

### 3-1-1

#### 3-1-1 ヴァーチャルマイクロフォン技術の独立ベクトル分析への適用

Applying Virtual Microphone to Independent Vector Analysis

☆木村凜乃(早稲田大), 上田哲也(早稲田大), 瀬川華子(筑波大), 牧野昭二(早稲田大)

- ◆補助関数法に基づく独立ベクトル分析 (Aux-IVA)はブラインド音源分離の一種であるが、ブラインド音源分離はマイクロフォンの数が音源数より小さい劣決定条件下では成立しない。
- ◆劣決定条件下における音声強調技術としてヴァーチャルマイクロフォン技術が提案されている。先行研究では、ヴァーチャルマイクロフォン (VM)をビームフォーマに適用することで音声強調性能を向上させることが知られている。
- ◆本稿では、 $\beta$ -ダイバージェンスに基づく補間を用いたヴァーチャルマイクロフォン技術を Aux-IVA に対して適用し、評価を行った。実験の結果、ビームフォーマと同様に音声強調性能の改善が確認された。

Table1: Comparison of speech enhancement performance of Aux-IVA.

音源数	マイクの数	VMの数	SDR	SIR
3	2	0	4.86	7.08
3	2	1	<b>5.94</b>	<b>10.34</b>

### 3-1-3

#### 3-1-3 Deep Complex-valued Neural Network-based Mask and Steering Vector Estimation for MVDR Beamforming

☆ Mohan Qin (Waseda Univ.), Li Li (NTT), Shoji Makino (Waseda Univ.)

- ◆MVDR Beamforming is widely used for multi-channel speech separation by minimizing the power of the noise under the constraint of keeping the desired speech distortionless.
- ◆In this paper, we propose a deep complex-valued network-based MVDR beamforming (DCN Beamformer) framework. Firstly, the network is designed with complex-valued layers to fully exploit the phase information. Secondly, we propose a Triple-Path mask estimation that takes spatial information into consideration. Thirdly, we design a complex-valued convolutional recurrent network (CCRN) to estimate the steering vector from the masked signal.
- ◆The experiment results show that by fully exploiting the spatial and phase information, the proposed method can achieve higher SDR score than the existing methods.

Method	SDR
IRM/IBM(2ch)	12.7/13.5
Deep Clustering(2ch)	12.9
ESPnet-SE(2ch)	20.5
<b>(Prop.) DCN Beamformer(2ch)</b>	<b>21.3</b>

Table.1: SDR [dB] results on spatialized anechoic WSJ0-2mix dataset

### 3-1-2

#### 3-1-2 干渉音源の移動範囲全体を抑圧する優決定条件オフライン幾何学的制約付き音源分離

Offline geometrically constrained source separation that suppresses the entire moving range of interfering sources

☆古長慎也(早稲田大), 後藤加奈(筑波大), 上田哲也(早稲田大), 山田武志(筑波大), 牧野昭二(早稲田大)

- ◆Offline GC-AuxIVA-ISS は独立ベクトル分析による音源分離と出力順序問題の解決を同時に行う指向性音声強調手法の1つであるが、すべての時間フレームの信号を用いて時不変な分離フィルタの推定を行うため、音源の移動に対応できない問題がある。
- ◆一方で、Online GC-AuxIVA-ISS は各時間フレームに対して処理を行うことで移動干渉音源に対応可能であるが、干渉音源の急な移動に対して分離性能が低下する問題が生じる。
- ◆そこで本論文では Offline GC-AuxIVA-ISS において、決定条件から優決定条件に拡張し、干渉音源の移動範囲に対して複数の幾何学的制約によるヌル制約を等間隔に与えることで、干渉音源の急激な移動に関わらず、移動範囲全体を抑圧する。音声強調実験を行った結果、提案手法は従来手法よりも分離性能が1.64 dB 高くなることを確認した。

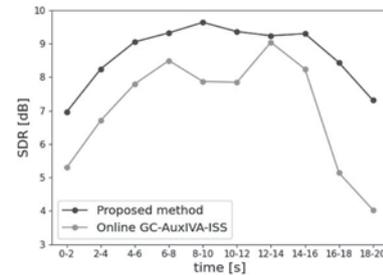


Fig. 1: Average SDR of the target source enhanced by each method in every 2 s. After 10 seconds, the interference signal moves in the opposite direction.

### 3-1-4

#### 3-1-4 A Mask-based Relationship Knowledge Distillation Method for Speech Enhancement with U-net Architectures

☆ Jiachen Wang(Waseda Univ.), Li Li(NTT), Shoji Makino(Waseda Univ.)

- ◆Knowledge distillation (KD) methods aim at compressing the huge deep neuron networks, which utilizing a complex teacher model to guide the training of a light weight student model.
- ◆However, current KD methods primarily concentrate on classification tasks, they didn't receive much attention in speech enhancement tasks.
- ◆We propose a masked based relationship knowledge for speech enhancement models with U-net architectures. It could solve the dimension mismatch problem and relieve the influence of the capacity gap to some extent.
- ◆According to the experiment result, the proposed method outperforms conventional KD methods and has better transfer ability.

Table.1: Performance comparison of various KD methods

Algorithm	PESQ	STOI	CSIG	CBAK	COVL
Student	2.71	0.93	4.00	3.40	3.36
KD	2.74	0.93	4.00	3.41	3.38
FitNet	2.71	0.93	3.98	3.39	3.35
<b>IRM (proposed)</b>	<b>2.81</b>	<b>0.94</b>	<b>4.03</b>	<b>3.44</b>	<b>3.43</b>

### 3-1-5

#### 3-1-5 エイリアス成分のブラインド信号分離とそれに基づく帯域復元

Blind signal separation of alias component and bandwidth restoration based on it

☆中村優希(徳山高専), 中嶋大志, 小野順貴(都立大), 宮崎亮一(徳山高専)

- ◆マイクロホンに適切なアンチエイリアシングフィルタが適用できない場合、観測信号にエイリアシングが生じてしまう。
- ◆単一音源を2マイクで観測する前提で、エイリアシングが生じている信号を元信号成分とエイリアス成分の混合信号と捉え、BSSによりこれらを分離する。さらに分離信号を周波数領域で再構成し、元の音源信号を復元する。
- ◆BSS手法にはFDICAを用いた。エイリアシングのモデルを基に元信号成分とエイリアス成分の位相差をモデル化し、それに基づいて分離信号のパーミュテーションを解いた。
- ◆実際にエイリアシングが生じた信号に提案手法を適用したときのスペクトログラムを Fig. 1 に示す。

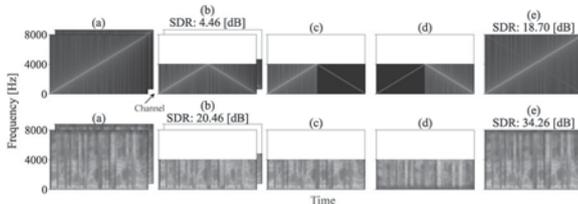


Fig. 1: Spectrograms of the proposed method. The upper and lower are the chirp signal case and speech signal case, respectively. (a) Observation signals without aliasing, (b) Observation signals with aliasing, (c) and (d) Separated signals (original component and alias component), (e) Restored signal.

### 3-1-7

#### 3-1-7 ヘビーテイル生成モデルに基づく独立低ランク行列分析におけるiterative source steeringを用いた分離行列の更新

Iterative-source-steering-based demixing filter updates of independent low-rank matrix analysis using heavy-tailed distributions

◎蓮実拓也, シャイブラー ロビン(LINE(株))

- ◆ブラインド音源分離手法の独立ベクトル分析 (independent vector analysis: IVA) では、近年、iterative source steering 1 (ISS1) や ISS2 に基づき分離行列を更新する手法が提案されている。
- ◆本稿では、IVA の生成モデルを Gauss 分布に拡張した独立低ランク行列分析 (ILRMA) に対して ISS2 を導出する。また、ヘビーテイルな生成モデルを仮定する ILRMA に対しても、ISS1 と ISS2 に基づく更新アルゴリズムを導出する。
- ◆音声と楽音のそれぞれで多チャンネル音源分離実験をおこない、各手法の分離性能を評価する。

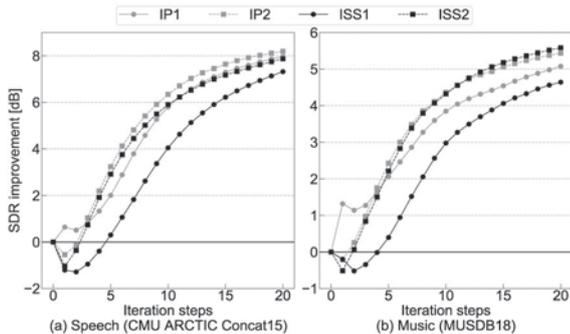


Fig. 1 Gauss-ILRMA の反復に対する source-to-distortion ratio 改善量の変化

### 3-1-6

#### 3-1-6 独立ベクトル分析によるオンライン音源分離・追跡のための高速最適化

Fast online algorithms for independent vector analysis

◎中嶋大志(都立大), 池下林太郎(NTT), 小野順貴(都立大), 荒木章子(NTT), 中谷智広(NTT)

#### 背景

- 1音源追跡: 移動音源(1音源と仮定)のステアリングベクトルを逐次的に推定する問題
- Iterative source steering (ISS): ステアリングベクトルを更新する最適化手法
  - » 音源追跡に適した手法!

#### 目的

- » オンラインISSの高速化

#### 提案手法: Online Source Steering (OSS)による音源追跡

- ⊆ オンラインISSと理論的に等価な挙動
- ⊆ オンラインISSの計算量を  $O(K^3)$  から  $O(K^2)$  に削減

#### 計算量・実験結果

(a) Complexity per frame.			(b) Runtimes [s].		
Algorithm	Complexity		Algorithm	Number of channels ( $K$ )	
	Update	Total		4	6
Online IP	$O(FK^3)$	$O(FK^3)$	Online IP	1.688	4.479
Online ISS	$O(FK^3)$	$O(FK^3)$	Online ISS	1.184	2.634
OSS	$O(FK^2)$	$O(FK^2)$	OSS	0.373	0.596
FastOSS	$O(FK)$	$O(FK^2)$	FastOSS	0.357	0.462

### 3-1-8

#### 3-1-8 Self-Remixing: 音源の分離と再混合による教師なし音源分離

Self-Remixing: Unsupervised Speech Separation via Separation and Remixing

◎西城耕平, 小川哲司(早大)

#### ◆分離・再混合を行う2つのモジュールを用いた教師なし学習

- Shuffler: 観測信号を分離し疑似混合音を作成 (solverの重みで更新)
- Solver: 疑似混合音を分離し観測信号を復元 (勾配法で学習)

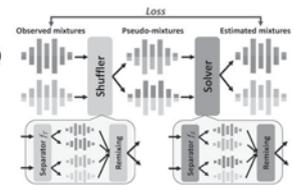


Figure 1. Proposed Self-Remixing

#### ◆評価実験

- WSJ-mix(シミュレーションデータ)を用いた教師なし学習
- WSJ-mix, LibriCSS(録音データ)を用いた半教師ありドメイン適応

#### ◆比較手法

- URCL: 1つのモデルで分離・再混合を2度繰り返す手法
- RemixIT: shuffler と solver の分離音同士で損失を取る手法

#### ◆実験結果: 2つのタスクで安定した学習・高精度な分離を達成

- RemixIT より学習が安定し (Figure 2), 半教師あり学習でも高性能 (Table 1)

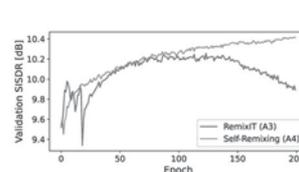


Figure 2. Validation SISDR of RemixIT and Self-Remixing on WSJ-mix

Table 1. WERs on LibriCSS.

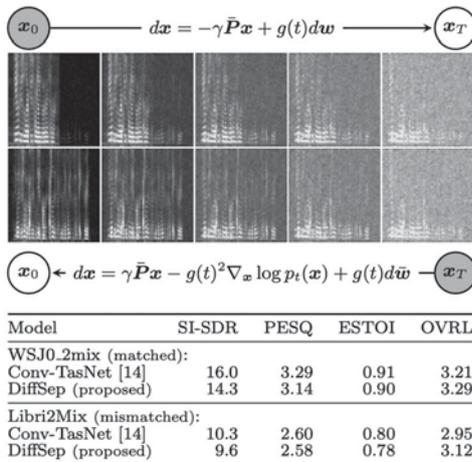
Method	WER[%]
PIT(WSJ-mix)	12.3
URCL	9.8
RemixIT	9.5
Self-Remixing	9.2

### 3-1-9

#### 3-1-9 Diffusion-mixing Process for Speech Source Separation

©Robin Scheibler (LINE), Youna Ji (NAVER), Soo-Whan Chung (NAVER), Jaekuk Byun (NAVER), Soyeon Choe (NAVER), and Min-Seok Choi (NAVER)

- ◆ We propose **DiffSep**, a new approach to single channel speech separation based on **score-based generative modelling**.
- ◆ We use the stochastic differential equation framework and design a diffusion-mixing process.
- ◆ A neural network is trained to approximate the score function of the reverse process and used to solve it, hence separating the sources.
- ◆ Experiments for speech separation and enhancement indicate natural sounding output speech.



### 3-1-11

#### 信号のスパース性とrank-1制約を利用したブラインド残響除去

Blind dereverberation with signal sparsity and rank-1 constraint

☆ 饒平名文希, 矢田部浩平 (農工大)

単チャンネルブラインド残響除去

・・・単チャンネルの残響音声のみを入力として残響を抑制する

残響音声をrank-1行列を用いて表現し、行列のグループスパース性を誘導する非凸最適化問題として定式化

- ▶ クリーン音声と室内インパルス応答を同時推定
- ▶ 残響時間が長い場合にWPEを上回る性能を達成

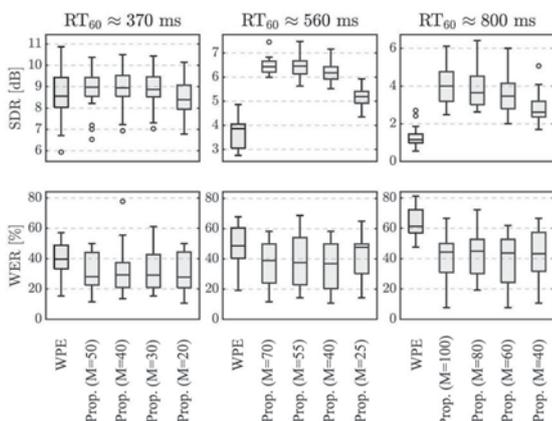


Fig: Performance comparison with WPE (M: length of RIR in STFT domain)

### 3-1-10

#### 3-1-10 深層学習に基づくaudio declippingと制約付きk-スパース最適化の融合による高性能化

Fusion of deep learning and constrained k-sparse optimization for high-performance audio declipping

© 田中僚郎 (早大理工), 矢田部浩平 (農工大), 及川靖広 (早大理工)

audio declipping = 振幅が制限されて歪んだ信号の復元

□ 目的: 幅広い歪みの大きさに対応できる手法の実現

	スパース最適化	深層学習	スパース最適化 + 深層学習
小さい歪み	✓ 信号のモデルを考慮可能	△ 信号のモデルも学習する必要あり	✓
大きい歪み	△ データドリブンでない	✓ データドリブン	✓

□ 提案手法: 複数の処理手法の結果を融合する枠組みを応用  
制約付きk-スパース最適化と深層学習による audio declipping を融合し、歪みの大きさによらない有効性を確認した。

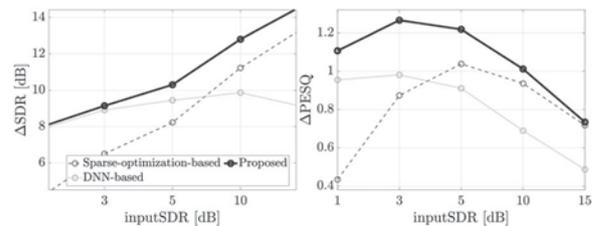


Fig.1 Comparison between conventional methods and proposed one

### 3-1-12

#### 3-1-12 多チャンネル時間差推定における性能限界の導出

Derivation of Performance Limitation in Multichannel Time Delay Estimation

©山岡洗瑛, 植野夏樹, 小野順貴 (都立大)

◆本発表では、マイクロフォンアレイを用いた時間差推定における性能限界について、クラメル・ラオの下界 (CRLB) を用いて議論する。

- ▶ 従来は、チャンネル毎に等しい雑音分散が仮定されることが多く、雑音分散がチャンネル依存の場合、CRLB は数値的に求められてきた。
- ▶ あるセンサ間の時間差推定精度が、他センサの観測によりどれだけ向上し得るかなどは議論されていない。

◆本発表では、閉形式のCRLBを導出し、チャンネル毎に雑音分散が異なる場合の時間差推定の性能限界について解析的に議論する。

- ▶ 導出したCRLBは相互相関関数の最大化に基づく時間差推定におけるCRLBの一般化である。

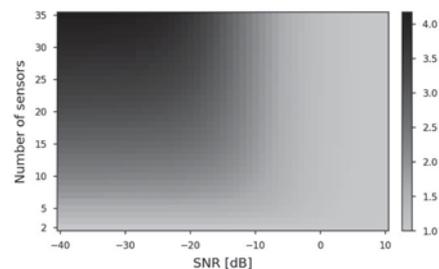


Fig. 1: Estimation gain in time delay estimation using multichannel observation. The color bar shows the ratio of performance limits when using all microphones to when using only two of them. The SNRs are the same for all microphones.

### 3-1-13

#### 3-1-13 正例とラベルなしデータからの学習に基づく非パラレル訓練データを用いた音響信号強調

Audio Signal Enhancement Using Non-Parallel Training Data Based on Learning from Positive and Unlabelled Data

○伊藤信貴、△杉山将(東大)

##### ◆本発表で扱うタスク：音響信号強調

- 信号(音声などの所望のクラスの音)に雑音(その他のクラスの音)が重畳した雑音を含む信号が観測されたときに、信号を推定
- 応用：自動音声認識、音楽情報検索、音響イベント検出

##### ◆従来アプローチ

- 雑音を含む信号とその信号成分の組からなるパラレル訓練データを用いた教師付き学習
- パラレル訓練データは、クロストークにより物理的に収録不可能→鏡像法などのシミュレーションにより合成 (Fig. 1(a))
- シミュレーションで実世界のデータを忠実に再現することは困難→音響信号強調を適用したい、実世界で収録した雑音を含む信号に対して、音響信号強調の性能が極端に低くなる場合があった

##### ◆提案アプローチ

- 実世界で収録可能な雑音を含む信号と雑音からなる非パラレル(組になっていない)訓練データ (Fig. 1(b)) を用いて訓練
- 弱教師付き学習の枠組みである正例とラベルなしデータからの学習 [du Plessis+2014] を利用
- MixIT (mixture invariant training) [Wisdom+2020] を上回る音声強調性能 SI-SNRI (scale-invariant signal-to-noise ratio improvement)

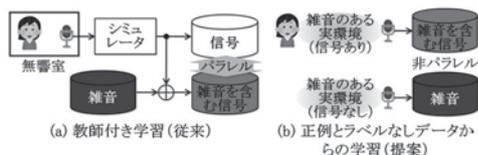


Fig. 1 従来および提案アプローチにおける訓練データ収集方法の比較

### 3-3-2

#### 声質変換チャレンジとASVspoofチャレンジ —話者モデリングの相互発展— 山岸順一(国立情報学研究所)

- Voice Conversion Challenge:
    - 音声の話者性を目標話者へ自動変換する声質変換技術のチャレンジ
  - VoiceMOS Challenge
    - 音声の主観評価値を予測するチャレンジ
  - ASVspoof Challenge
    - 話者認識の生体検知技術を競うチャレンジ
- これらがどう関係し発展してきたかを紹介する

### 3-3-1

#### 3-3-1 音声分離・認識の国際技術評価イベント CHiME チャレンジの紹介

CHiME Speech Separation and Recognition Challenge

○渡部 晋治 (CMU)

- ◆筆者は音声分離・認識の国際技術評価イベントであるCHiMEチャレンジに参加者・運営者双方の立場で関わっており、本稿ではその歴史・概要・技術動向を紹介する。
- ◆本チャレンジは第六回まで行われており、表1にあるように、データ収集及びタスクの複雑度の条件制御を逐次的に行なうことにより、チャレンジで得られる所見を明確にしている。
- ◆最新のCHiME-6では、世界初となる、多人数会話音声認識・音声強調・話者ダイアリゼーションを組み合わせた、挑戦的かつ実用的なタスクがチャレンジとして用いられた。
- ◆本稿では背景の他にも、チャレンジを企画する上での注意点や、各チャレンジで生まれた技術動向、最新のCHiME-7について説明を行う。
- ◆日本の研究機関は本分野において主導的な立場にあり、今後も、チャレンジでの好成績やチャレンジ運営などを通して、本分野における主導的立場を担い続けることを期待したい。

	実施年	語彙サイズ	データ	雑音の種類	発話スタイル	収録環境	発話区間情報
CHiME-1	2011	小語彙	模擬データ	家庭雑音	読み上げ	両耳	既知
CHiME-2	2012	大語彙	模擬データ	家庭雑音	読み上げ	両耳	既知
CHiME-3	2015	大語彙	模擬・実データ	公共雑音	読み上げ	単一マイクアレイ	既知
CHiME-4	2016	大語彙	模擬・実データ	公共雑音	読み上げ	単一マイクアレイ	既知
CHiME-5	2018	大語彙	実データ	家庭雑音	自由発話	複数マイクアレイ	既知
CHiME-6	2020	大語彙	実データ	家庭雑音	自由発話	複数マイクアレイ	未知

Table 1: CHiME チャレンジタスクの変遷

### 3-3-3

#### 3-3-3 計算機による環境音の理解・解釈に向けた統合的コンペティション DCASE CHALLENGE への招待

Invitation to DCASE CHALLENGE

○井本 桂右(同志社大学)

- ◆コンペティションの目的や概要、活用方法について紹介します
  - データセットや評価ツールを利用したい人
  - コンペティションに参加したい人
  - DCASE Challenge でタスクを主催したい人
  - 動向を把握しておきたい人
 それぞれに役立つ情報をお話します
- ◆研究発表会でしか聞けない秘情報やお得な情報も!?
- ◆DCASE2023 Challenge で扱われるタスク
  - 音響シーン分類(軽量モデル)
  - 音響イベント検出(弱ラベル, ソフトラベル, 生物音響)
  - 異常音検知(機器の状態監視)
  - 音響イベントの種類と到来方向の同時分析(実環境音)
  - 音響キャプション生成
  - 効果音・環境音合成

DCASE 2023 CHALLENGE への参加をお待ちしています!

### 3-3-4

#### 3-3-4 深層学習モデルを用いた言語特徴と音響特徴の後期融合による音声感情認識

Speech emotion recognition by late fusion of linguistic and acoustic features using deep learning model

☆佐藤清秀(山形大学/ピットアート), △岸恵志, 小坂哲夫(山形大学)

- ◆本研究では、言語特徴と音響特徴の後期融合でより実用的な音声感情認識の手法について検討する。
- ◆感情音声認識結果による言語特徴、統計量および時系列特徴による音響特徴の認識出力を用いて、DNN による後期融合を行ない、最終結果を決定する。
- ◆音響特徴と言語特徴の融合により各々の低認識率部分を補うことで高い精度の認識が可能であることが確認できた。また、重み付き融合法を上回る認識精度を得て、本システムの有効性が示された。

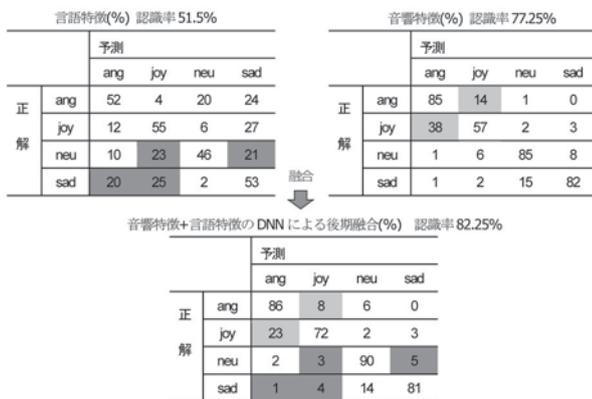


Fig.1 Confusion matrix of linguistic and acoustic features using deep learning model

### 3-3-6

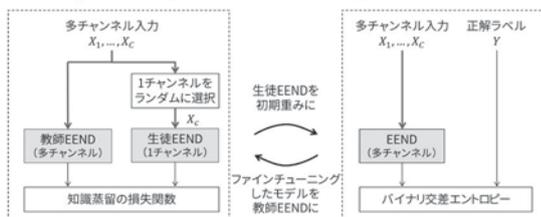
#### 3-3-6 1チャンネル及び多チャンネル話者ダイアライゼーションモデルの相互学習

Mutual learning of single- and multi-channel speaker diarization models

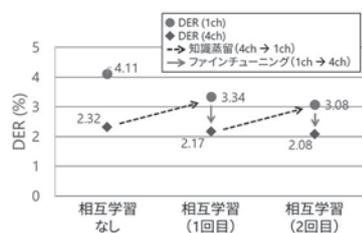
◎堀口翔太<sup>1</sup>, 高島悠樹<sup>1</sup>, 渡部晋治<sup>2</sup>, △Paola Garcia<sup>3</sup>

<sup>1</sup>日立 <sup>2</sup>CMU <sup>3</sup>JHU

- ◆1チャンネルモデルと多チャンネル音声処理モデルの相互学習を提案した。相互学習では以下の操作を交互に繰り返す。
  1. 多チャンネルモデルから1チャンネルモデルへの知識蒸留
  2. 1チャンネルモデルから多チャンネルモデルへのファインチューニング
- ◆End-to-end 型の話者ダイアライゼーション (End-to-End Neural Diarization; EEND) を用いて手法の有用性を検証した。



- ◆2話者対話のシミュレーションデータおよび実データを用いた実験の結果、提案手法は1チャンネルおよび多チャンネルEENDのダイアライゼーションエラー率 (DER) をそれぞれ削減できた。



### 3-3-5

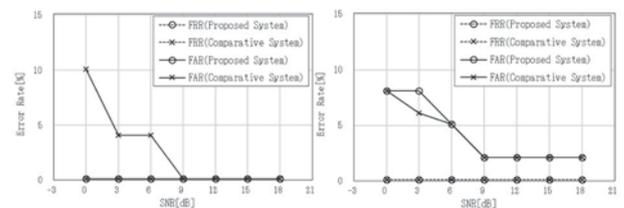
#### 3-3-5 干渉雑音除去後に所望音声を選択するオンライン通話手法の提案

Proposal of Online Calling Method with Speaker Identification after Interference Noise Removal

◎菓谷浩亮, 花野博司, 堀内俊治(KDDI 総合研)

近年オンライン通話アプリケーションの普及により、場所を問わずにオンライン通話を実施することが可能となっている。しかし、複数人が至近距離でオンライン通話をする時に、通話端末に他者の音声が入り込んでしまい、聴者は目的話者の音声のみが聞きとりにくくなってしま...

本稿における提案方式は、端末間が近接することにより生じる干渉雑音をブラインド音源分離により低減させ、目的話者のみの音声を相手端末に送信することを目的とする。ただし、音源分離前の混合音声に含まれる複数話者の音量のバランスが変化すると、分離音声情報の出力インデックスが変化してしまう。そのため、事前に話者の音声情報を記録し、複数の分離音声情報と事前情報とを、声帯特徴量に基づき照合して目的話者の分離音声情報を選択する手法を提案する。シミュレーションの結果、提案する話者照合手法は従来手法と比較すると、Fig.1のように異性による混合音声から話者照合を行う際の性能が優れていることを示す。



(a) Opposite-sex voice mixture (b) Same-sex voice mixture

Fig.1: Performances of the proposed system.

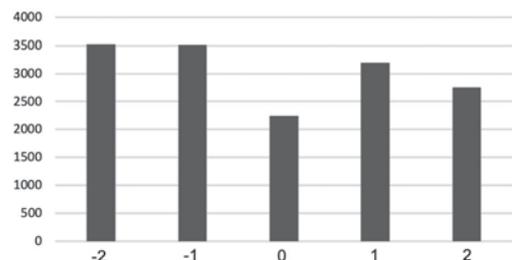
### 3-3-7

#### 3-3-7 感情演技発話動画へのクラウドソーシングを感情ラベル付与とラベルの分析

Labeling Emotional Utterance Videos Using Crowdsourcing and Analyzing Labels

☆ 堀井大輔, 伊藤彰則, 能勢隆 (東北大)

- ◆本研究では演技発話を収録し、日本語マルチモーダル感情コーパスの構築を行った。
- ◆映像収録後、クラウドソーシングを活用し、ネガティブ・ポジティブ、元気がない・元気がある、という2次元の感情次元ラベルの付与を行った。
- ◆発話文収集の際、テキスト感情認識を活用したことにより付与されたネガティブ・ポジティブラベルのバランスがよかった。



-2: Negative ~ 2: Positive

Fig.1 Total number of labels

- ◆ラベルと音響特徴量・表情特徴量の間で相関分析を行ったところ、AU12 (唇両端を上げる)の最大とラベルとの間で相関係数 0.4 程度の相関が確認された。

### 3-3-8

#### 3-3-8 若年話者判別のための x-vector と話者年齢埋め込みベクトルの比較

A comparison of x-vector and speaker age embedded vector for identification of young speakers

☆奥本佑哉, 西村竜一(和歌山大)

- ◆ 音声「大人っぽい声」か「子供っぽい声」かの自動判別を検討
- ◆ 話者年齢埋め込みベクトル(Age-vector)の導入を提案
  - 話者年齢を推定するようにニューラルネットワークを学習
  - 学習したネットワークを用いて発話音声から話者年齢埋め込みベクトルを抽出
- ◆ クラウドソーシングで収集した年齢・性別ラベル付き実環境発話を対象とした x-vector, Age-vector の比較実験
  - AgeVoxCeleb を用いて x-vector, Age-vector の抽出モデルを構築
  - 実環境発話音声から各特徴ベクトルを抽出
  - 大人と若年者の境界年齢として年齢閾値を設定し, 大人の男性・大人の女性・若年者の3クラスにラベリング
  - 各特徴ベクトルを入力としてラベルを予測する全結合 NN を構築
  - 年齢閾値を 9~17 歳に設定して交差検証
- ◆ 年齢閾値ごとの分類結果を Accuracy, F 値で評価・比較 (Fig.1)

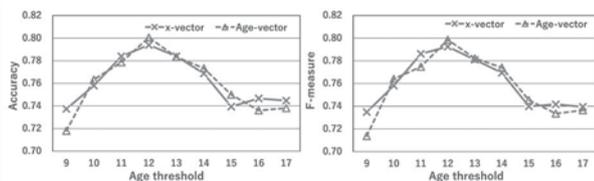


Fig.1: Accuracy and F-measure for each age threshold

講演取消

### 3-3-10

#### 3-3-10 シャドーイングを用いた英語学習者の聞き淀み計測とその自動推定

Automatic Prediction of English Learners' Listening Dysfluency Using their Shadowing Speech

☆高英翔, 峯松信明, 齋藤大輔(東大工), 中西のりこ(神戸学院大)

- ◆ リスニングは言語学習において重要な基本技能の一つである。しかし、学習者のリスニングが具体的にどこで聞き淀む(誤る)のかを検出する技術は少ない。
- ◆ 以前、聴取中の「聞き淀み」を計測する方法として、シャドーイング時の発話の乱れに着目し、単語単位で計測、モデル化し、新たな提示音声に対する「聞き淀み」を予測する研究がある[Kunihara+, 2022]。
- ◆ 本研究では同じ方法を用いて、予測モデルの変更、特徴量の追加を行った。さらに、様々な特徴量を選択的に利用することで、提示音源のどのような素性が予測に貢献するのかを分析的に検討した。
- ◆ 実験の結果、変更したモデルと追加特徴量の効果が確認された。また、具体的な特徴量の効用を解析し、更なる性能向上を実現した。分類精度は正解率が0.571から0.647まで上昇した。回帰予測では、相関が0.344から0.421まで上昇した。

### 3-3-9

### 3-3-11

#### 3-3-11 End-to-End 音声合成を利用したテキストまたは音声からの統合的な顔ランドマーク生成

Unified facial landmark generator from text or speech using end-to-end text to speech

©三井健太郎, 法野行哉, 沢田慶(rinna)

- ◆ 従来、音声と同期した顔生成にはテキスト・音声の一方を利用
- ◆ テキスト・音声から共通表現を抽出し、その共通表現から顔生成を行うことで、いずれのモダリティからも顔生成を実現
  - End-to-End 音声合成の学習によって得られる表現を利用
    - テキスト入力の場合には音声・顔ランドマークを同時生成
  - 音声の話者性を除去・感情表現を抽出する機構を内包
    - 音声の話者に非依存・感情に依存した顔ランドマーク生成
  - テキスト・音声のいずれを入力とした場合にも、既存手法と比較して高品質な顔ランドマークを生成可能

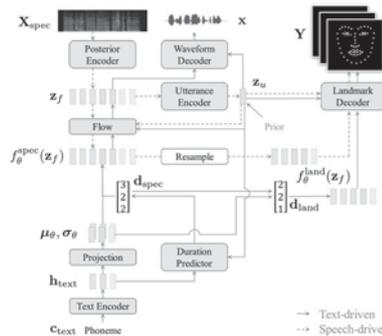


Fig.1: Conceptual diagram of inference of the proposed method.

### 3-3-12

#### 3-3-12 話者変換によるデータ拡張を利用した叫び声合成

Effect of data augmentation by speaker conversion for scream synthesis.

☆白鳥恵大, 有本泰子(千葉工大)

◆背景: 自然で表現豊かな音声対話の実現に向けて, 対話中の突発的に表出された叫び声を合成することを目指す

問題点:

- 叫び声は日常場面でも出現頻度の低い音声現象であり, データ不足により十分に学習できず, 不自然な叫び声を生成してしまう
- コーパス内に叫び声データが少ないことから, 特定話者のみの叫び声では, 表現の多様性に限りがある

◆目的: 特定話者の叫び声の表現に限定されずに, 複数話者の叫び声の表現を叫び声合成の学習に利用し, 自然な叫び声合成の実現する

◆方法: 別の話者の叫び声を話者変換して, 合成対象話者の叫び声をデータ拡張することを提案

◆結果: 話者変換によるデータ拡張した叫び声合成 (tvc) は男性を対象とした叫び声合成の自然性を改善する可能性が見られた

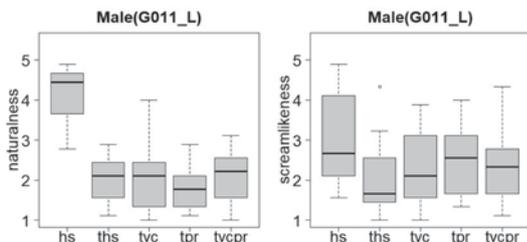


Fig.1: Results of naturalness evaluation and scream evaluation

### 3-4-1

#### 3-4-1 小鳥の音声伝えるもの: さえずりと地鳴きの仕組みと機能

What songbird vocal signals convey: Mechanism and function of song and call

○橋亮輔(東京大)

◆さえずる小鳥である鳴禽類の音声コミュニケーションについてジュウシマツとキンカチョウを中心に解説する。発声の種類や発声制御の仕組みを紹介しつつ, 発声を持つ社会的機能や情動伝達機能について検討する。

◆発声には大きく分けて, 歌(さえずり)とコール(地鳴き)がある。歌は学習性の系列発声であり, 主に求愛や縄張り防衛の機能を持つ。一方, コールは単純な短い発声であるが, 多様な音響パターンがある。警戒や餌ねだりなど多彩な機能があると考えられる。

◆社会的発声: 群れの成員間で頻りにコールを鳴き交わす。距離が近いと短く, 遠いと長くて強い声を出す。群れを形成して維持する機能を持つのかもれない。このような発声の社会的機能について議論する。

◆情動発声: トリがヒトと同じ情動状態を持つかは分からないが, 内的状態がネガティブかポジティブかといった大まかな判断は可能だろう。情動を反映すると思われる発声行動について議論する。

◆音声使用・音声発達の種間比較: ヒトの音声コミュニケーションには非言語的なやりとりが多分に含まれる。これを可能にする基盤能力を明らかにする上で, 鳴禽類との比較は非常に有効である。発声行動を他種間で比較することにより, 音声使用や発声学習の普遍的な仕組みについて検討する。

### 3-3-13

#### 3-3-13 振幅重み付けエネルギー関数を用いたボルツマンマシンによる位相復元

Phase recovery by temporal conditional Boltzmann machine

☆羽賀洋克(電通大), 矢田部浩平(早稲田大), 岸田拓也, 中鹿亘(電通大)

◆Griffin-Lim 法は位相復元に用いられるアルゴリズムの一種だが, 初期値に大きく依存し, 不自然なアーティファクトを生成してしまう。

◆過去の位相と振幅を入力するモデルによる位相復元が提案されているが, 品質の高い位相復元には不十分だった。

◆位相復元において振幅によって位相の重要度には差が存在しており, それを考慮することで位相復元の品質を向上させることができる。

◆条件項に位相と振幅の積を適用したボルツマンマシンを用いることで品質の高い位相復元が可能になった。なお, この条件項は検索スペクトログラムに相当する。

◆Pesq による評価では従来法と比べ大幅に性能の向上が確認できた。

◆Fast Griffin-Lim 法と組み合わせることでより高い効果が確認できた。

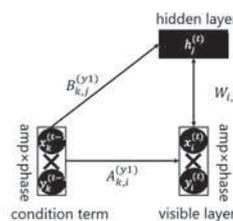


Fig.1 The proposed model

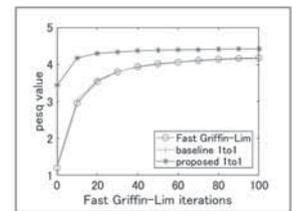


Fig.2: Variation of pesq values by the number of Griffin-Lim iteration

### 3-4-2

#### 3-4-2 ヒト乳児における音声の特異性, 情動性, 相互性と音声言語発達

Specificity, emotionality, and mutuality of speech for human infants and speech development

○皆川泰代(慶應義塾大・文)

◆ヒト乳児において, 初語が出る前の生後約1年間にも「声」は言語発達, 社会性発達のために重要な役割を果たす。

◆本稿では生後1年間の, 乳児への音声入力(乳児の知覚特性や養育者音声要因), 音声出力についてヒト声の特異性, 声を持つ情動性, そして養育者との相互性に着目する。

◆言語や社会性発達の遅滞がみられがちな自閉スペクトラム症リスク児と定型発達児を比較する行動研究, 脳機能研究を概観し, 特に声や相互性を持つ報酬性を考察しつつ Fig.1 の内容を説明する。

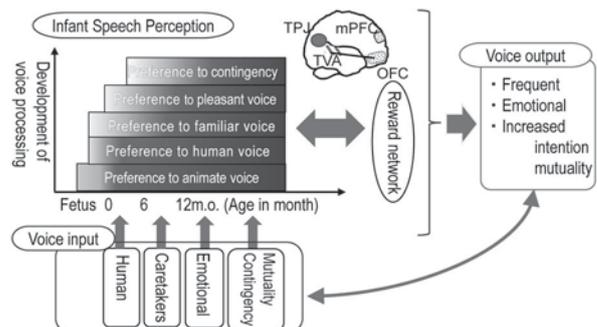


Fig.1. Influence of voice input on infant's speech perception and its resultant voice production (output)

### 3-4-3

#### 3-4-3 情動表出としてのマウス超音波発声—その発声の種類と個体差

Mouse ultrasonic vocalization as emotional expressions: Types of vocalizations and individual differences

○菅野康太(鹿児島大・法文)

- ◆マウスの音声コミュニケーションは超音波発声で行われます！
- ◆超音波発声は ultrasonic vocalizations、略して USVs
- ◆本発表ではマウスの超音波発声が情動表出であることを紹介します
- ◆主に、雄から雌への求愛発声の話します
- ◆求愛発声の他に、仔マウスが発するものや、雄→雄、雌→雌などがあります。
- ◆共同研究で作ってもらった超音波発声の自動解析技術(こちらの学会員の橋亮輔博士に感謝)や音源定位装置の話もします！
- ◆マウスを走らせたなら鳴きました(特定の遺伝系統で)
- ◆最近、マウス USVs の解説サイトを作りました

### 3-4-4

#### 3-4-4 母子分離が成体スナネズミの社会的コミュニケーションに与える影響

Effects of Maternal Separation on Social Communications in Adult Mongolian Gerbils

☆西堀諒, △新家一樹, △片山瑞月(同志社大),

玉井湧太(チュービンゲン大), △伊藤優樹, 小林耕太(同志社大)

- ◆幼少期の虐待は脳の発達に影響を及ぼす可能性があり、精神疾患や社会的コミュニケーション障害の発症リスクを増大させる。育児放棄を模した動物モデルとして、一定期間母と子を隔離する母子分離がある。
- ◆母子分離が社会的コミュニケーションに与える影響は未解明である。
- ◆本報告では音声の社会的機能が判明(Fig.1A)している成体スナネズミを使用し、母子分離が①接近行動を減少(Fig.1B)させ、②発声の音響的特徴を変化(Fig.1C)させることを明らかにした。更に健康個体の働きかけでコミュニケーションが改善する可能性も示唆された。

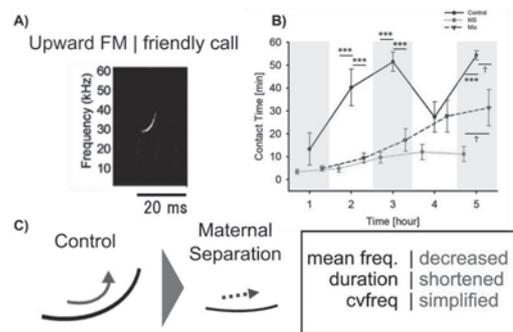


Fig. 1 (A) Friendly vocalizations of gerbils, (B) Contact time between the two gerbils, and (C) Changes in Acoustic Features. \*\*\*p<.001, †p<.10

### 3-4-5

#### 3-4-5 機械学習によるマーモセットの発声分類とそれに基づく発達に伴う発声変化の分析

Marmoset Calls Classification by Machine Learning and Its Application to Analysis of Their Vocal Development

☆上坂奏人, 河内秀人, 山岡洗瑛, 若林佑幸(都立大), 木下裕磨(都立大/東海大), 小野順貴(都立大), △野口潤, △渡邊恵, △一戸紀孝(国立精神・神経医療研究センター), △ベナー聖子(浜松医大/国環研), △山末英典(浜松医大)

- ◆自閉スペクトラム症 (ASD) の解明のため、ASD 症状を再現した ASD モデルマーモセットを用いた発声の分析が行われている。
  - しかし現状、マーモセットの発声の分析は人手で行われており、研究者らの負担になっている。
- ◆本研究ではマーモセットの自動発声分類に取り組み客観評価実験を行った
  - 通常のマーモセットの音声だけで学習した深層学習モデルを用いた場合でも、ASD モデルマーモセットの発声を、Macro F1-score 88.0 % の精度で分類できることが分かった。
  - また、人手でアノテーションした発声ラベルにより分析したマーモセットの発達に伴う発声変化の傾向 [Fig. 1 (a), (b)] が、学習した深層学習モデルを用いて推定したラベルからも確認できた [Fig. 1 (c), (d)].

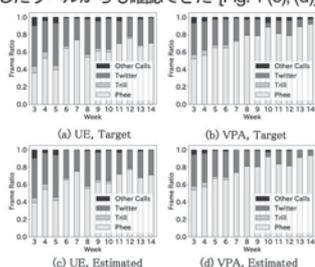


Fig.1: Weekly changes in the ratio of calls frames for target labels and estimated labels (UE Test set, VPA set)

### 3-4-6

#### 3-4-6 霊長類の群れ社会での音声交換と音声の情動

Vocal exchange and vocal emotion in primate societies

○香田 啓貴(東京大・総合文化)

- ◆霊長類の音声と、情動支配性、そのやり取りである鳴き交わしといった項目に対して、生物進化の視点を織り交ぜながら、その特徴を外観し、具体的な興味深い音声交換事例とその解釈方法、そして、今後の将来像、展望などの情報提供をする。
- ◆ヒト以外の霊長類では、発声学習がない点や条件づけ学習過程が成立しにくい点、またそれを裏付ける神経基盤のヒトとの違いから、発声運動が随意運動的でなく情動支配的とされている。
- ◆サルの音声コミュニケーションは、求愛や警戒といった特定の目的の音声の運用ではなく、単純に仲の良い個体間で音声をやり取りする音声交換が主流である。
- ◆情動によって音声が音響変化を受ける点、音声による情動状態の伝達が行われる点、情動の伝達は伝染などのような形式で自動的な過程として処理される点、などの特徴を考慮すると、サルの社会集団で観察できる音声コミュニケーションは情動情報の相互作用の動物行動モデルとして最適な研究対象であると考えられるだろう。

### 3-6-1

#### 3-6-1 音環境の可視化と可聴化を目指した音響収録・再生システム

Acoustic recording and reproduction system for visualization and auralization of sound field

©山内源太, 田部洋祐(日立), 亀川徹, 丸井淳史(東京藝大)

- ◆製品使用空間における音環境の評価に向けて、音の到来方向の把握(可視化)による物理評価と、高臨場再生(可聴化)による主観評価を目指した音響収録・再生システムについて述べる。
- ◆音環境は球で定義される音響要素の集合体と仮定すると、音響要素の中心は立方八面体の頂点と菱形十二面体の頂点と一致する。両立体の重心を一致するように重ねた立体のアレイを等軸晶アレイと呼ぶ。等軸晶アレイは球面アレイであり、実験により球面アレイは全方位の可視化と可聴化が可能であることを確認した。

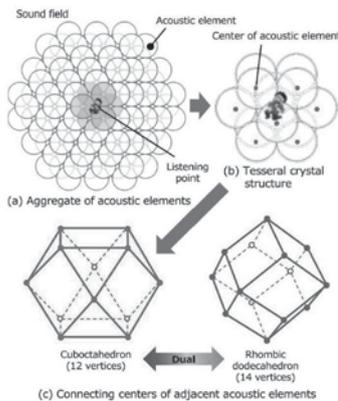


Fig.1: Geometry of tesser crystal structure.

### 3-6-2

#### 3-6-2 (招待講演)複合現実技術を用いた音場の計測と可視化

Sound field measurement and visualization using augmented reality

○及川靖広(早大理工)

- ◆近年、複合現実技術の発展が著しく、比較的簡易に用いることが可能となってきた。音響計測分野へ適用してきた。

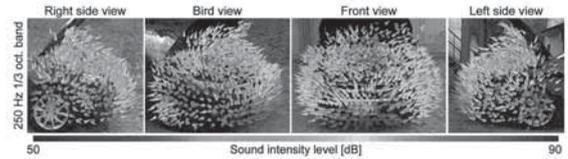


Fig.1: 自動車のエンジン音の可視化計測事例

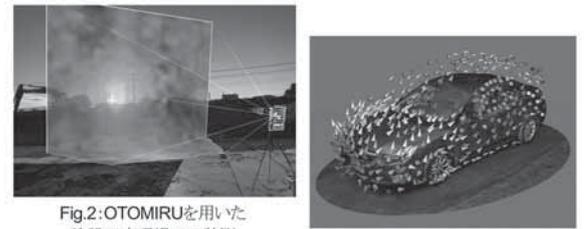


Fig.2: OTOMIRUを用いた建設工事現場での計測

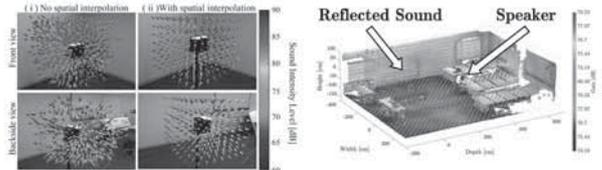


Fig.4: 信号処理技術の活用(音響インテンシティ計測データの補間)

Fig.3: LiDARを用いた3Dデータの活用

### 3-6-3

#### 3-6-3 三次元音源位置推定におけるLiDARを用いた3Dデータ利用の検討

Investigation of using 3D data with LiDAR for three-dimensional sound source estimation

©後藤昌彦, 及川靖広(早大理工),

井上敦敏, 寺岡航(早大理工/INSPIREI)

- ◆三次元音源位置推定において、Light Detection and Ranging (LiDAR)を用いた3Dデータの利用について検討
- ①LiDAR スキャナが搭載された iPad Pro で観測空間の三次元情報(3Dデータ)を取得 (Fig.1)
- ②3Dデータに基づき音源位置を仮定し、MVD R ビームフォーマの出力パワーを計算
- ③出力パワーを3Dデータに重畳することで音源の位置を推定
- ◆反射音の可視化実験では、壁からの反射音が可視化 (Fig.2)

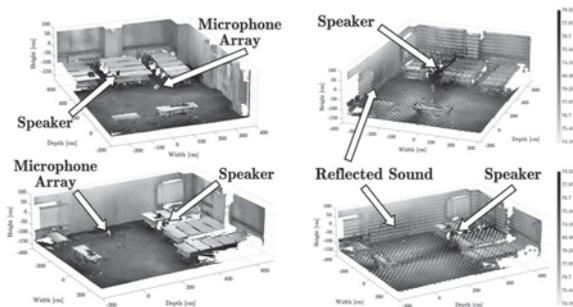


Fig.1 3D scan with LiDAR.

Fig.2 Visualization of sound source estimation.

### 3-6-4

#### 3-6-4 発音の空力音シミュレーションと指向性の可視化

Aeroacoustic simulation of speech production and visualization of acoustic directivity patterns

○吉永 司, 飯田明由(豊橋技科大)

- ◆仮想現実や対話システムでのより臨場感のある会話実現のため、音声の指向性の予測や、その指向性が生まれるメカニズムを明らかにすることが求められている。
- ◆本研究では、3次元圧縮性ナビエ・ストークス方程式を支配方程式とする空力音シミュレーションにより子音の発音の指向性を予測する技術を紹介する。
- ◆医療画像から形成した口腔形状に対して計算格子を構築し、高精度の有限差分法によりシミュレーションを行う。
- ◆唇から遠方10 cmでの音圧分布の指向性をプロットすることにより、唇の形状が摩擦音の指向性に大きく影響することがわかった。

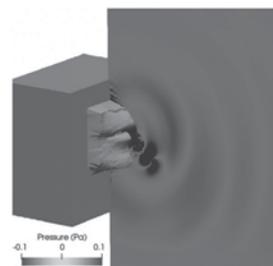


Fig.1: Pressure distribution of the aeroacoustic simulation.

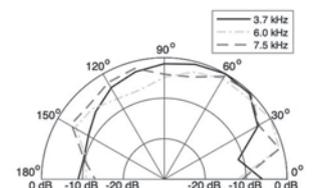


Fig.2: Acoustic directivity pattern at 10 cm from lips on the vertical plane.

### 3-6-5

**3-6-5 音空間レンダリングによる可視化・可聴化**  
 -シリコンコンサートホールからコウモリまで-  
 Visualization and auralization by sound field rendering  
 -from silicon concert hall to bats-

○土屋 隆生 (同志社大・理工)

音空間レンダリングでは、反射・散乱体の形状や反射率などの音響特性データを元に、音源から放射されたインパルスやドライソースに対し、波動性を考慮した音場計算を行う。本稿では、音空間レンダリングによる可視化・可聴化の例として、ホールシミュレーション、バイノーラルレンダリングによる音楽練習システム、シリコンコンサートホール、超音速旅客機の騒音予測、コウモリの超音波センシング解明への適用について解説する。

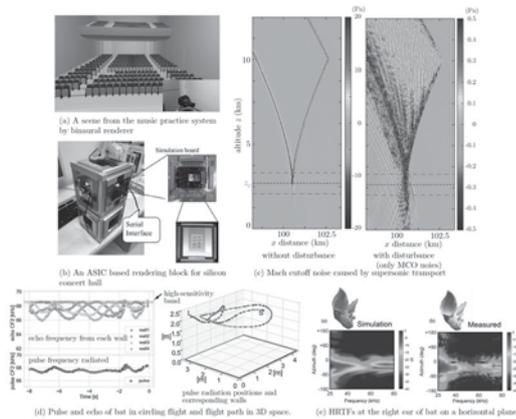


Fig.: Applications of sound field rendering.

### 3-6-7

**3-6-7 並列陽的時間領域 FEM の大規模建築空間のバイノーラル可聴化への応用**

Application of a parallel explicit time-domain FEM to binaural auralization of large-scale architectural space

○吉田卓彌, 奥園健, 阪上公博(神戸大・工学研)

- ◆ 高効率な並列陽的時間領域 FEM (TD-FEM) を用いた室内音響のバイノーラル可聴化法を新たに構築し、2000 m<sup>3</sup> のオーデトリウム内の音場の 5 kHz まで周波数成分を含む可聴化のデモンストレーションを行う。
- ◆ 両耳室内インパルス応答は 1 次アンビソニクス (FOA) と頭部伝達関数のハイブリッド法により生成する。
- ◆ アンビソニクスに必要な球面調和展開係数を有限要素の内挿関数の空間微分にに基づき算出する。
- ◆ 受音点の指向性と室内の音響指標に基づくオーデトリウム内音場の物理的な特徴とバイノーラル可聴化による聴感的な印象の良好な一致を提示し、本バイノーラル可聴化法の妥当性と有効性を明らかにした。

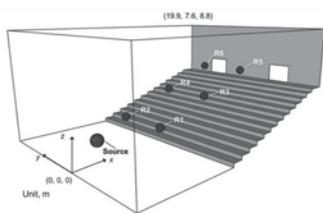


Fig.1 Auditorium model.

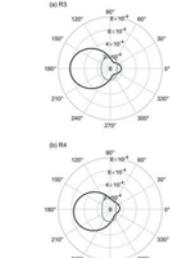


Fig.2 Directivity patterns in early (Black) and late time (Gray) on horizontal plane.

### 3-6-6

**3-6-6 室内音響の高精度かつ高効率な可視化・可聴化のための波動音響解析技術**  
 -PUFEM と時間領域 FEM-

Wave-based acoustics simulation of room acoustics for accurate and efficient visualization and auralization -PUFEM and time-domain FEM-

○奥園 健(神戸大院・工), 吉田 卓彌(安藤ハザマ技研/神戸大院・工),

- ◆ 近年、効率的な音響数値解析手法の開発によって、室内音響モデリングにおいて、波動音響解析を用いてモデル化できる空間の規模や周波数範囲が拡張されるとともに、さらなる効率化手法の開発が試みられている。
- ◆ 高精度なモデリングには室内音響の調整に重要な役割を果たす吸音や拡散を的確に扱うための各種音響材料のモデル化手法の整備も重要な課題である。近年、境界面での反射における位相の変化や材料内部の伝搬まで考慮した、より高度なモデル化に基づく室内音響モデリングの研究が活発であり、よりの確かな可視化・可聴化が可能となりつつある。
- ◆ 様々な波動音響解析手法が存在するが、本稿では、著者らが開発している幾つかの数値解析手法の中から、PUFEM による周波数領域での音響解析手法と陰的時間領域 FEM による音響解析手法について、可視化と可聴化例を簡単に紹介する。

### 3-7-1

**3-7-1 春季における日本海(Japan sea)の音波伝搬**

Sound propagation of NIHONKAI (Japan sea) in spring

○鶴ヶ谷芳昭 (三陽精工), 菊池年晃 (防衛大), 水谷孝一 (筑波大)

- ◆ 日本海は外洋と水深約 200 m の海峡でつながっている。また、冬季の冷却により、密度を増した海水は日本海の底部に沈み、冷水塊である日本海固有水を形成する。春季、対馬暖流の勢力が増大し、冷たい日本海固有水上を北東方向に流れる。したがって、水中音速は増大し、サウンドチャンネル軸深度は深くなり、余剰深度はなくなる。
- ◆ Fig. 1 に春季、大陸側から能登半島サイドまでの音波伝搬を示す。a) は海底が平坦である場合である。大陸側は余剰深度があるため、収束帯伝搬を形成する。また、サウンドチャンネル(SC)軸深度が浅いため、SC 伝搬も形成する。b) は海底地形を付加した場合であり、400 km 付近に大和堆がある。大陸棚を出た音波は大陸傾斜に沿って伝搬するため、SC 伝搬の傾向が強くなり、大和堆を超えて能登半島にまで音波は伝搬する。経路の途中にある大和堆は長距離伝搬に大きく影響する。

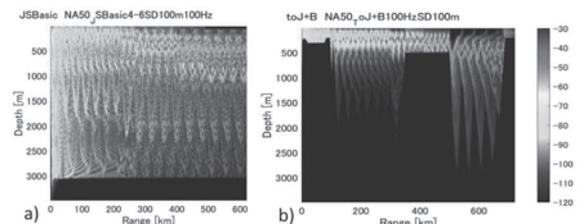


Fig. 1 Comparison of sound propagation without/with bottom topography from continental side a) without bottom b) with bottom SD;100m

### 3-7-2

#### 3-7-2 チャープ信号を用いた時間反転波による水路管変形の検出

Detection of water pipe deformation using time reversal wave of chirp signal

☆山本健介, 太田和彦(金沢工業大)

- ◆水路管の変形の有無を調べるために、チャープ信号を送信波形とした時間反転波を利用した。さらにトーンバースト波を送信波形とした場合と比較し、その違いを確認した。
- ◆一对の無指向性送受波器を用いて、外周 4.4m のダクト管(Fig.1)で実験を実施した。
- ◆収束パルス(Fig.2)に着目すると、トーンバースト波に比べ時間反転波の収束度は高くなることが確認できた。この結果から管内変形の有無の判別もより容易と考えられる。

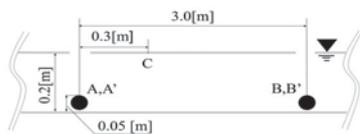


Fig.1 Experimental configuration

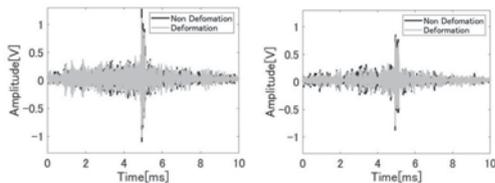


Fig.2 Comparison of waveforms (Left:Chirp, Right:Tone burst)

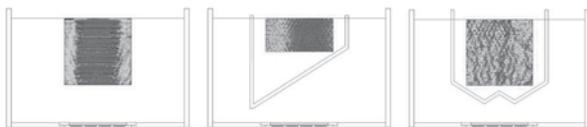
### 3-7-4

#### 3-7-4 ガラス容器を透過する超音波音場分布の計測

Measurement of ultrasonic field distribution transmitted through glass water tanks

○大野正弘(千葉工大), △青木 大(アトック)

- ◆周波数 400 kHz 程度の準高周波超音波洗浄機において、金属製外水槽の内部にガラス製内水槽を懸架する場合、超音波透過率を最適化するための内水槽底板厚みは、垂直入射では 7.5 mm 程度となるが、斜め入射を用いれば 3 mm 程度の板厚で高い超音波透過率を確保することができる。本研究では底板を斜めにした複数のガラス製内水槽を試作し、その内部の超音波音場を小型ハイドロフォンをスキャンする方法で実験的に計測した。
- ◆底面を単純に傾斜させた内水槽では高い透過率を示したが、ビーム位置が傾斜上方にシフトするという結果が得られた。容積を確保するため折れ曲がる形で底面を傾斜させた内水槽では、超音波透過率は理論値よりかなり低い値となった。



(a) No tanks (water only) (b) Slant-bottomed (c) Folded-bottomed  
Fig.1: Measured ultrasonic fields in the inner tanks.

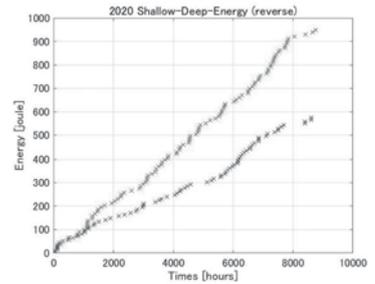
### 3-7-3

#### 3-7-3 紀伊半島南東沖地震の震源構造とエネルギー推移

Hypocenter structure and Energy transition of the Kii Peninsula southeast offshore earthquake

○菊池年晃(防大), 水谷孝一(筑波大)

先に、紀伊半島南東沖地震の震源振動構造を解析した。その際、震源深度の構造は2分性であることが判明した。即ち、この地域で発生したM 1.5以上の地震の震源深度を調べた。縦軸は発生した地震の震源深度を、横軸は元日からの経過時間である。震源深度は50km以浅と380km以深に2分化されている。この分布図は観測されている2002年より今日に至るまで変わらない。したがって、この2分性は紀伊半島の生成に起因した地勢学によるもので、これらの中に付加体が存在する。この付加体は浅震源域と深震源域とで共有しているので二つの震源域で発生する地震に関連性が生じる。この関連性を調べるために浅震源域と深震源域で発生する地震のエネルギー推移を調べた。上図は2020年の累積エネルギーで、浅震源域(x)と深震源域(x)のエネルギーに差が生じたことを表す。深震源域から、即ちプレートからのエネルギーが蓄積され、次の地震の源が形成されていく。これはプレートの挙動と関連しているので地震活動の監視に活用出来る。



### 3-7-5

#### 3-7-5 果物・野菜の縦波, 横波音速

Longitudinal and shear sound velocities in fruits and vegetables

○崔 博坤, △池田 敬 (明治大学)

- ◆果物や野菜では音波減衰量が大きいため、音速測定例は極めて少ない。我々は、ダイコンとスイカの縦波音速測定を 400 Hz—28 kHz で行った。
- ◆ヤング率で決まる伝搬モード、縦波弾性率で決まる伝搬モードの2種類が観測された。
- ◆Fig.1 で示す音速分散曲線の低周波極限はヤング率、高周波極限は縦波弾性率で決まる。得られた縦波、横波音速は、ダイコンで  $C_L=202$  m/s,  $C_S=61$  m/s, スイカでは  $C_L=82$  m/s,  $C_S=36$  m/s であった。
- ◆水中や空気中よりも小さい縦波音速の原因は、試料に数%含まれているガスのせいである。

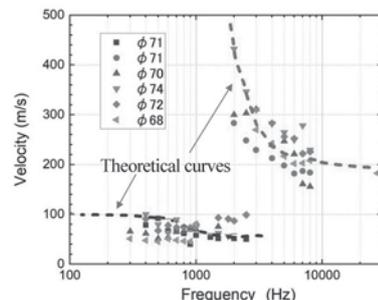


Fig.1: Dispersion of sound velocity measured in daikon cylinder. Symbols indicate the measured values with various diameters.

### 3-7-6

#### 3-7-6 熱音響システムの軽量化に向けた基礎検討 —局所的な管材質変化が熱音響プライムムーバーの性能に与える影響—

Fundamental study on weight reduction of the thermoacoustic system -Effect of local change in a tube material on the thermoacoustic prime mover performance-

☆宇治野 勝幸, 坂本眞一 (滋賀県立大)

- ◆ 熱音響システムの軽量化に向けた検討として、熱音響プライムムーバーを構成する管材質の変化を行った。
- ◆ 熱音響プライムムーバーの局所的な管材質変化を行うために、塩化ビニル樹脂で作成した管(塩ビ管)を用いた。(Fig. 1)
- ◆ 塩ビ管の長さを変更し、ループ管型熱音響プライムムーバーの音場に与える影響を検討した。
- ◆ その結果、管材質変化の有無や塩ビ管の長さの違い(100,200 mm)によって、音響インテンシティが変化した。(Fig. 2)

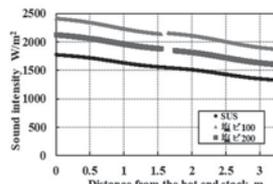
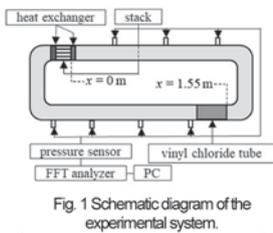


Fig. 2 Comparison of Sound intensity.

(K. Shiraki et al., Mech. Eng. Res. Dev. 43, 91, 2020.)

### 3-7-7

#### 3-7-7 同軸型熱音響システムの 発振条件探索に向けた基礎検討 —ループ管型熱音響システムの 安定性解析による試み—

Fundamental study for searching oscillation conditions of coaxial thermoacoustic systems -trial by stability analysis of loop-tube thermoacoustic systems-

☆小野 悟, 坂本眞一 (滋賀県立大)

- ◆ 同軸型熱音響システム(同軸型)の安定性解析をループ管型熱音響システム(ループ管型)のモデル化によって試みた。
- ◆ 同軸型管内の音波の流れに着目し、ループ管型モデルを作成した。[1]
- ◆ 外管と内管の流路断面積差を考慮するか否かで、異なる2種類のループ管型モデルを用意した。
- ◆ 安定性解析の結果、外管と内管の流路断面積差を考慮しない場合、発振温度の低下と周波数の上昇を確認した。
- ◆ 今後、同軸型の発振実験による検証やシステム条件を変化させ、同軸型の安定性解析に適したモデルの選定が必要だと考える。

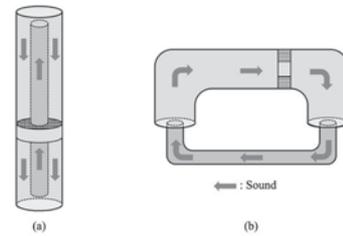


Fig. 1 Sound waves in the tube. (a) Coaxial type. (b) Loop-tube type.

[1] Y. Takeyama et al., Jpn. J. Appl. Phys. 57, 07LE14 (2018). [2] R. Onishi et al., Jpn. J. Appl. Phys. 60, S00002 (2021).

### 3-7-8

#### 3-7-8 薄いメダル状 Phase Adjuster の 断面積変化が熱音響プライムムーバーの 発振温度に与える影響

Effect of Cross-Sectional Area Change of Thin Medal-shaped Phase Adjusters on the Onset Temperature of a Thermoacoustic Prime Mover.

☆松本拓己, 坂本眞一 (滋賀県立大)

- ◆ 先行研究では、熱音響システムの発振温度を低下させるデバイスとして Phase Adjuster (PA) を提案した。
- ◆ 発振温度の低下を促した PA の物理的要因を調査するために、PA 端面のみを表現した薄いメダル状の PA を2枚用いて見かけの PA を構成し、熱音響プライムムーバー内に設置した。(Fig. 1)
- ◆ メダル PA の内径を変更し、熱音響プライムムーバーの発振温度を実験的に調査した。
- ◆ その結果、メダル PA の内径ごとに発振温度が変化したことから、PA の端面が発振温度に作用したことを示した。(Fig. 2)

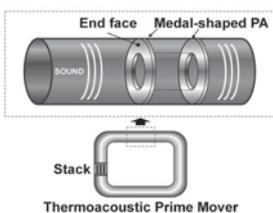


Fig. 1 Schematic of medal-shaped PA. (K. Sahashi et al., Proc. Acoust. Soc. Jpn., 2010, 3-2-2)

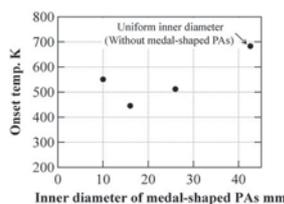


Fig. 2 Onset temperature for the inner diameter of the medal-shaped PAs.

### 3-7-9

#### 3-7-9 液体ピストン型熱音響エンジンへの双方向 タービンの実装と特性評価

Implementation and characterization of bi-directional turbine in liquid-piston type thermoacoustic engine

☆富平 嗣大(東北大), △庄司 衛太(東北大), 琵琶 哲志(東北大),

△高尾 学(松江高専)

- ◆ 液体ピストン型熱音響エンジンを比較的低温の熱源で動作可能な発電機へと発展させるために、衝動型の往復流型タービン(双方向タービン)を組み込んだ装置を作成した(Fig. 1)。
- ◆ タービンを気柱部分ではなく液柱部分に挿入したことで、タービンを回転動作させることに成功した。

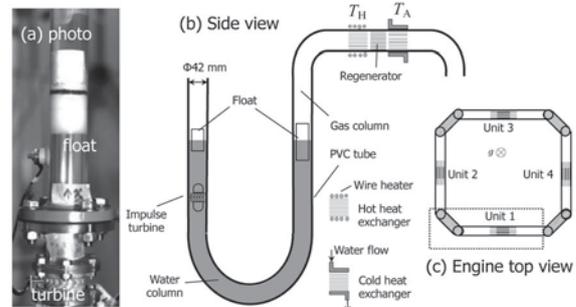


Fig. 1: (a) Photo of the impulse turbine and float installed in the liquid column. (b) Side view of the unit section of a liquid-piston type thermoacoustic engine. (c) Engine top view. The engine starts to operate with the heating temperature of 110°C. The time-averaged rotation speed of the turbine increases to 650 rpm by further heating.

3-7-10

3-7-10 予混合燃焼器における自励振動に対する線形安定性解析

☆高島耀介, △笠井湧斗, △庄司衛太, 琵琶哲志(東北大)

- ◆線形安定性解析によって予混合燃焼器における振動モードの励起を評価した。
- ◆音響の伝達行列法に、火炎伝達関数を用いて導いた火炎の伝達行列を加えて、系の特性方程式を求め、解 $\omega$ の虚部に注目することで、振動モードの安定性を予測した。
- ◆燃料の発熱量の変化に対する解の応答から、火炎の影響を受け、各振動モードの安定性がそれぞれ異なる変化をみせることが示唆された。

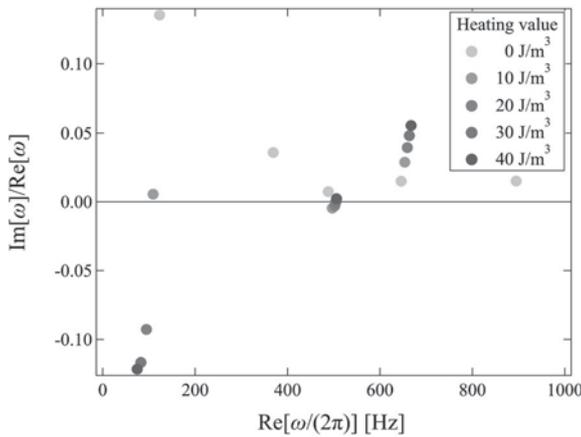


Fig. 1 The ratio of the imaginary to real parts of the eigenfrequencies obtained for five heating values of gas (0-40 J/m<sup>3</sup>).

3-10-2

3-10-2 油量及び食材の違いが揚げ音に与える影響

Effect of heating equipment and ingredients on frying sound

☆大沢公亮(神奈工大), 杉山洋介(神奈工大院/横浜創英大), 油井あまね, 上田麻理(神奈工大)

調理の際に発生する音は、調理をする上で重要な情報となっている。その音は、食材の水分含有量などの物性、油の温度などにより複雑に発生している。

調理のプロフェッショナルと管理栄養士養成学科の学生とでは、いったい音に関してどのような意識や感覚の違いがあるのか、調理中に発生する細かな音の違いに気が付けるのかを調べるため、調理支援アプリケーションを用いた調査と、天ぷら店へのインタビューを行った。調理支援アプリケーションの概要とともに、調査結果と今後の調理支援アプリケーションの活用可能性について報告する。



Fig.1: Screen of cooking support application

3-10-1

3-10-1 吸音材の導入による保育者の音環境の認識とその変化

-療育現場におけるアクションリサーチ-

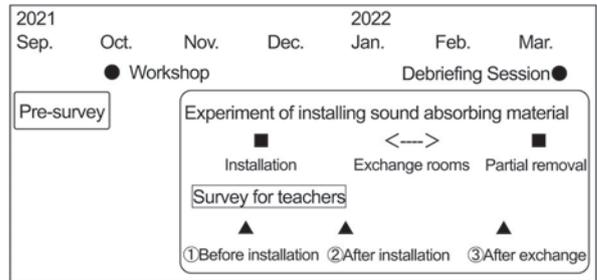
Teacher's perspective on the classroom acoustic environment and its change with sound absorption:  
Action research in a child development support center

◎野口紗生(明大/ADC), △松本知子(根洗学園),

△渡辺涼子(根洗学園/浜松学院大), 上野佳奈子(明大), 船場ひさお(横国大/ADC)

- ◆療育現場との協働で取り組んだ吸音性能向上実験 (Tab. 1) の中から、保育者の音環境の認識とその変化に着目して分析した。
- ◆保育室に吸音材を導入し、平均吸音率 0.18→0.30 (500, 1kHz 帯域の平均値) へと変えたところ、音の違いは認識されやすかった。
- ◆一方で、それが保育にとって持つ意味として理解されるには、保育者自身の問題意識や、部屋交換など環境変化の経験、また響きについて他の保育者や音の専門家と意見交換をするなど、音の視点を保育とつなげて考えるきっかけが必要であることが考えられた。

Tab. 1 Experiment flow



3-10-3

3-10-3 学生の言語コミュニケーション能力を高める学習教材の開発 -人のロボット化阻止に向けた情報学部における教育-

Proposal for a learning material to improve students' verbal communication skills - Education to stop the human robotisation in the Faculty of Informatics -

○上田麻理, 春日秀雄, 西口磯春, 松本一教, 田中哲雄(神奈川工科大)

タイトルの「ロボット化」という言葉は、「主体的に考えようとする」「機械的に処理する」と言った古典的な意味での常套句(クリシェ)として用いている。最前線のロボット研究者が目指しているであろう「ロボット」の概念を念頭に置いているわけではない。



Fig. 人間のロボット化阻止に必要な教育の実践と提案を行う

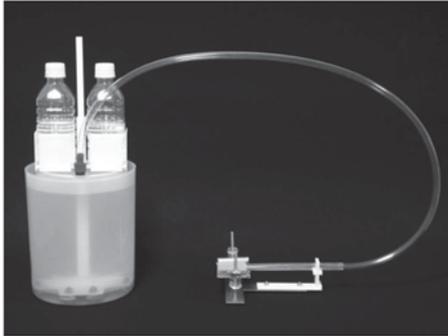
### 3-10-4

#### 3-10-4 声帯模型の音響分析に関する一考察

A Study on Acoustical Analysis of a Functional Vocal Cords Model

○青木 直史 (北大)

模型教材の活用は五感に対する訴求力に優れており、初学者の興味を引きつけるアプローチとして教育的効果が高い。本研究では、これまで検討事例が少なかった声帯をターゲットとして、機能に焦点をあてた実体モデルの作成を試みている。本発表では、実際に作成した声帯模型の特徴を把握するために行った音響分析の結果について報告する。



声帯模型に airflow を安定して注入する装置



デモ動画 : <https://youtu.be/WBn11kaXnA>

### 3-10-5

#### 3-10-5 「使える音響学」にする試み

Attempt to make it "usable acoustics"

○竹内京子(順天堂大), 青木直史(北大), 荒井隆行(上智大), △鈴木恵子(北里大), 世木秀明(千葉工大), △秦若菜(北里大), 安啓一(筑波技術大)

- ◆ことばのリハビリを行う言語聴覚士という職業の養成校では、音響学(聴覚心理学を含む)が必修科目である。そして、音響学は、学生が最も苦手と嫌いな科目である。
- ◆音響学の教師は、授業のカリキュラム作成の基準となる「国家試験出題基準」と「言語聴覚士テキスト」を参照し、実際にこれらの知識を臨床で使うであろうという前提で授業を行なっている。
- ◆しかしながら、現役の言語聴覚士の臨床では、「音響学」の知識が、ほとんど活用されていない。つまり、音響学の授業の目的が曖昧である。
- ◆「音響学を学ぶ目的」がないのならば、言語聴覚士の現状の環境で実行可能なことをゼロから探し、それを広める活動を始めた。本発表は、その経過報告と、現在分かっている問題点の紹介である。
- ◆これらの活動の目的は、臨床現場で「音響学を使う」ことはもちろんであるが、それだけではない。最終的には、養成校の学生に「音響学を学ぶ目的」の実例を示すことである。
- ◆教材作成や授業の工夫とともに、「使える音響学」を示す活動は、学生の学びを助ける大きなポイントではないだろうか。
- ◆今後は、言語聴覚士の臨床現場にある雑音環境でも使える録音・分析のための技術を探すとともに、何がどこまで使えるかを探究していきたい。さらに、録音・分析には、高価な機材が一般的であるが、できる限り安価な方法を追求していきたい。

### 3-11-1

#### 3-11-1 EV 加速を想定した周波数掃引音におけるマスキング特性

Masking characteristics in frequency sweep tone during EV acceleration

☆柴橋和真, 金澤立哉(中央大院), △金堂雅彦, 戸井武司(中央大)

- ◆本研究は、EV 加速走行時に車室内で知覚されるモータ次数による周波数掃引音(以下、FST : Frequency Sweep Tone)において、暗騒音からの可聴突出量(以下、 $L_p$ )に基づくマスキングラインを解明した。そして、定常音(以下、ST : Steady Tone)と FST とのマスキング特性の差異を明確化した。
- ◆~500 Hz では、100 Hz/s を超える FST における  $L_p$  は、ST における  $L_p$  と比較して有意に低減(\* $p < 0.05$ )し、そのレベルは 4~5 dB 程度である。高速掃引では、時間的な周波数移動により ST と差異が生じ、聴覚による知覚感度が高まる。
- ◆500 Hz~では、Fig. 1 より、周波数掃引速度の上昇に伴い、FST マスキングラインが高周波側へと移動する。高周波領域では聴覚の臨界帯域に基づき周波数分解能が低下するが、高速掃引においては単位時間当たりの周波数移動量が大きく、高周波領域でも周波数掃引を知覚しやすいためである。
- ◆10000 Hz 付近の高周波領域では、全 FST マスキングラインが ST マスキングラインに収束する傾向を見出した。聴覚の周波数分解能が低下するためである。

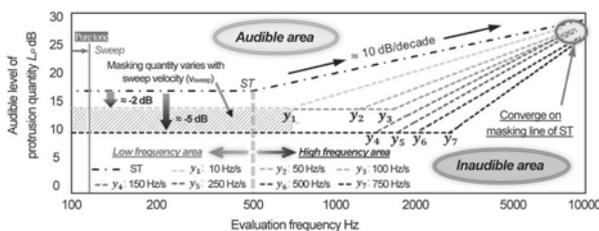


Fig. 1 Masking characteristics of Steady Tone (ST) and Frequency Sweep Tone (FST) by background noise

### 3-11-2

#### 3-11-2 視覚刺激を考慮した EV 車室内の周波数マスキングによる純音認知評価

Pure tone cognition evaluation by frequency masking in EV interior taken into consideration of visual stimuli

☆金澤立哉, 柴橋和真(中央大院), △金堂雅彦, 戸井武司(中央大)

- ◆本論文では、視覚からの負荷を主観と客観評価の両面より把握し、主観と客観の対応を明らかにする。その後、視覚刺激の負荷条件の差異がマスキングに与える影響を解明する。
- ◆初めに、提示する視覚刺激の負荷を、心電情報を基にしたローレンツプロット(LP)より評価する。提示する刺激は単刺激(走行音のみ)、複合2刺激(走行音+走行映像)である。複合2刺激の中でも、対向車の有無により視覚刺激の負荷を変化させた。条件別の評価結果を Fig. 1 に示す。刺激負荷に伴い、リラクセス度指数の低下を確認しストレスを感じていることが確認された。
- ◆同様の映像で可聴突出量の測定を行う。刺激数ごとに負荷の違う2条件を用いる。結果を Fig. 2 に示す。単刺激の条件(▲, ●)ごとでは差がないのに対し、開眼と複合2刺激(■, ◆)では平均 6 dB 程度の増加が見られ、視覚負荷の増加に伴う可聴突出量の増加が確認された。

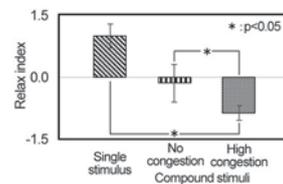


Fig. 1 Result of LP

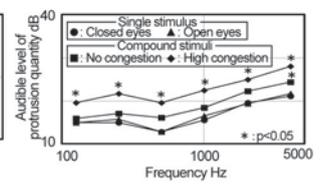


Fig. 2 Comparison of audible level of protrusion quantity for single and compound stimuli

### 3-11-3

#### 3-11-3 自動運転中に提示する警告音の吹鳴周波数上昇が運転引継ぎ行動に及ぼす影響

The influence of the rise in the blowing frequency of the warning sound on the behavior at take-over

☆速水亮(中央大), △阿部晃大, 加藤洋子,

関根道昭(交通安全環境研究所), 戸井武司(中央大)

- ◆自動運転から手動運転への引継ぎを行う際、吹鳴周波数が運転引継ぎを知らせる警告音の緊急感、運転を引継いだ後の操作に影響を与えることが示唆されている。低い吹鳴周波数では引継ぎ時のドライバの動揺が少ないものの引継後の車両の挙動が安定せず、高い周波数では引継ぎ後の車両挙動が安定するが、ドライバが警告音に対して驚愕感を抱く。本研究では低い吹鳴周波数から高い吹鳴周波数へ切り替わる音を警告音として使用し、両者の利点の両立を目指す。
- ◆吹鳴周波数 2 Hz 一定、6 Hz 一定、そして 2 Hz で鳴り始め、2 秒後に 6 Hz に上昇する音の 3 種類を用いて運転引継ぎ実験を行い、実験後に音に対する印象を評価させた。また、引継ぎ後の走行データから、前後の移動量に対する左右移動量を算出した。Fig. 1 に示す主観評価では 2 Hz、2.6 Hz、6 Hz の順で緊急感が高くなるが、Fig. 2 に示す左右移動量は自動運転中に行う運転以外の作業負担により、異なる傾向を示した。
- ◆負荷の低い動画課題の場合、2 Hz で鳴り始め、2 秒後に 6 Hz に上昇する音が両者の利点を両立し、有効な手段になることが示唆された。

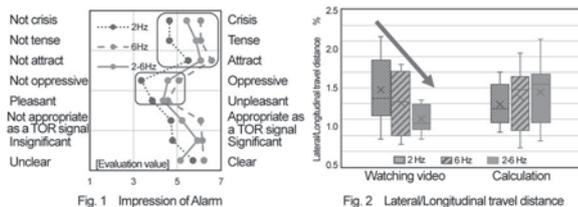


Fig. 1 Impression of Alarm

Fig. 2 Lateral/Longitudinal travel distance

### 3-11-5

#### 3-11-5 変速段ごとの変速サウンドデザインによる加速感向上

Improving acceleration by designing gearshift sound for each gearshift stage

☆五十嵐啓祐(中央大院), △花井奏太, 戸井武司(中央大)

- ◆内燃機関車(ICEV)および Fig. 1 に示す電気自動車(EV)走行音の違いを考慮し、変速段ごとに適切な変速サウンドを付加することで加速感を向上させる。
- ◆音圧は、ICEV、EV 走行ともに変速段ごとに増加する傾向の変速サウンドで加速感が向上する。また、ICEV 走行では走行音の音圧同等の狭い増加幅、EV 走行では走行音の音圧以上の広い増加幅の変速サウンドで加速感が向上する。
- ◆周波数範囲は、ICEV、EV 走行ともに変速段ごとに広がる傾向の変速サウンドで加速感が向上する。また、ICEV、EV 走行ともに走行音の主要周波数範囲より広い変速サウンドで加速感が向上するが、ICEV 走行では周波数範囲が広すぎると不快感につながる。
- ◆変速段ごとに音圧と周波数範囲が変化しない初期条件と良条件、悪条件を組み合わせて検証実験を行い、Fig. 2 に示す悪条件で加速感が低下し、良条件で加速感が向上することが確認できた。

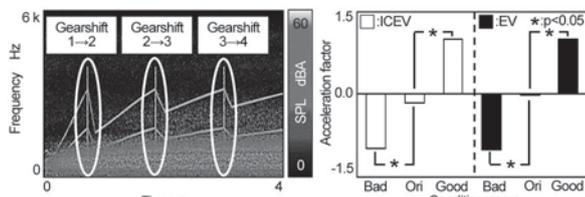


Fig. 1 EV driving sound

Fig. 2 Verification experiment

### 3-11-4

#### 3-11-4 自動運転中の車内環境における BGM の音色がドライバの反応時間と主観的印象に及ぼす影響

Effects on driver reaction time and subjective impressions by the timbre of background music in autonomous vehicle environment

☆尾崎真弘, 星野博之(愛知工業大学)

- ◆目的 本研究では、自動運転中の車室内を模した環境下で音色が異なるクラシック音楽を聴取しながら視覚反応タスクを行い、ドライバを快適かつ覚醒度の高い状態に保つ音楽を明らかにする。
- ◆方法 防音室内において高速道路走行中の映像と、ピアノ音の原曲(2個)またはその原曲から高周波成分を強調・減衰した音データ(第一実験:6個、第二実験:6個)を10分間流し、反応時間のタスク評価と、「眠い目覚めている」などのSD法による印象評価を12項目に対して行う。
- ◆結果 第一実験の反応時間グラフ(被検者12名の平均)を図1に示す。また、No.5の経過時間と反応時間の推移を図2に示す。第一実験ではシャープネス値の高い音データに対して覚醒維持効果を高める結果が得られた。

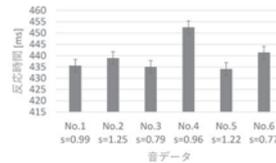


図1: 音データと反応時間 (s: シャープネス値 [acum])

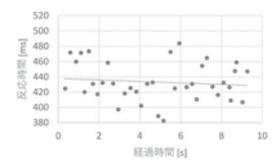


図2: No.5の経過時間と反応時間の推移

### 3-11-6

#### 3-11-6 気筒構造の差異に着目したディーゼルエンジンの快音化

Comfortable sound design of diesel engine combustion noise focused on differences in cylinder structure

☆武田幸広(中央大院), △藤齋晴輝, 戸井武司(中央大)

- ◆近年のディーゼルエンジンは燃費向上や排ガス規制が進む一方、燃焼加振力の増加やエンジン構造の振動特性の変化により、燃焼音の大きさだけでなく、音質も着目されている。そこで本研究では、ディーゼルエンジン燃焼音の変調度(DM)やラフネスといった変調成分に着目した快音設計を試みる。

先行研究では、ホワイトノイズを用いて疑似エンジン音を作成し、15~35 Hzの変調周波数が快適感に影響することを示している。本研究では、衝撃感に着目し、印象変化を Fig. 1 に示す。変調周波数を変化させることで、衝撃感に関する感度が10~30 Hz付近にあることを把握した。また、変動感成分の低減を狙い、CAEにてエンジンの構造変更を行い、構造変更前後の物理量を Fig. 2 に示す。変更前より変調度1次成分やラフネスを低減することができ、Fig. 3 に示すように燃焼音の快適感が向上した。

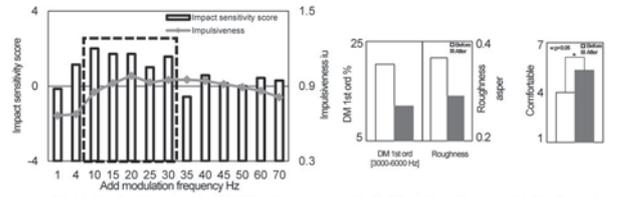


Fig. 1 Results of impact sensitivity(1000 rpm)

Fig. 2 Physical quantity

Fig. 3 Structural modification

### 3-11-7

#### 3-11-7 新しい時代の Driving Experience を支える 自動車のサウンドデザイン

Vehicle Sound design for new era

○桑田敏久, △金原俊一, 生越渉, 後藤昌也(日産自動車)

- ◆ 昨今、車の電動化や騒音規制強化等により車の音環境は大きく変化し静音化が進んでいる。一方、車は単なる移動手段としての役割だけではなく運転を楽しむ等、人生を豊かに過ごすパートナーとしての役割もある。
- ◆ 加速音は、車を運転する楽しさに大きな寄与を持ち、運転操作に合ったサウンドデザインが求められる。
- ◆ 従来、基本的には内燃機関から発せられる音を主にデザインしてきたが、新たな手段として 1980 年代後半に電氣的に発生させた音で騒音を打ち消すアクティブ騒音制御を開発し、その後、将来の電動化社会や車外騒音規制強化の動向も見越し、消音だけでなく加音することも制御範囲に加え進化を続けて来た。
- ◆ 本技術はこれまで黒子技術としてきたが、昨今、車両の電動化や騒音規制が運転する喜びをスポイルするのではないかといった懸念をお持ちの方もいらっしゃることもあり、本技術を適用することで、走る喜びを高めた車両の適用事例について紹介する。



Fig.1:Applicable model with new sound design (ICEV & BEV)

### 3-11-9

#### 3-11-9 車内騒音・車外騒音低減のための タイヤ技術開発動向

Tire technological development trend for reduction of interior and exterior noise

○齊藤由典(日本ミシュランタイヤ)

近年の自動車のパワートレイン電動化に伴う静粛性の向上や、環境騒音としての道路交通騒音による健康被害への関心の高まりから、タイヤと路面との接触により生ずるタイヤ路面騒音低減の要求は益々高くなっている。一方で葉書1枚程度の面積のみで接地し、あらゆる路面環境で車両の安全な走行を確保しながら、環境負荷を低減しなければならないタイヤには騒音性能以外にも多くの相反性能がある。本稿ではタイヤ路面騒音の発生機構を概説し、その低減技術開発動向を紹介する。

タイヤ路面騒音の発生機構は加振、応答、伝播の3つの系に分類して考えられる。加振系としてはトレッドパターンによる加振と路面粗さによる加振が主に存在し、タイヤの技術開発としてはトレッドパターンによる加振を最適化することで低騒音化を図る。トレッドパターンの最適化には様々な手法が存在するが、基本的な低騒音化技術の他、車内騒音による不快感を低減するための考え方を示す。

また、同一タイヤであっても車両によって車内騒音のレベル、音質も変化することを示し(図1)、シミュレーションを用いた今後のタイヤ技術開発に関しての展望を述べる。

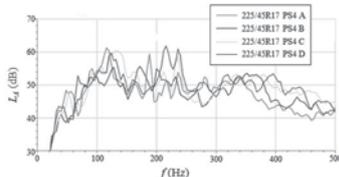


Fig.1:Interior noise spectrum measured in four different vehicles.

### 3-11-8

#### 3-11-8 機械振動のエネルギー伝達特性に基づく音 デザインの可能性

Possibilities for sound design based on energy transmissibility of mechanical vibration

○山崎徹(神奈川大・工)

- ◆ 「音デザイン」≡エンジンサウンドなど音の強調と音の付加を連想。固体音(機械騒音)は騒音(ノイズ)、音量の低減が先決。
- ◆ 音振問題の二段階設計の紹介(図1)。基線の制御=第一段階設計(素性のよい設計)+ピークの制御=第二段階設計(味付け設計)。
- ◆ 第一段階設計をベースとした多性能適正化(形で考えない研究会)の取り組み紹介(図2)。
- ◆ 第一段階設計を基にした「音デザイン」の考察・議論題材の提示。

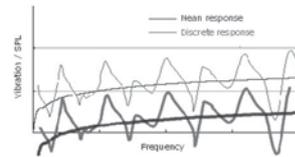


Fig.1 Two-step design concept for low noise and vibration

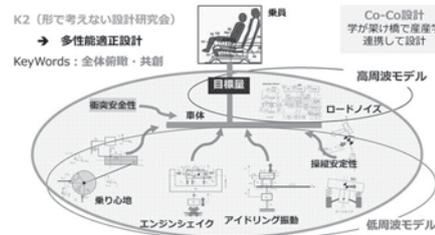


Fig.2:An approach to multi-performance appropriate design in conceptual design

### 3-11-10

#### 3-11-10 インバータ駆動モータの音質改善技術

Sound quality improvement technology of motor driven by inverter

○渡部真徳, 田部洋祐, △原崇文(日立)

- ◆ カーボンニュートラル(CN)への流れに伴い EV(電気自動車)の普及が進んでおり、モータ音の音質改善を含む静粛性の要求が高い。
- ◆ PWM インバータ駆動のEV用モータドライブユニットを対象として、モータ音の発生・伝搬メカニズム解明と音質改善に取り組んだ。
- ◆ 構成要素同士を瞬時エネルギーで繋ぐ物理機能モデルを活用し、加振力と加振源から騒音評価点までの振動騒音の伝搬経路を明確にした。
- ◆ 騒音の分析結果から、モータに起因する電磁音とインバータに起因するキャリア音が支配的であることを明らかにした。
- ◆ キャリア音の改善手法として、ランダムキャリア制御を提案し、シャープネスとラフネスにおいて音質改善を確認した(Fig.1, Fig.2)。

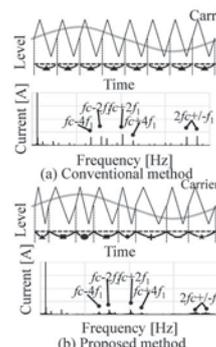


Fig.1 Random carrier control.

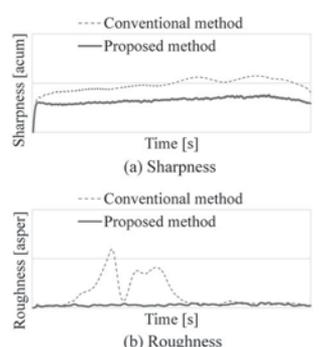


Fig. 2 Sound quality evaluation result.

### 3-4P-1

#### 3-4P-1 骨伝導における耳栓効果が単音節明瞭度に及ぼす影響: 話者性別の比較

Influences of the occlusion effect with bone-conduction on monosyllable articulation: An effect of speaker gender

☆美和 あす華, 大塚 翔, 中川 誠司(千葉大)

- ◆外耳道を閉塞した状態で骨伝導音を聴取すると、低域のラウドネスが増大する“耳栓効果”が生じる。
- ◆著者らは頭部の様々な部位への骨伝導刺激に対する耳栓効果を調べてきた。一部の顔面部位(特に鼻骨(Nasal)や眼窩下部(Infraorbital))は、従来部位(乳様突起(Mastoid)、顎状突起(Condyle)等)を凌ぐ耳栓効果を示すことを明らかにしている。
- ◆また、前報では、乳様突起、顎状突起、鼻骨および眼窩下部において、耳栓効果が女声の単音節明瞭度を上昇させることを明らかにした。
- ◆本研究では、前報と同じ部位において耳栓効果が女声および男声の単音節明瞭度に及ぼす影響を評価し、比較した。
- ◆女声および男声、ともに耳栓装用により明瞭度は上昇したが、男声の方が明瞭度の上昇が大きかった。構成周波数が女声よりも低いため、低周波で顕著に見られる耳栓効果の影響が大きかったと考えられる。

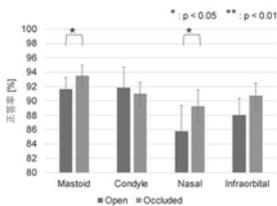


Fig. 1 Percent correct of the articulation test with female voice (Mean ± S. D.).

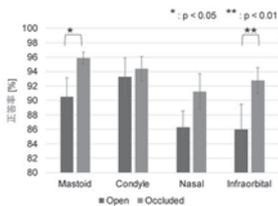


Fig. 2 Percent correct of the articulation test with male voice (Mean ± S. D.).

### 3-4P-3

#### 3-4P-3 EEGNetを用いた脳磁界データによる想起時の音声識別の検討

Sound classification of brain magnetic fields using EEGNet

☆山名莉央, 北条直樹, 矢野肇, 滝口哲也,

高島遼一(神戸大), 中川誠司(千葉大)

- ◆従来のBCI開発では、自由度の高い意思伝達はできない、ユーザの意思をより汎用的に認識するBCIの実現に向け、脳磁界データから音声想起の識別を試みた。
- ◆脳磁界データにおける音声想起識別が困難な理由として、音声想起時の脳活動には明らかな音響特徴の鈍化・欠落が生じており、単語を識別するのは容易ではないことや、被験者実験を伴うデータ計測が長時間行えないことにより、データ数が極端に少なくなってしまうことがあげられる。
- ◆そこで本研究では、脳磁界データにおける想起音声識別の精度を向上させるため、神経生理学的信号から解釈可能な抽象的特徴を抽出するだけでなく、少ないデータセットで信頼性の高いモデルを構築することが可能であるEEGNetを用いて音声想起識別を行った。

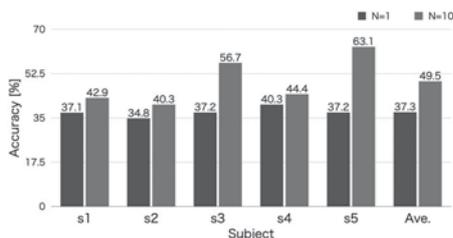


Fig. 1: Discrimination accuracy of the imagined sounds.

### 3-4P-2

#### 3-4P-2 顔面に呈示された骨伝導音の音声伝達特性: 単音節明瞭度試験による評価

Characteristics of speech perception by bone-conduction presented to the human face: Assessments by a monosyllable articulation test

☆上村 昂, 大塚 翔, 中川誠司(千葉大)

- ◆骨伝導では一般に側頭骨の乳様突起や下顎骨の顎状突起に刺激呈示されるが、近年は鼻骨、頬骨といった顔面への呈示も検討されている。
- ◆顔面は複雑な構造を持っており、振動子の呈示位置のわずかな変化により骨伝導音の知覚・伝搬メカニズムが変化する可能性が高い。
- ◆顔面における骨伝導の音声の伝達特性を明らかにするため、鼻骨、眼窩下部(上顎骨)、頬骨、下顎角(下顎骨)、およびオトガイ隆起(下顎骨)に呈示された骨伝導音による単音節明瞭度試験を行った。
- ◆頬骨および下顎角では従来部位の骨伝導音とおおよそ同等の明瞭度が得られ、特に頬骨は気導音と比べても大きな低下が確認されなかった。
- ◆鼻骨、眼窩下部およびオトガイ隆起では有声子音の異聴が増加する傾向があった。これらの部位は聴覚器から離れており、頭部内を伝搬する過程で高周波の減衰を含む比較的大きな歪が生じた可能性がある。

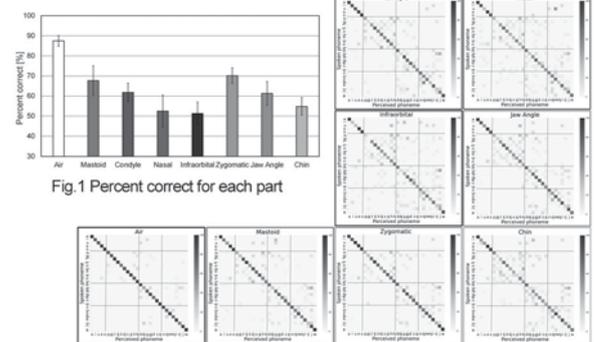


Fig. 1 Percent correct for each part

Fig. 2 A confusion matrix for each part

### 3-4P-4

#### 3-4P-4 うつ症状検出に適した発話タスク選定に向けたうつ病患者の音声分析

Analyses of Utterance Tasks for Automatic Depression Detection from Speech.

○村田有実子, 安藤厚志(NTT)

- ◆音声からうつ症状の有無を自動検出するうつ症状検出においては、検出モデルの改良だけでなく、うつ症状の特徴が表れやすい特定の発話課題(発話タスク)の選定が検出精度向上において重要であるが、先行研究においてその比較は十分に行われていない。
- ◆本研究では、うつ症状検出に適した発話タスクの選定に向け、発話タスクごとの音響特徴量の統計分析を行った。従来・提案の発話タスクに対し、同一の音響特徴量を用いて分散分析を行い、うつ病有無の項目に有意差があった特徴量数を集計した。
- ◆精神科への通院歴をもつ話者を用いた評価実験の結果、自由発話、感情付き文読み上げ、一人しりとり、音韻繰り返し、早口言葉、単語連続ゲームにおいて有意差がみられた特徴量数が多く、これらの発話タスクがうつ症状検出に有効である可能性が示唆された。

Table 1: Numbers of acoustic features (spectrum / time features) for each utterance task, which had the effect of depression or the interaction of gender.

	感情	スペクトル特徴		時間特徴	
		#うつ	#相互	#うつ	#相互
文読み上げ		0	1	1	0
段落読み上げ		1	5	1	0
自由発話		4	1	2	2
音韻繰り返し		2	0	7	0
感情付き文読み上げ	hap	3	0	3	0
	sad	2	0	1	0
	ang	3	3	2	0
感情付き文読み上げ(平静発話差分)	hap	3	0	1	0
	sad	3	0	0	1
	ang	2	3	0	0
早口言葉		1	0	5	1
単語連続ゲーム		0	3	6	0
一人しりとり		3	5	3	1

### 3-4P-5

#### 3-4P-5 頭皮上で検出される骨伝導音声の明瞭度 : 子音種類毎の検討

Intelligibility of bone-conducted speech detected on the scalp : Examination at each consonant type

☆南里聡志, 大塚翔, 中川誠司(千葉大)

- ◆骨伝導マイクロホン (BCM) は周囲騒音を検出しにくいという特長を持つが、装着性に問題があった。対して我々は、BCM をヘルメットに内蔵させた新型骨伝導デバイスを提案した。また、その最適化に必要な、頭皮上から検出される骨伝導音声 (頭皮上骨伝導音声) の特性解明に取り組んでいる。
- ◆前報では、頭皮上骨伝導音声の調音素性情報の伝達率を推定し、前頭部 (Fpz)・頭頂部 (Cz)・頬骨突起 (Che) では子音の平均伝達率が高い傾向を見いだした。しかし、子音種毎の明瞭性は明らかでない。
- ◆前報の明瞭度試験の結果を用いて、子音種毎の明瞭度を算出した結果、接近音を除く各子音種で検出部位間の差異が確認された。
- ◆例えば、摩擦音の明瞭度は他の子音種に比べ低下したが、前頭部 (Fpz) と頭頂部 (Cz) では他部位より高い値が得られた。これらの部位では、高周波成分が減衰しにくいことが示唆される。

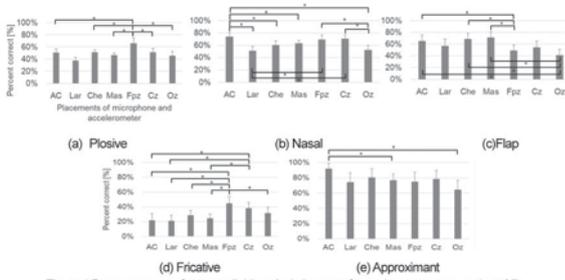


Figure 1 Percent correct of mono-syllable articulation tests for each consonant type (p < .05).

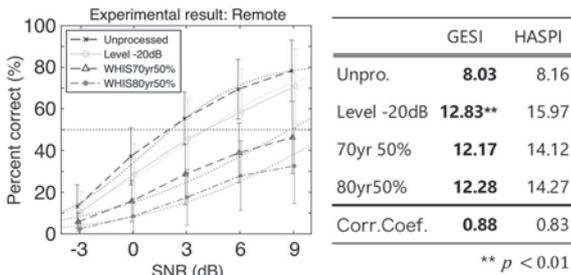
### 3-4P-7

#### 3-4P-7 客観評価指標 GESI による 模擬難聴音声了解度の個人別予測

Intelligibility prediction of simulated hearing loss sound from individual subject's data using GESI

◎山本絢子(和歌山大院), 宮崎美紀, 田丸萌夏, 入野俊夫(和歌山大)

- ◆高齢難聴者個人ごとの音声了解度予測を目指して、客観評価指標 Gammachirp Envelope Similarity Index (GESI) を開発
- ◆模擬難聴音声の主観評価実験を防音室とクラウドソーシングで実施 + GESI で原音声から個人別に予測できるか検証・従来手法と比較
  - 聴取環境 (防音室とクラウドソーシング) による了解度の違い
    - 提案手法 GESI : 予測可能 / 従来手法 HASPI : 予測困難
  - 話者性別 (男声と女声) による了解度の違い
    - 提案手法 GESI : 予測誤差小 / 従来手法 HASPI : 予測誤差大



Speech intelligibility (%) of simulated hearing loss sounds. Figure: Intelligibility of male speech in the crowdsourced experiment, Table: Mean of RMSE across the subjects and correlation coefficient between subjective score and prediction.

### 3-4P-6

#### 3-4P-6 振幅変調方式の違いが 遠位呈示骨導超音波の単音節明瞭度に 及ぼす影響の検討

The effect of amplitude-modulation methods on monosyllable articulation by distantly-presented bone-conducted ultrasound

☆高橋尚也, 大塚翔(千葉大), 中川誠司(千葉大)

- ◆骨伝導で呈示された 20 kHz 以上の高周波音 (骨導超音波) は、音として明瞭に知覚される。また、振幅変調することで音声の伝達が可能である。
- ◆骨伝導音は一般的に乳様突起等に呈示されるが、骨導超音波は頭部から離れた位置に呈示 (遠位呈示) された場合も明瞭に知覚可能である。
- ◆遠位呈示では受容器までの距離が長く複雑になることから、伝搬経路上に存在する生体組織の非線形性によって生じる自己復調成分が従来の呈示方法よりも発生しやすくなると考えられる。
- ◆自己復調成分に影響を与える振幅変調方式および音声の話者性別が、遠位呈示骨導超音波の間こえに与える影響を評価するため、単音節明瞭度を測定した。
- ◆受容器から遠い部位ほど正答率が低下した。また、DSB-SC, SSB, Transposed, DSB-TC の順に正答率が上昇した。さらに、全ての呈示部位、変調方式に対して、女声に対して男声よりも高い正答率が認められた。
- ◆遠位部位における変調方式の効果は、乳様突起のものと同様であった。一方、乳様突起では明確に認められなかった話者性別の効果が観察された。自己復調特性が影響した可能性があるが、そのメカニズムの解明にはさらなる検討が必要である。

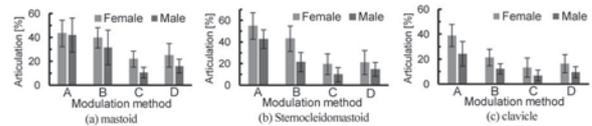


Fig.1 Scores of the monosyllable-articulation tests at each stimulus placement. A: DSB-TC, B: Transposed, C: DSB-SC, D: SSB.

### 3-4P-8

#### 3-4P-8 クラウドソーシング聴取実験のための 効果的な事前参加者スクリーニング

Effective pre-screening technique for crowdsourced listening experiments

☆宮崎美紀, 山本絢子, 土庵晋太郎, 入野俊夫(和歌山大)

- ◆コロナ禍のため、クラウドソーシングを用いた聴取実験を実施している。昨年度行った男声の模擬難聴音声を用いた了解度実験 (男声実験) の知見から、トーンピップ・Huggins Pitch (HP)・語彙数の3項目で事前スクリーニングを設計し、300名分のデータを得た。
- ◆トーンピップ平均聴取数が13個より多い人、Huggins Pitch の正答数が5問以下の人を除外し、登録された機材情報の遮音性・信頼性などをもとに116名の参加者を選定した。そのうち95名に女声の音声了解度実験を直接依頼して、29名分のデータが収集できた。
- ◆結果
  - 男声実験結果と比較したところ、事前スクリーニングの効果は認められた。しかし、300名で29名分のデータしか得られず、費用対効果は改善しなかった。音声了解度実験以外では使えるが対策が必要である。

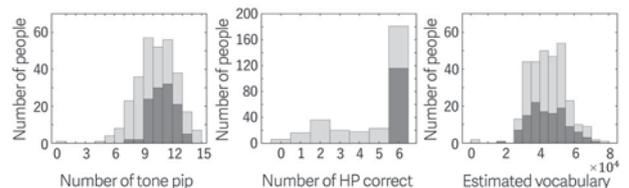


Fig.1: Distribution of participant responses for each item. light gray: all participants (300 people) dark gray: selected participants for experiment of female voice (116 people)

### 3-4P-9

#### 3-4P-9 調波複合音の基本周波数の変調に対する発声の不随意応答: 階段状ピッチシフト実験との比較

Involuntary vocalizations corresponding to modulated fundamental frequency: compare with step pitch-shift stimulus

☆摩嘉慧(豊橋技科大), 河原英紀(和歌山大), 松井淑恵(豊橋技科大)

- ◆ランダムな基本周波数変調のある刺激音を作成し、それを聴取しながら発声した際の発声の不随意応答を調査した。
- ◆刺激音の構成による影響を調べるため、純音 SINE と高調波を重ねた複合音 SINES (1~8 倍), fullSINES (1~20 倍), MFND (2~20 倍), MFNDH (8~20 倍)を用いた。
- ◆SINE 以外では刺激音のパルスの際に、それと反対方向に発声ピッチの補償応答が見られた。fullSINES と MFND での応答が最も大きい。
- ◆同じ種類の刺激で階段状ピッチシフト実験を行ったが、応答が見られなかった。

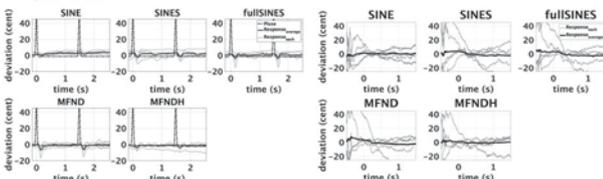


Fig 1. Pulse recovery analysis for each modulated stimulus. Dotted line: stimulus. Black solid line: subjects averages. Gray solid line: individual averages.

Fig 2. Pitch analysis for each step pitch-shift stimulus. Black solid line: subjects averages. Gray solid line: each subject.

### 3-4P-11

#### 3-4P-11 Audio Spectrogram Transformer を用いた Non-intrusive 音声了解度予測の検討

A Study of Non-intrusive Speech Intelligibility Prediction Using the Audio Spectrogram Transformer

☆中澤和司、近藤和弘(山形大院)

- ◆先行研究の結果から self-attention 機構が音声了解度推定に有効であると推測
- ◆Transformer ベースの環境音識別モデル Audio Spectrogram Transformer (AST) を用いて Non-intrusive 了解度推定を検討
- ◆画像識別タスクや環境音識別タスクで事前学習済みのモデルと、初期モデルを了解度推定タスクで学習し精度比較
- ◆事前学習済みモデルを使用することでより高い精度で了解度推定可能

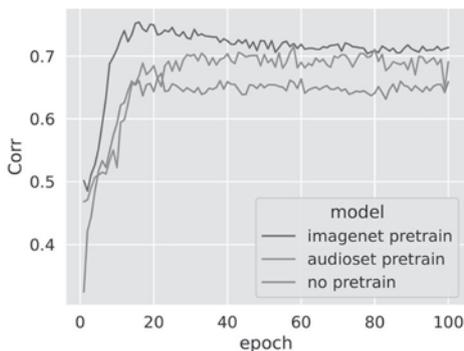


Fig.1:Linear correlation coefficient between estimated intelligibility and subjective intelligibility for each epoch

### 3-4P-10

#### 3-4P-10 ヘッドフォンを装着して発声する際に生ずる側音と提示音声信号の漏洩について

Sidetone and test signal leakage measurement in testing voice response to auditory stimulation with headphones

○河原英紀(和歌山大), 矢田部浩平(東京農工大), 松井淑恵(豊橋技科大), 程島奈緒(東海大), 水町光徳(九工大), 柳原健一(北海道医療大)

- ◆ここでは、発声中にヘッドフォンを装着した場合、どのような刺激が聴覚的に提示されているのか、また、聴覚刺激として提示した信号が漏洩することで、音声を収録するマイクロフォンにどの程度収録されるのかを調べた。
- ◆併せて、発声中に聴こえる自分の声(側音)が、ヘッドフォンの装着によりどのように変化するかを調べた。
- ◆測定の結果は、外耳道を塞がない『空気伝導イヤホン』を用いることで、自然側音に影響を与えずに、聴覚刺激を追加することによる発声の変化を調べることができる可能性を示した。

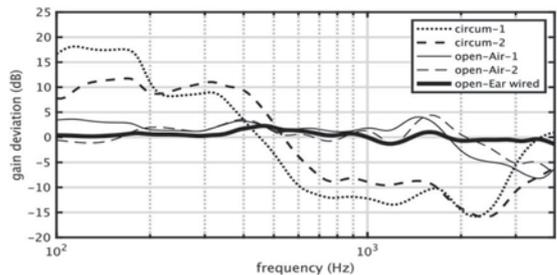


Fig.1:Sidetone deviations from the natural sidetone level due to wearing headphones. Open-Ear type one effectively introduces no deviation.

### 3-4P-12

#### 3-4P-12 音の再生手法が音像定位の反応時間に及ぼす影響

Effect of sound presentation method on reaction time of sound image localization

○増田光新, 木村敏幸(東北学院大学)

- ◆本報告では音の再生手法(ステレオ方式、Multiple Vertical Panning (MVP) 方式、直接再生)が音像定位の反応時間(音が再生されてから回答するまでの時間)に及ぼす影響をもぐらたきゲームを用いて検証した。
- ◆結果、画面中央の場合、画面端と比較して反応時間が短くなった。また、直接再生や MVP 方式の方がステレオ方式と比較して反応時間が短いことが分かった。

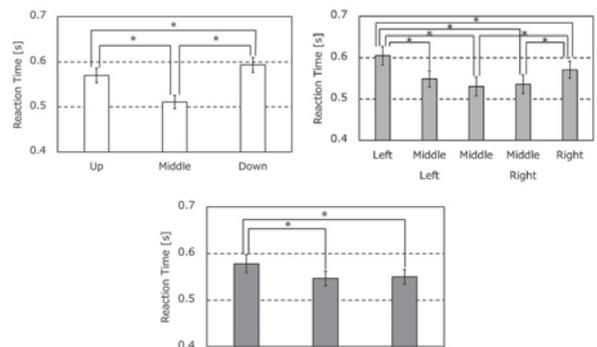


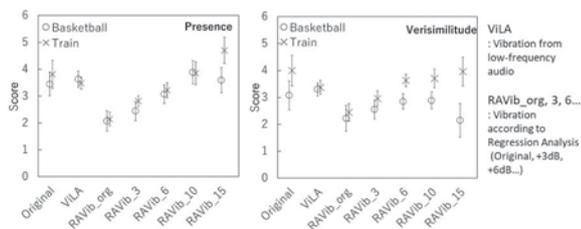
Fig.1:Results of the reaction time (Upper left: Vertical direction, Upper right: Horizontal direction, Under: Sound presentation method)

### 3-4P-13

3-4P-13 映像音響コンテンツの特徴量から合成した  
全身振動情報が臨場感と迫真性に与える影響  
Effects of whole-body vibration information synthesized  
from audio-visual features on the sense of presence and verisimilitude

☆阿部翔太†, 山高正烈††, 坂本修一†, 鈴木陽一†  
†: 東北大通研 院情科研, ††: 愛知工科大, ††: 東北文化学園大

- ◆音響特徴量を用いた重回帰モデルによって、視聴覚コンテンツの体験に最適な振動レベルを推定する方法を先行研究で提案した。
  - 本報告では、この推定法に従い合成された全身振動情報 (RAVib\_org) が臨場感、迫真性に与える影響を検討した。
- ◆実験の結果、RAVib\_org では、臨場感、迫真性を十分に生起させることはできなかった。
  - 先行研究において、最適振動レベルの推定に誤差が生じる場合があり、本実験で用いたコンテンツにおける推測誤差を考慮したところ (Basketball : RAVib\_6 相当, Train : RAVib\_10 相当)、臨場感、迫真性ともにより高い値となった。
- ◆このことは、最適振動レベルの正確な推測が臨場感、迫真性により強く寄与する可能性を示しており、音響特徴量に加えて、その他の情報、例えば対象物体や事象の情報も考慮した推定法の構築の必要性を示唆している。



Subjective evaluation of each vibration condition

### 3-4P-14

3-4P-14 マルチチャンネル音響システムにおける  
異なるパンニング手法の印象評価  
—ベクトル・ベース振幅パンニングと  
アンビソニック・パンニングを比較して—

Impression Evaluation of Different Panning Techniques in Multichannel Audio System: Comparison of Vector-Based and Ambisonic Amplitude Panning

☆森永実季, 亀川徹, 丸井淳史(東京芸大)

本研究では、アンビソニックパンニングを評価するための評価語抽出と2つの主観評価実験を行った。まず、レパートリーグリッド法 (RGT) を用いた評価語抽出を行った。その結果 123 個の評価語対が抽出され、最終的に 13 個の評価語を採用した。

そして 5 つのパンニング手法 (実音源・VBAP・アンビソニックス 1 次・3~5 次) に、3 つの音素材 (スピーチ・ギター・トランペット) と 2 つのパンニング角度 (0 度・-30 度) を組み合わせた刺激を用いて、類似度についての実験と評定尺度法による実験を行った。類似度については、どの音素材・パンニング角度においても高次アンビソニックスの類似度が高く、実音源、VBAP、1 次アンビソニックスはそれぞれ印象が異なることが分かった。評定尺度法の実験では、RGT による 13 個の評価語と「好み」について評価を行った。その結果、アンビソニックパンニングと相関が高いと考えられる評価語として「包まれ感がある」「広がりがある」「輪郭がにじむ」「奥行き感がある」「鈍い」「厚みのある」「豊か」「上方向に定位」が挙げられた。一方、実音源や VBAP と相関が高いと考えられる評価語として「定位が明瞭」「音像幅が狭い」「クリアな」「芯がある」が挙げられた。そして特に 1 次アンビソニックスと「包まれ感がある」の相関が高いことが伺えた。

また各刺激に対する、両耳間相関度および前後・左右・上下方向のエネルギー率の影響について検討した。