR・AR

音場の計測と制御

計測

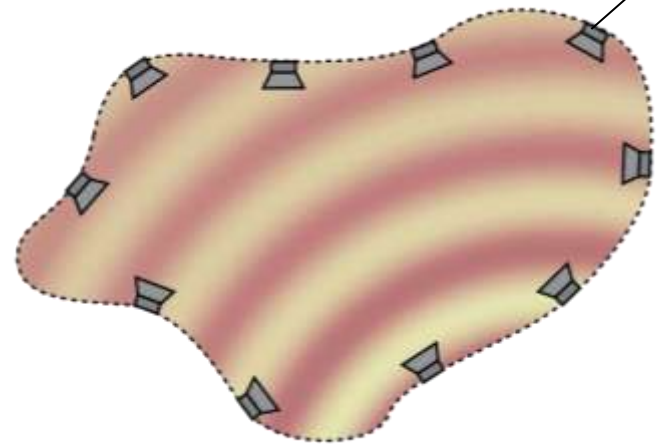
マイクロフォン



- 音場の可視化・再構成
- 音源位置・室内音響パラメータ等の推定
- 空間音響再生のための收音

制御

スピーカ

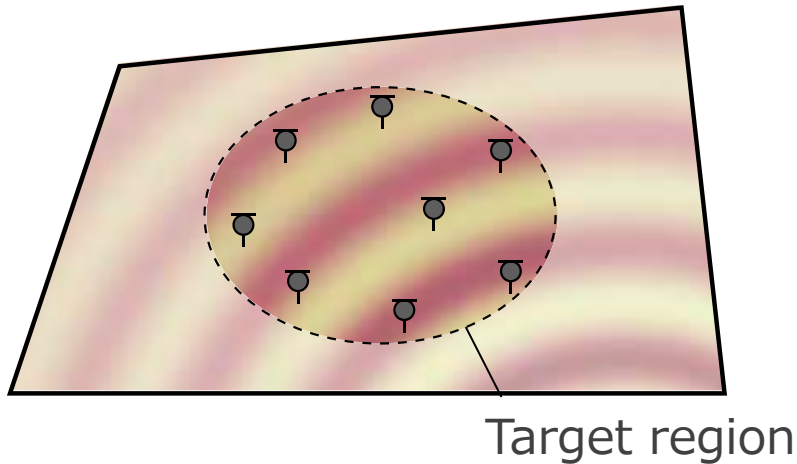


- 高臨場感音響再生
- 指向性制御・エリア再生
- 空間能動騒音制御

音場の計測と制御に関わる信号処理・逆問題を
基礎から応用まで扱う

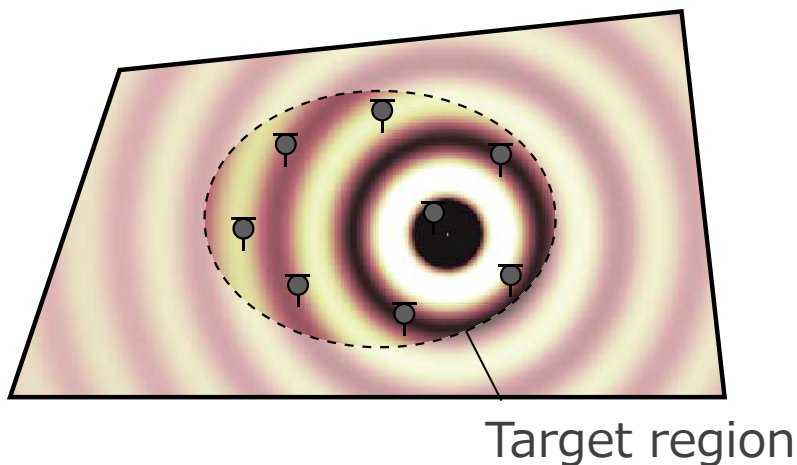
音場の計測

離散配置されたマイクの観測信号から、連続的な音場を再構成



音源を含まない領域の再構成

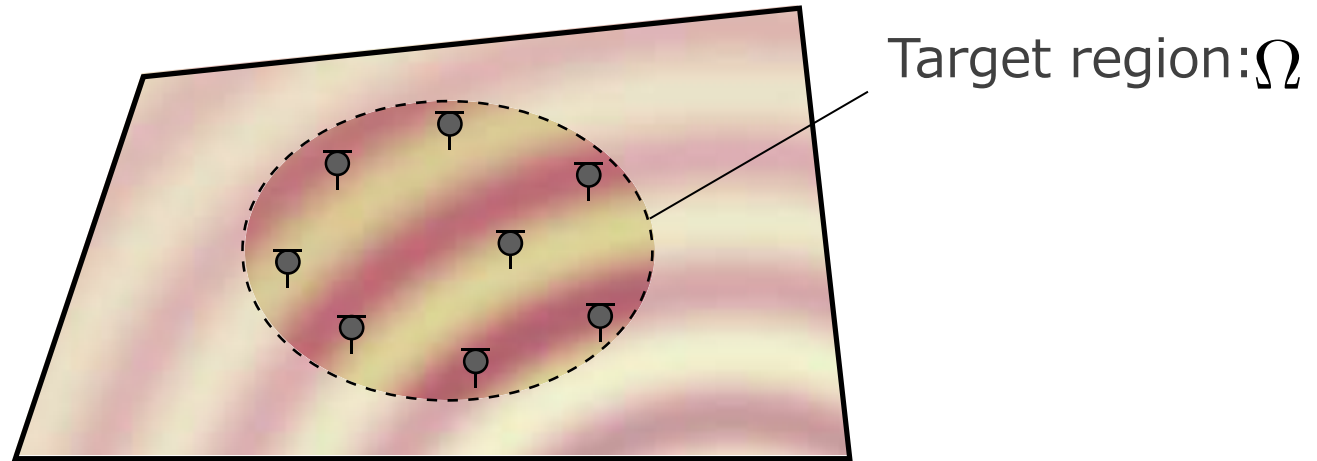
- 無限次元調和解析に基づく再構成 [Ueno+ IEEE SPL 2018]
- 音場のスパース表現のための最適化法 [Murata+ IEEE TSP 2018]



音源を含む領域の再構成

- 音源のスパース性に基づく再構成法 [Koyama+ IEEE JSTSP 2019]
- Reciprocity Gap Functionalに基づく音場分解 [Takida+ Elsevier SP 2019]

音源を含まない音場の再構成

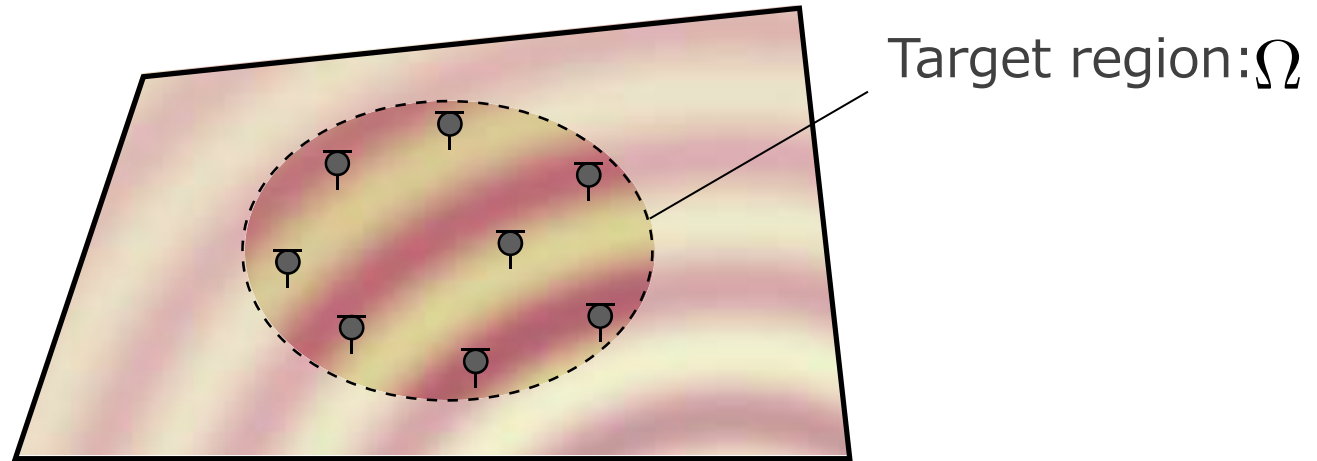


- 音場は斉次Helmholtz方程式に従う

$$\left\{ \begin{array}{l} (\Delta + k^2)u(\mathbf{r}, k) = 0, \quad \mathbf{r} \in \Omega \\ \text{室境界における未知の境界条件} \end{array} \right.$$

音圧分布

音源を含まない音場の再構成



- 再構成音場が波動方程式を満たすように推定したい

マイク位置での音圧

観測信号

$$\underset{u \in \mathcal{H}}{\text{minimize}} \sum_{m=1}^M |u(\mathbf{r}_m) - s_m|^2 + \lambda \|u\|_{\mathcal{H}}^2$$

$$\mathcal{H} = \{u \in \mathcal{C}^2(\Omega; \mathbb{C}), \text{ s.t. } (\Delta + k^2)u = 0 \mid \|u\|_{\mathcal{H}} < \infty\}$$

斉次Helmholtz方程式

音場のカーネル補間法

➤ 解空間とカーネル関数を構成

あらゆる方向から到来する平面波

解空間:

$$\mathcal{U} := \left\{ \frac{1}{4\pi} \int_{\mathbb{S}_2} \tilde{u}(\boldsymbol{\eta}) e^{j\mathbf{k}^\top \mathbf{r}} d\boldsymbol{\eta} \mid \tilde{u}(\boldsymbol{\eta}) \in L_2 \right\}$$

カーネル関数:

$$\kappa(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} J_0(k\|\mathbf{r}_1 - \mathbf{r}_2\|), & \text{in 2D} \\ j_0(k\|\mathbf{r}_1 - \mathbf{r}_2\|), & \text{in 3D} \end{cases}$$

➤ カーネルリッジ回帰による再構成

グラム行列

カーネル関数のベクトル

$$\hat{u}(\mathbf{r}) = \left((\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{s} \right)^\top \boldsymbol{\kappa}(\mathbf{r})$$

観測信号ベクトル

[Ueno+ IEEE SPL 2018, IWAENC 2018]

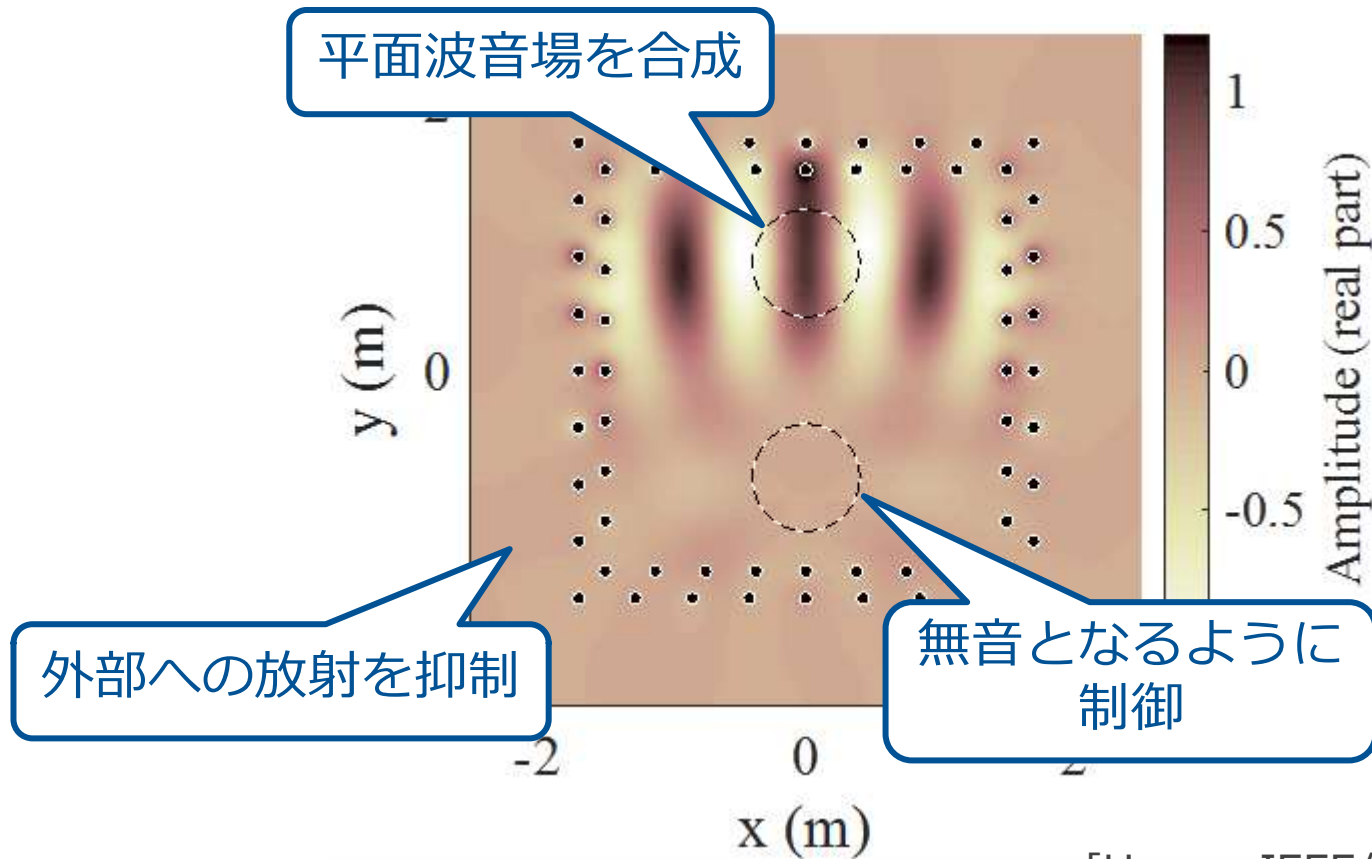
複数のスピーカを用いて対象領域内に所望の音場を物理的に合成

所望音場を合成するためのスピーカ駆動信号を求める逆問題

- 波数領域信号変換による収音・再現 [Koyama+ IEEE TASLP 2013, etc.]
- 音場のスパース表現による超解像化 [Koyama+ JASA 2018]
- 重み付きモードマッチング法による音場合成 [Ueno+ IEEE/ACM TASLP 2019]
- 音場制御におけるスピーカ・制御点配置の最適化 [Koyama+ IEEE/ACM TASLP 2020]

重み付きモードマッチング法

➤ シミュレーション実験例



[Ueno+ IEEE/ACM TASLP 2019]

音場再現だけでなく、マルチゾーン再現や
外部放射抑圧などにも適用可能

仮想再現音場例1（音源がスピーカの手前）

物理的なスピーカ列はここ



研究紹介4. 統計的音声合成

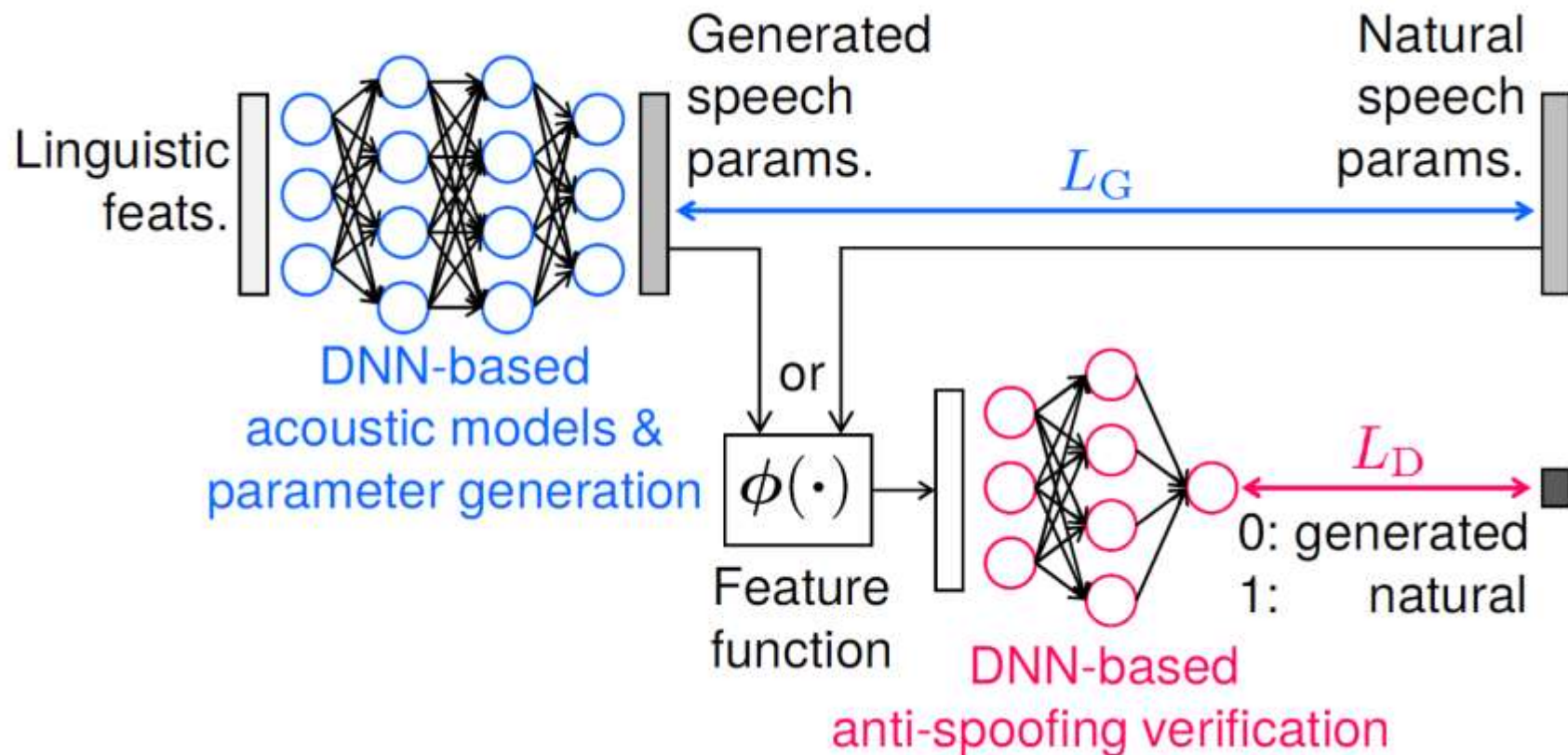
- テキストもしくは自分の声を入れるだけで、誰の声でも、どんな訛りでも、何語でも、喋ることが出来るような統計的音声合成システムを実現する。
- 深層学習 (Deep Neural Net) の枠組みを活かし、「AIオレオレ詐欺師」と「AI防犯課刑事」を対決させて、お互いに精度を高めるAnti-Spoofing 敵対学習理論を独自に提唱

ビッグデータ
深層学習
敵対学習



Anti-Spoofingと敵対する音響モデル学習理論

[ICASSP2017 SPL Student Grant賞・IEEE SPSJ 学生論文賞他]





リアルタイムDNN声質変換



2019年3月 日経xTECHプレスリリース

とにかく音が好きな人は集まれ！

2014年以降の研究教育実績

- ・原著論文：Top論文誌IEEE, ASA, EURASIP, ISCA 26編, 他19編
- ・国際・国内会議発表：**数えきれないので略**
- ・教員・指導学生が2014年以降に**71件の学術賞**を受賞



音の情報処理に興味がある人
統計的信号処理の数理に興味がある人
機械学習を使って生のデータを扱いたい人



一期一会なスモールデータの研究をしたい人
波動現象を数理的に考えたい人 etc.
そんなあなたにお勧めです。

