

3-1-1

3-1-1 Physics-informed neural networks を用いた 光学的音響計測のための三次元音場再構成 Three-dimensional sound field reconstruction from optical projections using physics-informed neural networks

◎伊藤陸人(早大理工), 石川憲治(NTT),
谷川理佐子(NTT/早大理工), 及川靖広(早大理工)

- 目的:
音圧線積分値から空間上の各点の音圧を求めるための音場再構成
→ Physics-informed neural networks (PINNs) の活用
- 提案法:
■ Sinusoidal Representation Neural Network (SIREN) を使用
✦ 音圧線積分値とニューラルフィールドの線積分値の誤差
✦ モデル出力がヘルムホルツ方程式を満たすかに関する誤差
一両者を小さくするように学習
■ 外挿能力, ノイズへの頑健性等の PINNs の特長を保持

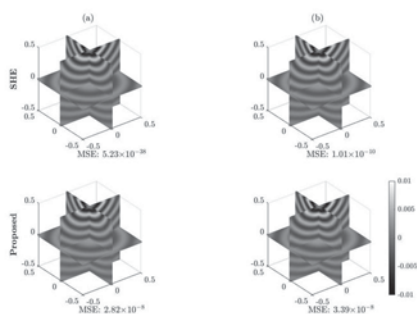


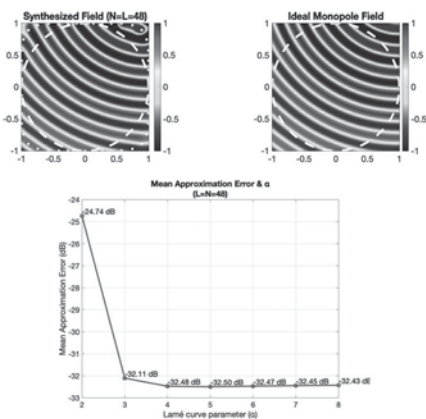
Fig1: Reconstructed fields using Spherical Harmonic Expansion (SHE) and the proposed method for (a) noise-free and (b) noisy cases.

3-1-3

3-1-3 平面波分解に基づいた音場合成における ラメ関数を用いたスピーカアレイ配置 Loudspeaker array configuration using Lamé function for sound field synthesis based on plane wave decomposition

☆Gao Ming, 坂口智弘, 及川靖広(早大理工)

- ◆ 実環境での音場制御を考慮する場合、従来の円形配置に限定されない理論の構築が求められている。
- ◆ ラメ関数は次数を上げることで、円形から矩形に近づけることができる代数曲線である。
- ◆ 本研究ではラメ関数を導入し、円形アレイ配置による点音源音場合成時のスピーカ位置の表示を置き換える。
- ◆ シミュレーションを通じて誤差を計算し、提案手法が良好な合成精度が得られることが確認できた。



3-1-2

3-1-2 接線法に基づくスピーカアレイ: 電子ピアノを想定した検討 Tangent line method-based loudspeaker array: Application to digital piano

☆真下人和, △倉内喬乃介(工学院大), △水品隆広, △濱中秀雄,
△中山健二(カシオ計算機), 貝塚勉(工学院大)

- ◆ 複数のスピーカを一行に並べたアレイを使い、円弧の形状の音響ビームを再生する。円弧の長さを規定できるため、音の届く方向のみならず、距離も制御できる。
- ◆ 電子ピアノへの適用を想定し、スピーカアレイを設計・試作した。ステレオ再生を想定し、奏者の両耳へ2本のビームを別々に届ける。
- ◆ 提案のスピーカアレイによって作り出される音圧の空間分布を評価し、周囲への音漏れを抑えつつ、奏者の耳元へ音を届けられることを確認した。

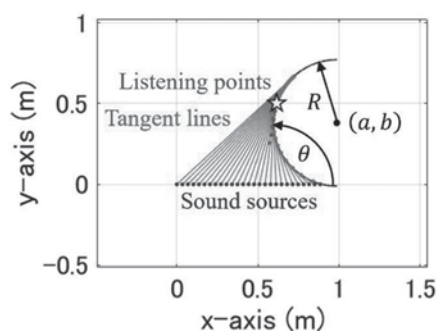


Fig. 1 Geometry of sound sources, listening points, and circular beam as envelope of tangent lines.

3-1-4

3-1-4 ラメ関数による定式化を用いた 矩形形状スピーカアレイの音場生成 Study on sound field creation using radius formulation with Lamé function for rectangular array.

◎坂口智弘, 及川靖広(早大理工)

- ◆ 実在する室内形状における音場制御を考慮し、矩形形状での音場生成およびその評価をシミュレーションにて実施した。
- ◆ ラメ関数の次数ごとのアレイと波面を示す (Fig.1)。中央部では理想波面に近い波面が生成されている。一方で領域拡張による波面生成領域は拡張しないことがわかった。
- ◆ 理想波面との差を可視化するためにアレイに内接する平面で誤差分布を示した (Fig.2)。中心部分では誤差が少ないことがわかる。一方でそのエリアを外れると急激に誤差増加することがわかった。

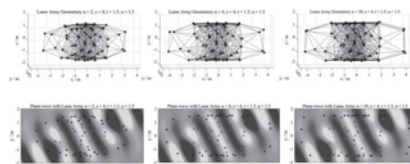


Fig.1 The geometries and wave forms with rectangle shape

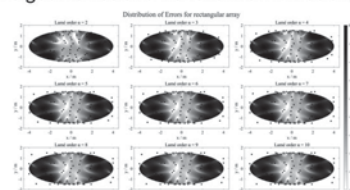


Fig.2 The error distribution map with rectangle shape

3-1-5

3-1-5 優決定系 BoSC システムにおける 正則化パラメータの最適化による 音場再現性能の改善

Improvement of sound field reproduction by optimizing regularization parameters in overdetermined BoSC systems

○伊勢史郎 (東京電機大)

- ◆音場再現システムを小型化するためにスピーカ数を減らした場合について逆システム設計法を確立する必要がある。
- ◆スピーカ数がマイクロホン数よりも少ない場合には MIMO 逆システムが優決定系となるが、最適な正則化パラメータを求めることにより安定した逆システムが得られ、音場再現精度が 20dB 改善した。

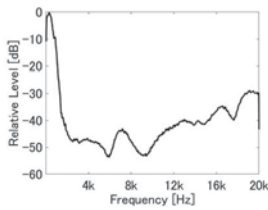


図1 LMSによる逆システムの周波数特性

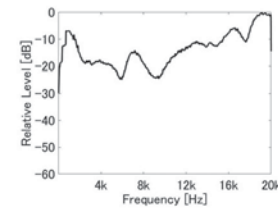


図3 最適化LMSによる逆システムの周波数特性

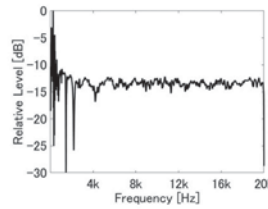


図2 LMSによる再現されたパルスの周波数特性

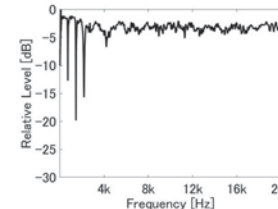


図4 最適化LMSによる再現されたパルスの周波数特性

3-2-2

3-2-2 TV-CAR 分析の ARMA 分析への拡張に関する検討

A study of the extension of ARMA of the TV-CAR analysis

○舟木慶一(琉球大学)

解析信号を用いた音声分析は、低周波数領域においてより正確なスペクトル推定を実現する。我々は時変複素 AR(TV-CAR) 音声分析を構築している。これらのアルゴリズムは、最小二乗誤差 (MMSE) の他、一般化最小二乗法、 l_1 ノルム正則化、 l_2 ノルム正則化、TLS 基準方式などである。音声信号は反共振を持つため、特に鼻音において、ARMA パラメータ推定が音声処理と音声研究のツールとして有用である可能性がある。ARMA 分析として 1980 年代に拡張カルマンフィルタに基づく適応的方式が提案されているが、適応的アルゴリズムを TV-CAR モデルに適用するのは困難なので、本稿では、励振信号を残差信号から推定されるパルス列と仮定する IV(補助変数) 法に基づいて ARMA スペクトル推定を行う手法を提案する。提案手法がロバストな ARMA スペクトル推定を実現できることを示す。

3-2-1

3-2-1 MOS-Bench: 音声品質評価モデルの汎化能力に着目したベンチマーク

MOS-Bench: Benchmarking Generalization Abilities of Subjective Speech Quality Assessment Models.

OHuang Wen-Chin¹, COOPER Erica², 戸田智基¹ (¹名古屋大学, ²NICT)

本稿では、MOS-Bench と呼ばれる学習および評価用データセットを紹介する。さらに、MOS-Bench を使いこなすため、SHEET(Speech Human Evaluation Estimation Toolkit)と呼ばれるオープンソースツールキットを開発し、公開した。



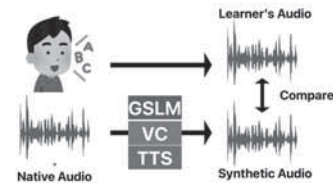
<https://github.com/unilight/sheet>

3-2-3

3-2-3 Modeling and Predicting Individual Learners' Performance of Prosody Imitation and its Use for Material Selection

学習者による韻律模倣精度の予測モデルとそれを用いた音源選択

☆Xiai Cheng, Haopeng Geng, Kentaro Onda, Daisuke Saito, Nobuaki Minematsu (Univ. Tokyo)



◆ Model Analysis and Difficulty Prediction.

Table 1. The input data types for model training.

	Native Speech	Learner's Speech
GSLM	Audio	X
VC	Audio	Audio
TTS	Text	Audio

- 1) Training GSLM, VC model, TTS model on different data types.
- 2) Verification of similarity between Synthetic Audio and Learner's Audio.
- 3) Verification of models feasibility in predicting difficulty.

◆ Application Scenarios for the Three Models Based on Previous Results.

- GSLM can predict for all learners without learners' data.
- TTS/VC can predict for individual learners.

3-2-4

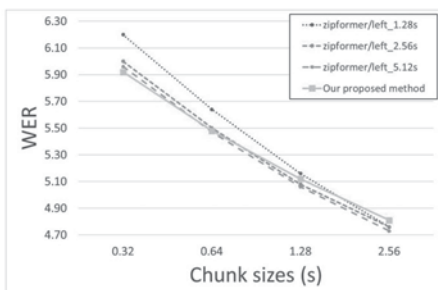
3-2-4 Combining Linear RNNs, Self-Attention, and Convolution for Lightweight Streaming Speech Encoding

☆ Wen Shen TEO, Yasuhiro MINAMI (UEC)

- ◆ Incorporates MAMBA-2 into Zipformer, a Conformer-based streaming acoustic encoder.
- ◆ Eliminates left-context key-value caches of Self-Attention modules in Zipformer.
- ◆ Carries left-context information via smaller, fixed-size hidden state of MAMBA-2

Results

- ◆ Our proposed method performs on par with conventional Zipformer at smaller chunk sizes, despite smaller model and cache sizes.



3-2-6

3-2-6 大規模事前学習済みモデルと Mixture-of-experts を用いた多言語音声認識 Multilingual Speech Recognition Using Mixture-of-experts and Large-scale Pretrained Models

◎品山健太郎, 松浦孝平, 三村正人(日本電信電話株式会社)

- ◆ 大規模事前学習済みモデル (XLS-R) に基づく多言語音声認識において、図1のように Mixture-of-Adapters を導入。
- ◆ アダプタ選択の偏りを軽減するため、balancing loss とアダプタのパラメータドロップアウトの2つのテクニックを導入。
- ◆ フレーム単位の gating と発話単位の gating の組み合わせを提案。
- ◆ 表1の通り、mixture-of-adapters が単一アダプタの性能を上回り、balancing loss やドロップアウトの導入により性能がさらに改善。

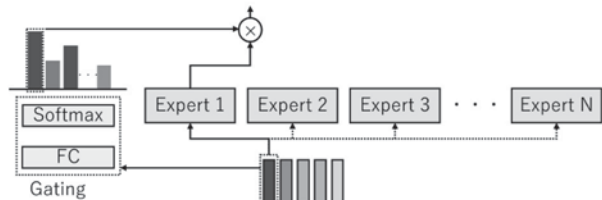


Fig. 1: Structure of Mixture-of-Adapters

	学習データ	Gating	balancing loss	dropout	# trainable params	WER(%)					平均
						en	de	fr	it	es	
Full finetune	mono	-	-	-	311.33M	23.5	19.3	25.8	9.4	12.4	18.1
	multi	-	-	-	311.33M	22.1	18.5	26.4	9.8	11.5	17.6
Single adapter	mono	-	-	-	1.78M	25.9	20.1	29.4	10.9	12.9	19.8
	multi	-	-	-	1.78M	26.3	22.7	30.5	12.5	13.2	21.0
Mixture of adapters	multi	frame	✓	✓	6.89M	26.2	22.7	30.3	12.1	13.2	20.9
	multi	frame	✓	✓	6.89M	25.8	21.0	29.9	12.1	13.2	20.4
	multi	発話(全層)	✓	✓	6.89M	25.8	21.9	30.5	11.7	13.5	20.6
	multi	発話(後半層)	✓	✓	6.89M	25.7	21.4	29.8	11.6	13.1	20.3
	multi	発話(後半層)	✓	✓	6.89M	25.7	21.4	29.8	11.6	13.1	20.3
	multi	frame	✓	✓	6.89M	22.2	18.9	27.6	10.5	11.9	18.2
multi	発話(後半層)	✓	✓	6.89M	23.8	20.0	28.9	11.4	12.3	19.3	

Table 1: Comparison of Learning Methods across Five Languages (Evaluation Metric: Word Error Rate, Dataset: FLEURS)

3-2-5

3-2-5 大規模自己教師あり学習モデルによる日本語方言の音声認識

Automatic Speech Recognition of Japanese Dialects Using Large-scale Self-supervised Learning Models

○瀧澤大吾, 中村友彦, 須田仁志, 深山覚(産総研)

- ◆ 本研究では、方言を含む大規模日本語音声データで事前学習した日本語自己教師あり学習モデルが、日本語方言音声認識に有効であることを検証した
- ◆ 方言音声のデータ量は限定的であるため、日本語方言音声認識においては自己教師あり学習モデルを用いる手法が有用であり、特に多言語で学習したもの比べて、日本語音声で事前学習した自己教師あり学習モデルを用いた場合に低い文字誤り率を実現できた
- ◆ 日本各地の方言で性能良く音声認識ができることは、音声認識技術の公平性の観点で重要であり、今後本研究で構築した自己教師あり学習モデルを改良しつつ、日本語方言の音声認識性能向上に取り組む

Table 1 音声認識モデルにおける COJADS の評価結果

音声認識モデル	SSL モデル	事前学習言語	継続学習	CER (%)
Conformer	-	-	-	41.2
E-Branchformer	-	-	-	42.9
SSL モデル+Conformer	HuBERT Large	英語	なし	45.1
	XLSR-53	53 言語	なし	41.0
	HuBERT Large (rinna)	日本語	なし	33.3
	HuBERT Large (AIST)	日本語	なし	33.3
			あり	<u>32.7</u>

3-2-7

3-2-7 マルチトーカー重畳音声認識及び話者ダイアライゼーションの統合自己回帰モデリング

Joint autoregressive modeling of multi-talker overlapped speech recognition and speaker diarization

◎牧島直輝, 河田尚孝, 庵愛, 田中智大, 折橋翔太, 安藤厚志, 増村亮(NTT)

- ◆ 本稿では、複数話者の重畳音声を入力として、誰がいつ何を話したかを推定する話者ダイアライゼーション及びマルチトーカー音声認識の統合モデリングを扱う。
- ◆ 従来の話者ダイアライゼーション及びマルチトーカー音声認識の統合モデリングでは、まず非重畳音声区間を推定し、その区間の話者特徴量をクラスタリングすることで、話者を推定する。
- ◆ しかし、重畳音声では非重畳区間が非常に短いケースが存在しており、そのような場合において従来法は精度が大きく劣化する。
- ◆ 本稿では、この問題を改善するため、マルチトーカー音声認識と同時に、重畳音声から直接話者特徴量を推定可能な枠組みを提案する。

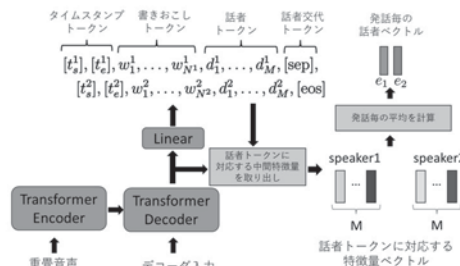


Fig.1 Overview of the proposed method.

3-2-8

3-2-8 NAIST System for the CHiME-8 Distant Meeting Transcription Challenge

☆Yuta Hirano (NAIST), △Mau Nguyen (JAIST), Kakeru Azuma, △Jan Meyer Saragih, Sakriani Sakti (NAIST)

- ◆NOTSOFAR-1 task of the CHiME-8 challenge focuses on advancing meeting transcription systems that work with a single distant microphone.
- ◆Aiming to develop an efficient system, we worked to improve speech recognition accuracy while reducing inference speed by modifying the ASR module.
- ◆Proposed system significantly reduced the inference time by **up to 83.4%**.
- ◆Proposed system earned **3rd place** in the multi-channel track.
- ◆Proposed system earned **4th place** in the single-channel track.

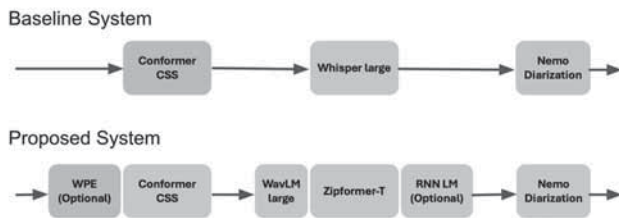


Fig.1: Overall architecture of NAIST system for NOTSOFAR-1 task

3-2-10

asj3-2-10 自然言語による音響区間検索

Language-based audio moment retrieval

◎宗像北斗, △西村太一, △仲田勝太, 小松達也 (LINE ヤフー)

- ◆本研究では、音響区間検索というタスクを新たに提案する。
- ◆音響区間検索におけるデータセットは存在しないため、新たなデータセットを構築する。各データは長い音響データと音響データ内のある場面の説明文およびタイムスタンプからなる。本研究ではモデルの学習と評価のため、シミュレーションと人手でのアノテーションの二種類の方法でデータセットを構築する。
- ◆さらに本タスクに向けた新たなモデルを提案する。単純なアプローチに基づくモデルと比較し、このモデルは音響データのフレーム間の依存関係を捉え、クエリに対応する区間をより正確に予測することが可能である。

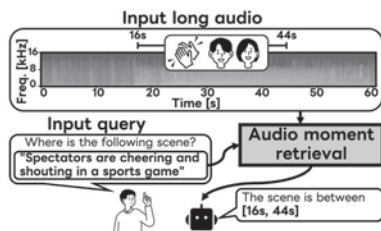


Fig. 1 An overview of audio moment retrieval.

3-2-9

Efficient Environmental Sound Classification based on Real-time Block Pruning

○王昌竜, 伊藤彰則, 能勢隆(東北大・工)

- ◆事前学習済みモデルは、環境音認識を含む下流タスクで優れた性能を発揮しますが、再適応においてモデルが過大化することがよくあります。環境音認識に対する効率的な研究が不足している中、Real-time Block Pruning に基づく効率的な環境音分類のための適応型モデル圧縮手法を提案しました。
- ◆本手法は、汎用の分類モデルを深さ方向に分解し、各層での分類結果を活用して不要な層を段階的に剪定する仕組みを構築しました。また、グローバル最適化フレームワークを用いることで、許容可能な性能劣化の範囲内で最短の構造を選択します。

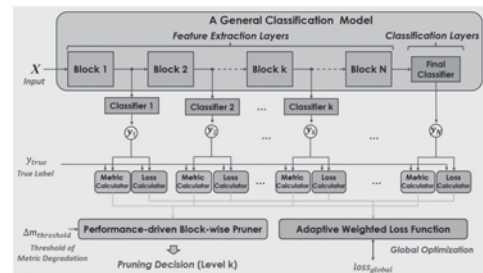


Fig.1: The framework of depth-wise classification and global optimization for real-time block pruning

- ◆本手法は、異なるタスクやモデル構造にも適用可能であり、追加の経験的調整を必要とせず、一度の訓練プロセスで効率的に実行できます。

3-2-11

3-2-11 反復した環境音の学習による擬音語認識システムの開発

Development of onomatopoeic recognition system based on training repeated environmental sounds

☆松浪 優斗, 西村 竜一 (和歌山大)

- ◆End-to-End 型深層学習による擬音語認識システムを開発
 - モデルの学習に環境音を連結した連続環境音の音響信号を使用
 - ◇ 擬音語 (テキスト) の情報が付与されているデータセット RWCP-SSD-Onomatopoeia の環境音サンプルを使用
 - 繰り返しを考慮し、環境音を連結する際に次の2条件を検討
 - ◇ 無作為型: 環境音 10 個を無作為に連結
 - ◇ 反復型: 2 つの環境音をセットにして 5 回反復して連結
 - 2 条件のモデルを比較する ESC-50 サンプルの認識実験を実施
 - ◇ 無作為 100%: 無作為型のみで学習したモデル
 - ◇ 反復 50%: 訓練データ 50% を反復型にして学習したモデル
- ◆反復型を含めたことで「時計のアラーム音」「時計の針の音」等の反復性を持つ環境音に対する擬音語の出力精度の向上を確認 (Table 1)

Table 1: 妥当な出力の割合 (主観評価)

	無作為100%	反復50%
ドアのノック音	50.0%	22.5%
マウスのクリック音	15.0%	5.0%
キーボードの音	45.0%	32.5%
ドアが軋む音	7.5%	5.0%
缶を開ける音	17.5%	0.0%
洗濯機の音	27.5%	32.5%
掃除機の音	67.5%	27.5%
時計のアラーム音	30.0%	32.5%
時計の針の音	12.5%	37.5%
ガラスが割れる音	27.5%	17.5%

【擬音語認識の出力例】
「時計の針の音」
無作為 100%モデル: チャッガッパ
ッガッガッガッガッガッガッ
反復 50%モデル: ガッコンガッ
コンガッコンガッコンガッコン

3-4-1

3-4-1

聴覚心理学の授業で何を習った？

What did you learn in the auditory psychology class?

○竹内京子(順天堂大), 青木直史(北大), 荒井隆行(上智大), 大金さや香(国際医療福祉大), △鈴木恵子, 秦若菜, △村上健(北里大), 世木秀明(千葉工大), 安啓一(筑波技大)

- ◆言語聴覚士養成校では、音響学、聴覚心理学が必修科目。音響学は、学生が最も苦手の科目である。
- ◆聴覚心理学は、音響学と同じ科目内で行われることも多いが、専門家も少なく、教えることのできる教師も少ない。音響学教師でさえ、学ぶことが難しい状況である。教科書や参考書もほぼない。
- ◆2021年より開催の現役言語聴覚士対象の講習会「STのための音響学」の受講者アンケートでは、聴覚心理学は、養成校の授業で「記憶がない科目」という結果も見られた。
- ◆教える内容が、教師にも分からない難易度ということは、学生の記憶に残らなかったのではなく、実際に授業で扱っていない可能性もある。本発表では、現役言語聴覚士に、養成校時代の聴覚心理学の授業の記憶と、現在の理解度の自己評価をしてもらった結果を示す。
- ◆言語聴覚士養成校の聴覚心理学の現状調査は、授業の「記憶がない」という感想を持つ理由が少し理解できる結果となった。今後は、より詳細を検証するとともに、学生だけでなく、聴覚心理学を学びたい教師の援助方法も考えたい。

3-4-3

3-4-3 Intelligibility of Checkerboard and Interrupted Speech: The effects of level differences between two talkers

☆Jun HASEGAWA, △Rina MATSUURA, Kazuo UEDA (Kyushu Univ.), Hiroshige TAKEICHI (RIKEN), Gerard B. REMIJN, and Emi HASUO (Kyushu Univ.)

The effects of signal-to-noise ratios (SNRs) between a target and a distractor on the target's intelligibility of two-talker interrupted and checkerboard speech stimuli (Figure) were investigated with native Japanese listeners ($N = 22$; age range: 19–25). SNRs (-6, 0, 6 dB) did not alter the target's intelligibility for interrupted, 2-, and 4-band checkerboard speech stimuli. However, reducing SNRs degraded intelligibility for 8- and 16-band checkerboard stimuli. The results suggest the number of frequency bands categorically affects perceptual segregation between the target and distractor.

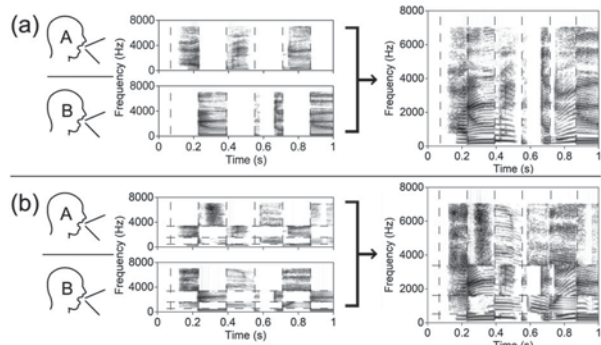


Fig. A schematic diagram depicting the process of stimulus generation. Two-talker (a) interrupted and (b) 4-band checkerboard speech stimulus.

3-4-2

3-4-2 聴覚フィードバック情報に基づく母音の発話運動制御における話者の母語の影響

Influence of speaker's native language on speech motor control of vowels based on auditory feedback information

☆村上正悟, 上江洲安史, 鶴木祐史 (JAIST)

疑問：母音の聴覚-発話運動制御が話者の母語に依存するか不明

目的：母音/e/のF1における聴覚-発話運動制御の学習に、話者の母語母音/e/と英語母音/æ/の位置が影響するか検討する

方法：日本語、中国語、ベトナム語母語話者各10名を対象にF1 摂動(上昇・下降)フォルマント変形聴覚フィードバック実験を実施

結果：

母語話者	母音位置の関係	F1変化パターン	平均F1補償量
日本語	母語 = 英語	F1 上昇 < F1 下降	母語 = 英語
中国語	母語 ≠ 英語	F1 上昇 > F1 下降	母語 = 英語
ベトナム語	母語 = 英語	母語 > 英語	母語 = 英語

➤ 3言語の母語話者間において、平均F1補償量が異なった

まとめ：母音/e/のF1における聴覚-発話運動制御の学習には、話者の母語母音と英語母音の位置関係が影響することがわかった

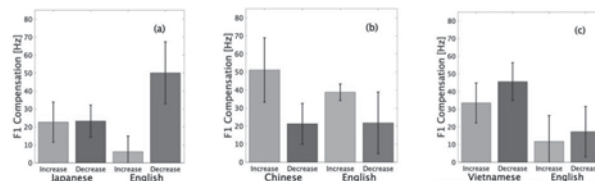


Fig.1: Averaged F1-compensation (a) JPN, (b) CHN, and (c) VNM

3-4-4

3-4-4 単耳マスク量の低下しない条件における両耳間音圧差による音声の了解度の変化

Change in speech intelligibility due to interaural level difference added without decreasing masking in unilateral ear

☆花木涼太郎, 森川大輔, モクタリ パーハム, 岡崎聡(富山県立大・工)

- ◆片耳のマスク量とは独立に、両耳間音圧差 (ILD) が目的音 (音声) の聴き取りやすさに及ぼす影響は不明である。
- ◆そこで、異なるILDを、目的音の単耳マスク量が低下しない条件で目的音と妨害音 (ピンクノイズ) に付与し、目的音の音声了解度を測定した。
- ◆目的音と妨害音に付与したILDの差 (ILD_{diff}) が15 dBの際の了解度が、ILD_{diff}が0から10 dBの際の了解度に対して有意に増加した。
- ◆単耳のマスク量が低下しない条件でも、目的音と妨害音のILDの違いが了解度に影響を及ぼすことがわかった。

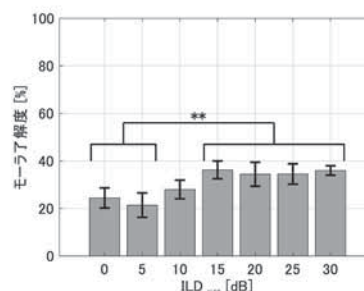


Fig.1 Relation between mean mora intelligibility and ILD_{diff}

(**: $p < 0.01$)

3-4-5

3-4-5 宣伝音声の話者と聴取者の属性が購買意欲の媒介モデルに及ぼす影響

Effects of speaker and listener attributes in advertising speech on mediation model of purchase intention

◎長野瑞生, 井島勇祐, 廣谷定男(NTT)

◆宣伝音声は購買意欲に与える影響

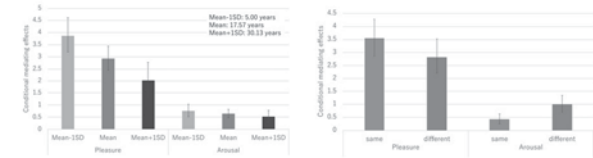
- 感情や音声印象は音声の購買意欲への影響を媒介[Nagano+, 2025]
- 販売員と顧客の性別等の属性が類似することで購買意欲が向上[Smith+, 1998]
- 話者と聴取者の年齢差や性別の一致/不一致が、感情や音声印象の媒介効果を調整し、購買意欲に影響を及ぼす可能性がある

◆合成音声を用いた大規模な宣伝音声聴取実験による検証

- 話者クラスタリングにより選出した20-70代の男女話者26名に対し、1話者あたり13文章を音声合成で作成
- 話者と聴取者の年齢差や性別の一致/不一致が感情・音声印象の媒介効果に及ぼす影響を大規模な主観評価実験により調査

◆年齢差、性別の一致/不一致の影響を調整媒介分析によって検証

- 快感情、明るさ、温かさは年齢差が小さく性別一致のとき媒介効果が高い
- 声の落ち着いたさや迫力の有無は年齢差が大きい場合に媒介効果が高い
- 覚醒、声のかすれ、迫力の有無は性別が不一致の場合に媒介効果が高い



3-4-6

3-4-6 時間振幅包絡線の瞬時変調成分を制御した雑音駆動音声の緊迫感知覚の変化

Study on urgency perception of noise-vocoded speech with controlled instantaneous modulation components of temporal amplitude envelope

☆房野早希, 郭太陽, 磯山拓都, 木谷俊介, 鶴木祐史(北陸先端大)

◆音声の時間振幅包絡線(TAE)の瞬時変調成分(IMCs)を用いて雑音駆動音声(NVS)の緊迫感知覚の変化を説明できるかどうかを検討する。

◆日本語母語話者14名に対しNVSを用いた聴取実験を行った結果、次のことがわかった。

- (1) TAEは音声の緊迫感知覚の手がかりを含む
- (2) TAEに対する変調フィルタバンクの利用は、NVSの緊迫感知覚に影響を与えない
- (3) 緊迫感知覚に重要な変調周波数帯域は4 Hz から 16 Hz
- (4) TAEの時間反転処理によってNVSの緊迫感^は低減する
- (5) TAEの時間伸長圧縮処理を用いて時間反転したTAEのIMCsを模擬するようにIMCsを制御することで、時間反転したTAEを持つNVSと同様の緊迫感の低減が見られる (Fig.1)

◆音声の緊迫感知覚の変化を説明するにあたって、TAEのIMCsが直接的な手がかりとして重要な働きを考えると考えられる。

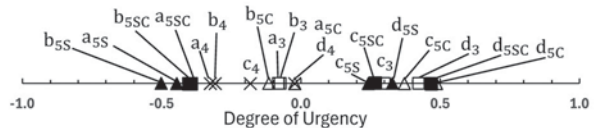


Fig.1:Urgency perception of NVS.

(NVS (□: a₃~d₃), NVS with Reversed TAE (×: a₄~d₄), NVS with Compressed TAE (△: a_{5c}~d_{5c}), NVS with Stretched TAE (▲: a_{5s}~d_{5s}) and NVS with Stretched and Compressed TAE (■: a_{5sc}~d_{5sc}))

3-4-7

3-4-7 「歌う雑音」: 多層動的バンドノイズ音声から見える世界

“Singing Noise”: Multi-layered dynamic band-noise speech synthesized by the Phantom Skeleton method.

○内田照久(大学入試センター)・森勢将雅(明大・総合数理)・

津崎 実(電通大院・情報理工)

目的 ピッチ知覚研究で利用できる声の生成を試みる。調波構造状スペクトルを備えた多層動的バンドノイズ音声を生成する。

方法 疑似ささやき声の生成法 Phantom Silhouette 方式を拡張。通常の歌声から抽出したスペクトルを元に白色雑音で駆動、さらに低域を抑制する Phantom Skeleton 方式を提案。

結果 生成した実験音声を首都圏の国公立大学(14大学)の1年生256名(男性166名,女性84名,無回答6名)が評価。声種はささやき声と知覚され、品質は疑似ささやき声と同等かつ、メロディを感じる歌声として聞こえると評価された。

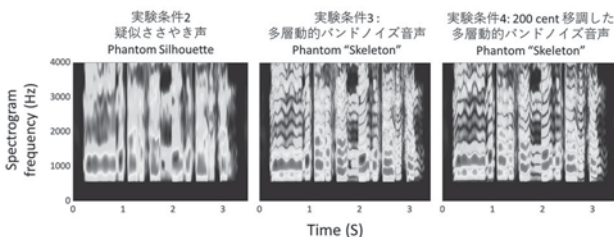


Fig. 1 Modified spectrum used to generate experimental voice.

3-4-8

3-4-8 音声のピッチパターン弁別に老人性難聴が与える影響: 高齢者・模擬難聴者・若年健聴者による調査

Effects of age-related hearing loss on speech pitch pattern discrimination: with elderly, simulated hearing impaired and young normal hearing people

☆森澁斗(豊橋技科大院), △原楓子(豊橋技科大), 河原英紀(和歌山大), 松井淑恵(豊橋技科大院)

◆加齢による聴覚末梢機能の低下が音声のピッチパターンの弁別に与える影響を調べるため、高齢者・模擬難聴者・若年健聴者の音声弁別閾値の測定を行った。

◆原音声組には、男女5組の4モーラ単語を使用した。刺激音声はWORLDベースの音声モーフィングと模擬難聴システムWHISによって作成した。

◆高齢者群・模擬難聴者群のピッチパターン弁別閾値の上昇は、純音聴力だけでは説明できないことが明らかになった。

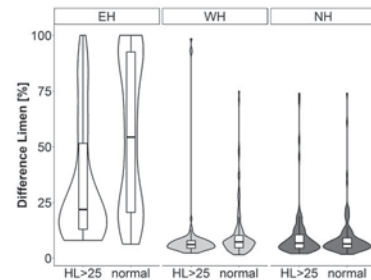


Fig. 1: Speech discrimination thresholds for each participant group (EH: elderly, WH: simulated hearing loss, NH: young healthy hearing). Each group is divided into normal hearing (normal) and mild hearing loss (HL > 25) corresponding to hearing level of EH listeners.

3-4-9

3-4-9 音声のスペクトル包絡の弁別に 老人性難聴が与える影響： 高齢者・模擬難聴者・若年健聴者による検討

Effects of age-related hearing loss on the discrimination of spectral envelope in voices: Investigation by elderly listeners, simulated hearing-impaired listeners, and young normal-hearing listeners

△原楓子(豊橋技科大), ☆森澁斗(豊橋技科大),
河原英紀(和歌山大), 松井淑恵(豊橋技科大)

- ◆加齢に伴う聴覚の末梢機能の低下が音声のスペクトル包絡の弁別に与える影響を、高齢者、模擬難聴者、若年健聴者を対象に音声弁別閾を測定することで調査した。
- ◆4モーラ単語2つを男女5組用意し、WORLDベースの音声モーフィングで刺激を作成した。その後、模擬難聴システムWHISで処理した。
- ◆高齢者群の弁別閾値が他の群より有意に大きいことが示された。高齢者群のスペクトル包絡パターンの弁別閾値上昇は、純音聴力のみでは説明できないことが明らかになった。

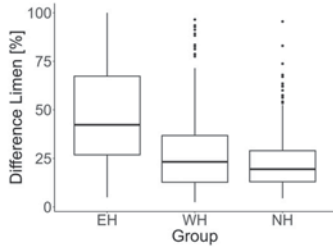


Fig. 1: A boxplot of voice discrimination thresholds for participant groups (EH: elderly, WH: simulated hearing impaired by WHIS, NH: young normal hearing).

3-4-11

3-4-11 振幅包絡情報と音声基盤モデルを用いた 難聴者のための音声了解度予測

Speech intelligibility prediction for hearing-impaired listeners using auditory envelope information and speech foundation models

○山本克彦, 郡山知樹(サイバーエージェント)

- ◆先行研究：音声基盤モデルを用いた音声了解度の予測モデル (SFM)
 - リファレンス信号を必要としない non-intrusive な評価指標
 - 補聴器音声の了解度の予測コンペ (CPC2) で一位の予測精度
 - 問題点：難聴者の特徴量はオーディオグラムのみ
 - 難聴者特性と入力音を考慮した特徴量抽出が不十分
- ◆提案手法：SFMに模擬難聴処理済みの聴覚モデル特徴量を追加
 - 前回：聴覚フィルタバンク+振幅包絡抽出+変調フィルタバンク
 - 今回：聴覚フィルタバンク+振幅包絡抽出
- ◆実験：CPC2のデータセットを用いた評価 (15 試行)
 - データセット全体で予測誤差 (RMSE) が有意に減少 ($p < .05$)
 - 振幅包絡情報がSFMの予測精度を改善できることを示した

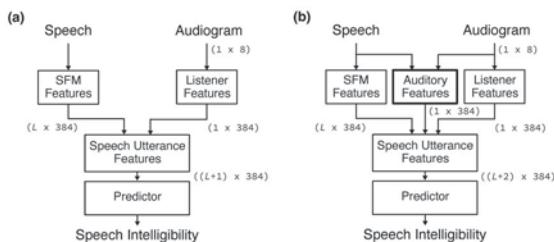


Figure 1: Overall structure of the conventional model (a) and our proposed model (b).

3-4-10

3-4-10 先頭子音の音声強調による 加齢性難聴者の単語聞き取り改善

Improvement of word comprehension in patients with age-related hearing loss by emphasis of initial consonants

☆上原千晴(中央大), 菅田康平(中央大院), 高橋伸弥(福岡大),
森山剛(東京工芸大), △小野博, 山口雅夫, 戸井武司(中央大)

- ◆現在多くの補聴器に使用されている高域強調の音声加工は、聴力が低下した周波数帯域の聞き取り向上を図っているが、十分な効果が得られていない。そこで、先頭子音に着目し、聞き取り改善を行った。
- ◆評価音声は4 moraの単語で、先頭子音に対し摩擦音は「高域強調」、破裂音と破擦音は「高域強調+振幅強調」、鼻音と弾音は「高域強調+伸張」の強調加工をし、無加工音声、高域強調音声および先頭子音の強調音声の3種をスピーカから提示して、正答率を評価した。
- ◆Fig.1は破裂子音「カンマツ」の音圧レベルの時間変化であり、Fig.2に示す高域強調+先頭子音の音声強調をすることで、正答率が向上した。特に、高域の聴力が低下する高音障害漸近型や山型の被験者には効果があることを明らかにした。

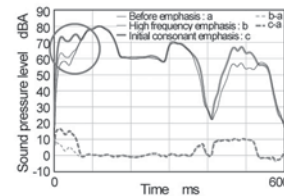


Fig.1: Time trend of sound pressure level.

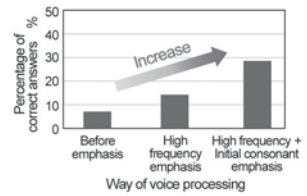


Fig.2: Number of correct answers for "Kanmatsu".

3-4-12

3-4-12 一側性難聴者が残響下での聴取に適応 していく機序に関する検討 —ガンマトーンフィルタバンクによる 刺激を用いて—

An investigation into how individuals with unilateral hearing loss adapt to hearing in reverberant environments: Using stimuli processed by a gammatone filter bank

◎辻慎也, 荒井隆行(上智大・理工)

- ◆我々がこれまでに行ってきた検討により、1) 一側性難聴では特にその発症直後、残響により聴取が影響されること、2) 残響下での聴取には次第に適応していくことが示唆された。
- ◆一側性難聴者 (UHL)・その発症直後の模擬 (MNH)・両耳聴 (BNH) を対象に、無響/残響の音声にガンマトーンフィルタバンクを適用した刺激によって文章了解度を測定し、次の結果を得た (Fig. 1)。

- 1) 無響の刺激では、Broadband・High frequencyでUHLはMNHと比べ有意に優れた成績を示した。
- 2) 残響を付与した刺激では、Low frequencyで残響による影響が顕著であったが、それを含むBroadbandでもUHLの音声明瞭度は保たれた。
- 3) 高域への適応と同時に、残響の影響が大きい帯域からの音声明瞭度への影響を抑えるような適応が生じることが示唆された。

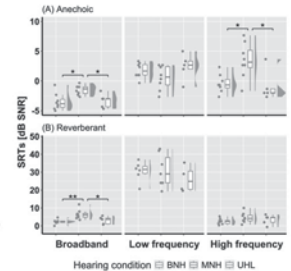


Fig. 1 Raincloud plots for the speech reception thresholds (SRTs) including raw jittered data, box-whisker plots, and split-half violins for three hearing conditions: BNH (n=8), MNH (n=8), and UHL (n=5). The solid horizontal lines indicate the median. ** $p < 0.01$, * $p < 0.05$.

3-7-1

3-7-1

STの音響分析現状調査 2024

Survey on the practical status of acoustic analysis
by speech therapists 2024

○竹内京子(順天堂大), 青木直史(北大), 荒井隆行(上智大),
大金さや香(国際医療福祉大), △鈴木恵子, 秦若菜, △村上健(北里大),
世木秀明(千葉工大), 安啓一(筑波技大)

- ◆言語聴覚士の養成校では、音響学が必修科目で、学生が最も苦手な科目である。この傾向は卒業後も変わらない。
- ◆2021年より、現役言語聴覚士対象の講習会「STのための音響学」を開催してきた。講習会では、養成校で習う音声の音響分析の実践の可能性を多方面から追求してきた。
- ◆その結果、臨床での音響分析の現状は、2021年調査から3年間で変化したのだろうか。2024年度実施のアンケート結果を報告する。
- ◆防音室、聴覚検査室での録音が増えた。音声の録音も耳だけの評価から、音響分析を意識することによって、できるだけ騒音のないところという流れになってきている。
- ◆パソコン、タブレット、スマートフォンによる録音が増えた。コロナ後の音響学の授業は、音響分析実習が、各段にやり易くなった。それと同じ現象が臨床にも波及しているのかもしれない。
- ◆3年間で、急激に「臨床の音響分析の環境」が変わったということはないかもしれない。しかしながら、「音響分析で具体的に何がしたいか」について回答する者が増えてきているように感じる。
- ◆今後は、これらの希望がかなうように、実現可能な方法を示していければと考えている。

3-7-3

3-7-3

飲食店の音環境評価アプリケーションの開発その3
-USEN社員を対象としたアプリケーションの試行-
Development of an application for evaluating the sound
environment of restaurants -Part 3.

○上田麻理, 田中哲雄(神奈川工科大),

池上卓也, △松本茂雄(USEN), △津本祐一(乃村工藝社)

我々はより良い音環境を構築するためには、音の問題や快適さを自分事化することも重要であると考えている。前報では空間デザイナーを対象として自作のアプリケーションによる飲食店の音環境評価を報告した。今回は、音(環境)を運営・技術的サポートする企業の社員を対象としてアプリケーションを用いた音環境評価を行ったのでその概要を紹介する。



Fig. 今回は USEN 株式会社の社員を対象として音環境の意識付けのための飲食店の音環境評価を試した(写真は USEN の新しいスタジオ)

3-7-2

3-7-2

「聴能形成」のトレーニング成績と騒音感受性の関係

Relationship between training performance of Technical Listening Training and Noise Sensitivity

○金森克洋(神奈川工科大), 廣江正明(小林理研), 上田麻理(神奈川工科大)

- ◆「聴能形成」とは、音に関わる業務で活躍する人材に必要な音の聴き分け能力や、音を数値化する力を養う能力の育成である。
- ◆筆者らは情報メディア学科 2 年生を対象とする演習授業内で、この「聴能形成」訓練を 40 人程度の学生に対し実施している (Fig.1)。
- ◆訓練の成果には大きな個人差が存在するが、その要因については検討されておらず音に対する敏感さも 1 つの要因と推定される。
- ◆2022 年度と 2023 年度にそれぞれ半期にわたって聴能形成トレーニングを実施し、同時に音に対する敏感さとして騒音感受性も継続的に調査して両者の関係を調べたのでその結果を報告する。

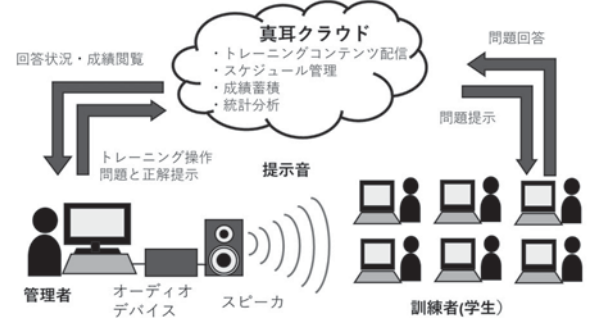


Fig.1 Image of Technical Listening Training class using "Shinji-online" system

3-7-4

3-7-4 感覚特性に配慮した音環境調整に関するワークショップの実践

Workshop on sound environment adjustment
considering sensory characteristics

☆志水祐貴, 上野佳奈子(明治大), △高橋秀俊(高知大)

- ◆学習障害 (LD) 等の発達障害に関する研究・臨床・教育の関係者が多く参加する日本 LD 学会大会において、感覚特性に配慮した音環境調整に関するワークショップ (以下、WS) を実施した。
- ◆WS 前半は、騒音レベル、室内の響き、遮音と吸音、聴覚特性の多様性 (特に自閉スペクトラム症の聴覚的非定型性)、局所的な音環境調整の方法や活用事例について、ミニレクチャー (20 分) を行った。
- ◆WS 後半は、3 グループに分かれ、パネルで囲った空間での吸音効果の体験、自作可能な音環境調整用補助具の体験 (Fig.1)、スマホアプリによる騒音レベルの測定の 3 つ体験ワーク (25 分) を行った。
- ◆2 日間計 8 回 (1 日 4 回) 実施し、各回約 20 名、合計 155 名が参加した。参加者の感想から、WS を通じて音環境に対する理解が促進され、環境調整の導入の一助とすることができたことが窺えた。



Fig.1: Spatial settings of sound environment adjustment in the workshop

3-7-5

3-7-5 音声と音響の可視化における 3Dプリンターとロボットアームの応用について

Using a three-D printer and a robot arm for visualization of speech and acoustics
○河原英紀(和歌山大), 榊原健一(北海道医療大), 北村達也(甲南大)
水町光徳(九工大), 矢田部浩平(東京農工大)

- ◆3Dプリンターとロボットアームを導入したことにより、音声・聴覚、音響に関連する様々な現象を可視化と併せて体験することが容易になっている。ここでは、その使用例の幾つかを紹介した。
- ◆これらに用いたプログラムや関連情報をGitHubで公開しているので利用し、フィードバックをお願いしたい。

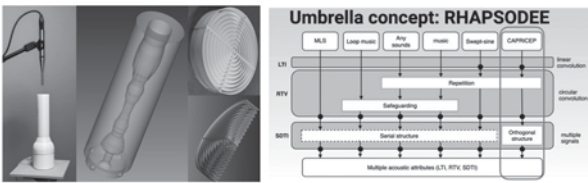


Fig.1:3D-printed acoustic components and underlying concept of measurement and visualization (top left: acoustic field measurement, top right: head model for binaural recording and open ear headphone, bottom left: vocal tract model and measurement, and acoustic lens, bottom right: underlying concept of measurement and visualization.)

3-8-2

3-8-2 弾性表面波バイオセンサを用いたヘルスマニタリング

Health monitoring using surface acoustic wave biosensor

○近藤 淳(静岡大・創造), △鄭家亘, △谷津田博美(静岡大・創造, tst-bio)

- ◆弾性表面波 (SAW) デバイスは、スマートフォンなどのフィルタ・デュプレクサとして現在社会に欠かすことが出来ない。
- ◆圧電結晶表面を伝搬する SAW は、表面上の物理的または化学的变化によりその伝搬速度や振幅が変化する。これらの変化量を検出するとセンサに応用できる。液体を測定対象とする場合、横波型弾性表面波 (SH-SAW) と呼ばれるモードを利用しなければならない。
- ◆SH-SAW センサの重要な応用の一つは酵素反応や免疫反応を検出するバイオセンサである。
- ◆図1は3種類のコレステロールをSH-SAW センサを用いて測定した結果である。この結果の表現を変えると、コレステロール分子サイズが評価可能となる。

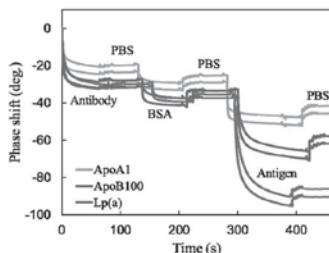


Fig.1: Measurement results of three kinds of cholesterol. PBS: phosphate buffer solution, BSA: bovine serum albumin

3-8-1

3-8-1 MEMS へのいざない 驚きの技術と広がる応用

Invitation to MEMS
Their Surprising Technology and Expanding Application
○田中秀治(東北大・工)

MEMS (Micro Electro Mechanical Systems) について全体像を30分で話して欲しいという光栄な「無茶ぶり」に対して、どうすべきか考えた結果、この講演を私のインターネット記事とリンクさせることにした。予稿には記事にアクセスするためのQRコードが24個、掲載されている。その半分以上のリンク先はXのポストであるが、1つの画像と140字以内の短い説明文で構成されるので、さっと概要を掴めると思う。大量生産するコンシューマー向けMEMSから、少量高付加価値の産業向けMEMSまで様々な種類のMEMS製品があるが、この講演では両極端を紹介する。具体例として取り上げるMEMSは、日本音響学会での講演であることも考慮し、マイクロフォン、BAWフィルター、データセンター用光スイッチ、および半導体製造装置に用いるMEMSである。予稿ではジャイロセンサーにも触れた。本講演を通じて皆さんがMEMSにいざなわれ、引き続き興味を向けて下さると幸いである。

3-8-3

3-8-3 弾性波素子を活用した原子時計の開発とその展望

Developments and prospects of atomic clock using acoustic wave devices

○原 基揚(情報通信研究機構)

原子時計は原子の吸収スペクトルを用いて、発振器を安定化する技術である。この原子時計をチップレベルにまで小型化できれば、あらゆる端末に周波数標準を内蔵することが可能となる。通信端末に目を向けると、通信端末は基地局との同期によって、高い時刻精度を持っている。ただし、これは、基地局からのランチを前提としたツリー状ネットワークを仮定したものであり、我々が開発している、端末間での双方向通信による時刻同期技術を活用すれば、無線端末のD2D (Device-to-Device)の接続を介して、相互補完的でシームレスなメッシュ状の時刻同期網を構築できるであろう。そして、これらに小型の原子時計を用いた高安定な周波数源を援用できれば、複数端末の相互時刻同期によるアンサンブル処理を用いて、一個の端末では実現できないレベルでの安定性を示す同期時刻系を生成・維持することも可能となる。6Gのような次世代の通信網においてISAC(Integrated Sensing & Communication)やAIの活用(Integrated AI and Communication)が新たなサービスの柱として注目されているが、これは、基地局からの電波放出を通信のみではなく、センシングにも活用し、様々なユースシーンの創出を促すことを企図している。これらを実現するための鍵は、無線基地局間の高い時刻同期であり、仮に、前出の時刻同期技術の無線ネットワーク全体への波及が実現されるならば、上記の6Gの戦略においても新たな可能性が提供でき、ユースシーンの充実と利便性のより一層の向上が期待される。

3-8-4

3-8-4 BAW 共振子・フィルタの最新研究動向 -高周波動作への検討-

Latest trends in BAW resonators and filters for high-frequency operation

○鈴木雅視(山梨大)

本講演では次世代通信規格に向けた BAW 共振子・フィルタの高周波化に対する課題と様々な研究機関での検討(デバイス構造や性能)を紹介するとともに、山梨大で開発中の分極反転 SiAlN/AlN 多層膜 BAW 共振子について報告する。Al 極性 AlN 膜と N 極性 SiAlN 膜を交互に積層することで分極反転構造が形成でき、分極反転 SiAlN/AlN 多層膜 HBAR では高次モード共振を確認した。また AlN 単層膜より大きい膜厚と約 9 GHz 動作を両立する 2 次モード共振分極反転 SiAlN/AlN 膜 SMR の形成にも成功し(Fig.1), FEM 解析で発見した分極反転構造によるデバイス性能改善を実験的に実証し、分極反転構造が BAW 共振子の高周波動作に有効であることを明らかにした。

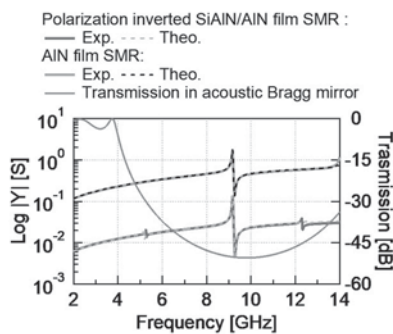


Fig.1 Admittance frequency characteristics in two-layer polarization inverted SiAlN/AlN film and single-layer AlN film SMRs.

3-8-6

3-8-6 4 mm 角の放射面を有する 40 kHz 用ボルト締めランジュバン型振動子の 基礎検討

Basic study of bolt-clamped Langevin type transducer for 40 kHz with 4 mm square emitting surface

☆塚本悠介, 三浦 光, 浅見拓哉(日大・理工)

- ◆本報告では、空中を伝搬する超音波の半波長以下となる 4 mm 角の放射面を有する 40 kHz 用ボルト締めランジュバン型振動子の初期検討として、FEM を用いた設計並びに試作結果について報告する。
- ◆Fig.1 は振動子の概形である。ネジの先端形状はテーパとなっており、振動子の中心にある圧電素子と銅電極をステンレス製ブロックを介して固定している構造である。有限要素法解析を用いて、振動子が 40 kHz 付近で縦振動の半波長共振となる振動モードを求めた。そして、求めた振動モードの電気機械結合係数を求めることで評価を行い、適切な寸法を求めた。その結果、ネジの長さが 23.5 mm、テーパの長さが 2.0 mm、振動子の長さは 61 mm、共振周波数は 39.915 kHz が適切な寸法となった。

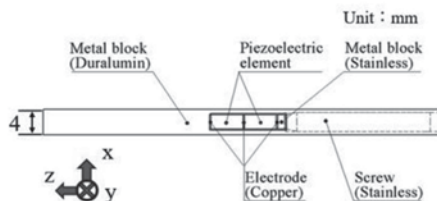


Fig. 1 Schematic of bolt-clamped Langevin type transducer

3-8-5

3-8-5 Society 5.0 実現のための環境発電の 最新動向

Recent Progress in Eenergy Harvesting for Society 5.0

○山田駿介, △本田崇(九工大・工)

- ◆様々なものを無線センサネットワークにつなぎ、センサにより得られる情報とその AI 解析により豊かな社会 Society 5.0 の実現を目指す取り組みが進められている。
- ◆しかしながら、無線センサ端末の数は年間数兆個と試算されており、その膨大な数の端末のメンテナンスを人力で行うことは困難であり、電池の交換ができないという課題が存在する。
- ◆そこで近年注目を集めているのが、環境中の存在するエネルギー源から発電し蓄電素子を充電できる環境発電素子である。
- ◆本発表ではは振動発電素子の最新動向を紹介するとともに、タイマー型やイベントドリブン型など無線センサ端末の駆動方法を紹介する。最後に、無線センサ端末の研究動向に関して求められる特性に関して述べる。



Fig. 1 Schematic image of Society 5.0 with wireless sensor networks.

3-8-7

3-8-7 積層型超音波アクチュエータを用いた 広帯域超音波音源の開発 -アルミニウムフィルムの振動特性の検討-

Development of a wideband ultrasonic sound source using a multilayered ultrasonic actuator

-A study on vibration characteristics of aluminum film-

☆高橋晴輝, 浅見拓哉, 三浦 光(日大・理工)

- ◆本研究では、空中超音波による非接触での非破壊試験の欠陥検出性の向上を図り、空中超音波をチャープ波として発振させるために広帯域で駆動可能な超音波音源の開発を検討している。
- ◆Fig. 1 は開発する広帯域超音波音源の概要である。積層型超音波アクチュエータに音波放射のための振動板としてアルミニウムフィルムを接着した構造である。本検討では 40 kHz におけるアルミフィルムの振動特性について、有限要素法を用いた周波数応答解析を行った。
- ◆その結果、最適なアルミフィルムの寸法は一邊が 6.0 mm、厚さが 0.1 mm と求めた。この時、アルミフィルムはその中心で最大変位が得られるような振動モードであった。

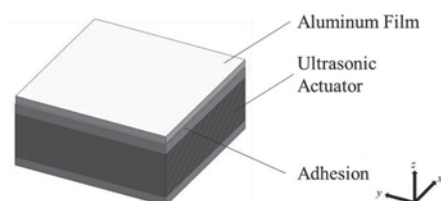


Fig. 1 Wideband ultrasonic sound source.

3-8-8

3-8-8 駆動時間が空中超音波音源の特性に与える影響に関する基礎検討

Basic study on the effect of driving time on the characteristics of airborne ultrasonic sound sources

☆柿沼優太, 浅見拓哉, 三浦 光(日大理工)

- ◆当研究室では、超音波技術の発展に向け、種々の超音波音源の開発に取り組んでいる。特に、商用化を視野に入れた技術の実用化に向けては、音源の耐久性に関する評価が不可欠である。本研究では、空中超音波音源の耐久性を定量的に評価することを目的としている。本稿では、初期検討として、当研究室で開発した 58 kHz 空中超音波音源を用いて、駆動時間に対する音源の特性の変化について検討した。
- ◆Fig.1 は駆動時間に対する音源の特性変化の評価方法である。音源の各特性の測定と 48 時間の駆動の一連の手順を繰り返すことで、駆動時間に対する音源の特性変化を評価した。また、48 時間の音源駆動中は、音源内部に生じる最大のフォンミーゼス応力が 120 MPa となるよう制御を行った。120 MPa に着目した理由は、本研究で取り扱う音源の材質であるアルミニウム合金 A2017-T4 の疲労強度が、一般的に 120 MPa であると報告されているためである。
- ◆その結果、音源に最大のフォンミーゼス応力 120 MPa を加えた場合でも、音源の電気的特性に明確な変化は見られなかった。

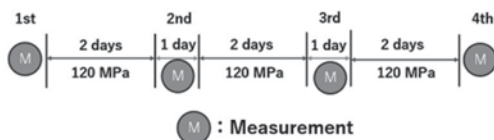


Fig. 1 Measurement method

3-8-10

3-8-10 100 kHz 以上の周波数での微小物体の空中超音波浮揚

Ultrasonic levitation of small object in air at beyond 100 kHz

☆橋本 花香, 和田 有司, 中村健太郎(東京科学大学)

- ◆空中定在波超音波による微小物体や液滴の浮揚がこれまで様々検討されてきた。それらのほとんどが数 10 kHz での実験であった。
- ◆より小さな物体の浮揚には高周波化が必要であり、100 kHz におけるチップ部品の浮揚が検討されている。
- ◆本報告では超音波周波数と浮揚物体の大きさの関係を確認して、100 kHz を超える周波数での浮揚を試みた。
- ◆共振周波数が数 10 kHz のボルト締めランジュバン振動子を高次モードで動作させることで発泡スチロール片を浮揚させることができた。
- ◆より高周波数での浮揚実験のため 200 kHz 振動系を試作した。

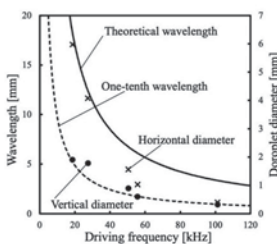


Fig. 1: Relationship between driving frequency and droplet size.

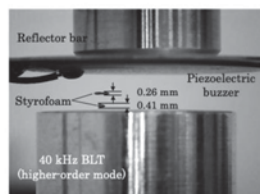


Fig. 2: Levitation at 158 kHz.

3-8-9

3-8-9 金属 3D プリンターにより造形した工具の性能測定 —超音波振動工具のトポロジー最適化(7)—

Vibration measurement of tools fabricated by metal 3D printer —Topology optimization of large ultrasonic tools (7)—

○和田有司, 中村健太郎(東工大・未来研)

- ◆超音波接合等を目的とする作用面が平坦に振動する工具をトポロジー最適化を用いて設計する方法を検討している。
- ◆複雑な三次元形状を有するトポロジー最適化の結果を造形する方法の一つとして、金属 3D プリンターが注目されているが、造形物中に少なからず空隙が発生することが知られており、母材金属と同様の弾性定数を有するかどうかは定かではない。
- ◆Fig. 1 のような金属棒を切削加工および 3D プリンターで積層方向を変えて造形し、共振特性を測定した。ヤング率は切削材に比べて奥行方向積層で 1%、長さ方向積層で 10%低下した。
- ◆Fig. 2 のようなトポロジー最適化による振動工具を造形し、振動分布を測定した。造形した工具は切削加工の工具に比べて振動モードに若干の変化がみられた。

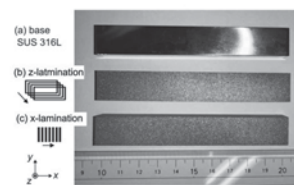


Fig. 1: Photograph of the base and 3D printed blocks.

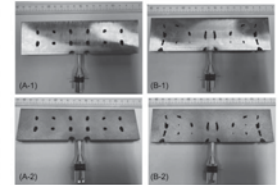


Fig. 2: Photograph of (A-1)(B-1) machined and (A-2)(B-2) printed ultrasonic tools of uniform vibration.

3-8-11

3-8-11 水中での MHz 帯超音波の受信 超音波受信のためのカーボン・トランスデューサとその特性(2)

Receiving MHz-ultrasounds in water, Carbon-based transducer for receiving ultrasound and its characteristics (2)

☆松本 一伸, 和田 有司, 中村健太郎(東京科学大学)

- ◆超音波イメージングの空間分解能の向上のためには高周波化が必要だが、減衰も増大するため限界がある。そこで、感度の高いトランスデューサが必要である。
- ◆かつて可聴域で使われたカーボン型トランスデューサは増幅器無しで受話器を鳴らすことができるほど高出力である。
- ◆これまで、カーボン型が超音波領域で使われたことはほとんどなかったが、前報において新たな薄型構造のカーボン型トランスデューサを考案し、数 10 kHz の空中超音波が受信可能であることを示した。
- ◆本報告では、この構造により水中での MHz 超音波超音波の受信を試みた。
- ◆受信した信号の時間波形には距離に応じたピークが見られ、水中超音波を受信することができた。

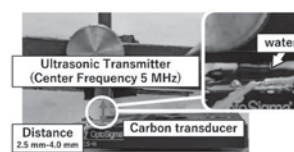


Fig. 1 Experimental setup for receiving MHz ultrasonic signals in water.

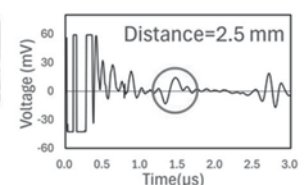


Fig. 2 Received ultrasonic signal

3-8-12

3-8-12 超音波振動と液晶を用いた可変NDフィルタの基礎検討

Basic study of the variable ND filter using ultrasonic vibration and liquid crystal

☆水野稜山¹, 吳沢農¹, △江本顕雄², 小山大介¹(¹同志社大, ²徳島大)

- ◆透明電極と機械的可動部を必要としない、超音波振動とねじれ液晶(TN)配向を用いた可変NDフィルタを開発した。
- ◆超音波非印加時は画像全体が暗い(Fig. 1(a))が、超音波印加によって画像中央部が明るくなった(Fig. 1(b))。
- ◆NDフィルタに印加する電圧が大きいくほど、画像の平均輝度が増加した(Fig. 1(c))。
- ◆TN配向した液晶層に対して超音波振動を励振することで、透過光強度を調整可能な可変NDフィルタとして動作した。今後、光学機器の薄型化・低価格化が期待できる。

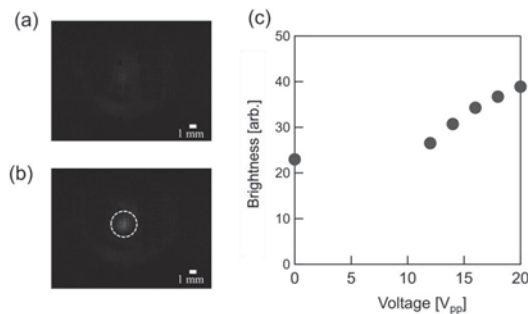


Fig. 1: Photographs through the LC ND filter (a) without and (b) with ultrasound excitation. (c) Relationship between input voltage and brightness level in photographs.

3-P-1

3-P-1 状態空間モデル Mamba に基づく end-to-end ニューラル話者ダイアライゼーション

End-to-end neural speaker diarization using state space model Mamba

○俵直弘, Plaquet Alexis, デルクロア・マーク, 堀口翔太, 安藤厚志, 荒木章子 (NTT)

- ◆状態空間モデルのニューラルネットワーク拡張であるMambaを用いたend-to-endニューラル話者ダイアライゼーションシステムを提案
- ◆提案法はEENDセグメンテーションモデル(Fig. (a))のデコーダにMamba(Fig. (b))を用いることで、より長いコンテキストを考慮可能
- ◆複数の自然対話音声データセットで、従来のlong short-time memory (LSTM) やアテンションに基づく手法よりも高い性能を達成

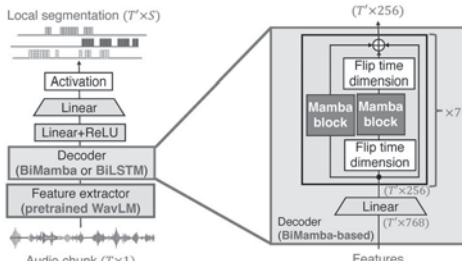


Fig.(a) Architecture of local segmentation Fig.(b) Proposed Mamba-based decoder

Table: Performance comparison with Diarization error rate (DER)

System	Macro average DER [%]	
	In-domain	Out-domain
LSTM	16.7	17.9
Attention	18.2	24.9
Mamba (proposed)	16.2	16.7

3-8-13

3-8-13 超音波振動がC2C12細胞の成長に与える影響

Effect of ultrasonic vibration on C2C12 cell growth

☆市川崇, ○小山大介(同志社大・理工), △桑田昌宏(京大・生命)

- ◆再生医療分野における細胞培養制御技術として、超音波振動を用いた細胞培養制御について検討した。
- ◆培養環境下の細胞に超音波振動刺激を与えることが可能な超音波細胞培養ディッシュを開発した(Fig. 1(a))。
- ◆ディッシュ底面に超音波振動を発生させ、細胞分裂および分化を観測することにより超音波が細胞成長に与える影響を評価した。
- ◆C2C12細胞を培養ディッシュに播種してから48時間超音波振動を与え、48時間毎に計8日間細胞を観察した。
- ◆本評価手法では、超音波振動がC2C12細胞の成長に与える有意な影響は確認できなかった(Fig. 1(b))。今後、超音波振動を与えるタイミングを検討する必要がある。

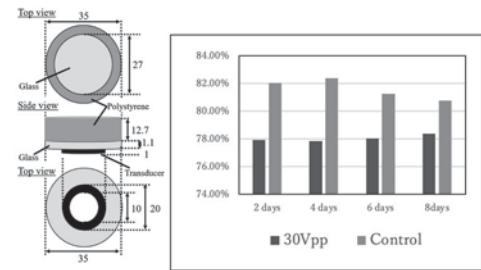


Fig. 1(a) Ultrasonic cell culture dish and (b) changes in the C2C12 cell area with respect to time.

3-P-2

3-P-2 話者遷移確率に基づく話者ダイアライゼーションのためのデータ生成

Data generation for speaker diarization based on speaker transition probabilities.

☆市川奎吾, 上乃聖, 李晃伸(名工大)

- ◆End-to-end neural diarizationでは、複数人の単一話者音声を重畳させることにより生成した音声を学習に利用する。
- ◆本研究では生成データの話者交替を実際の多人数会話に近づけるために、話者遷移確率に基づくデータ生成法を検討する。
- ◆評価実験では、3, 4話者音声において性能改善がみられた。先行研究SC法の話者順決定が生成に利用するデータセットに依存することによるEENDの性能低下が引き起こされる可能性が示唆された。

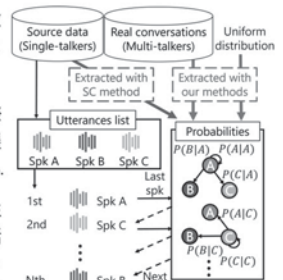


Fig. 1: overview of determining speakers with probabilities

Table 1: DERs (%) evaluated on the CALLHOME 4-speaker scenario

Speaker transition	Miss	False	Conf.	DER
SC [baseline]	8.2	7.2	17.8	33.2
AMI	8.5	7.6	16.4	32.4
AMI + CALLHOME 2	8.7	6.4	18.9	34.0
AMI + DIHARD 3	9.2	7.0	17.5	33.7
CALLHOME 2 + DIHARD 3	8.8	6.3	17.5	32.7
AMI + CALLHOME 2 + DIHARD 3	9.0	7.4	16.1	32.4
CALLHOME 1 (oracle)	7.7	8.2	17.6	33.5
Uniform	8.3	7.0	17.0	32.3

3-P-3

3-P-3 DNN 音声強調フロントエンドを組み合わせた End-to-end ニューラルダイアライゼーションモデルの検証

A verification of combining DNN-based speech enhancement front-end with end-to-end neural diarization model

★山口拓生(都立大), 俵直弘(NTT), 佐藤宏(NTT), 塩田さやか(都立大)

- ◆話者ダイアライゼーションは、複数話者の会話音声から各話者の発話区間を推定するタスク。
- ◆残響や雑音を多く含む実環境下の音声では精度が低下。
- ◆DNN に基づく音声強調フロントエンドによる雑音・残響除去でダイアライゼーションの精度改善が期待される。
- ◆本検証では、既存の話者ダイアライゼーションシステムに音声強調フロントエンドを導入。
- ◆マルチドメイン遠隔音声認識タスクに適用し、Figure 1のA~Dすべてで音声強調することの有効性を確認。

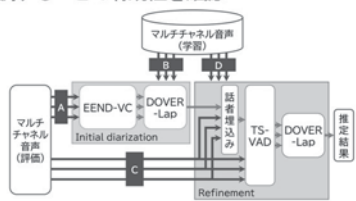


Figure 1: An overview of the proposed diarization system

Table 1: DER using EEND-VC initial diarization with speech enhancement

Speech enhancement on Initial diarization		DER (%)
Evaluation (A)	Training (B)	
✓		17.7
	✓	18.0
✓	✓	17.6

Table 2: DER using TS-VAD refinement with speech enhancement

Speech enhancement on Refinement		DER (%)
Evaluation (C)	Training (D)	
	✓	15.4
✓		15.2
✓	✓	15.1

3-P-5

3-P-5 Towards Effective Spoofing Detection in Non-Native English Speech: Indonesian and Thai Case Study

★ Aulia Adila, Candy Olivia Mawalim, Masashi Unoki (JAIST)

- ◆Aim: Investigate the development of spoofing countermeasures (CMs) for non-native English, focusing on Indonesian and Thai.
- ◆Problem: Current CMs face challenges detecting spoof speech due to non-native pronunciations, compounded by a lack of tailored datasets for these contexts.
- ◆Solution: Developed a comprehensive dataset with English native and Indonesian-Thai non-native speech and constructed two types of CMs—Native CM and Combined CM—using common features and classic machine learning classifiers.
- ◆Evaluation: Assessed CMs on native and non-native speech using minimum detection cost function (minDCF) and EER.
- ◆Summary: Results reveal the limitations of Native CM in addressing non-native speech. Incorporating non-native speech data improved the Combined CM's detection performance, emphasizing the need for tailored datasets.

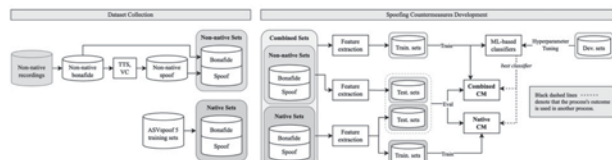


Fig.1: Block diagrams of our proposed method

3-P-4

3-P-4 Challenges in Speech Spoofing Countermeasures for Southeast Asian Languages

◎ Candy Olivia Mawalim (JAIST), Kasorn Galajit (NECTEC), Dessi Puji Lestari (ITB), Win Pa Pa (UCSY), Masashi Unoki (JAIST)

- ◆Aim: To investigate the development of speech spoofing countermeasures for Southeast Asia (ASEAN) languages.
- ◆Problem: The significant challenges in speech spoofing countermeasure development for ASEAN languages include limited high-quality datasets, inconsistent data quality, the emergence of sophisticated attacks, and the limited generalizability of existing models.
- ◆Solution: Developing dedicated datasets, utilizing advanced models, and continuously researching and adapting to evolving spoofing techniques are crucial for robust speech spoofing countermeasures for ASEAN languages.
- ◆Evaluation: The spoof countermeasures were evaluated based on equal error rate (EER) and accuracy.
- ◆Summary: Despite promising initial results, the existing datasets exhibit limitations in terms of the diversity of spoofing methods employed and the number of speakers represented.

Table 1: Existing Spoofing Countermeasure Datasets in ASEAN Languages

Name	Year	Language	Dialects	Spoof Type	Environment	#Speakers
ThaiSpoof	2023, 2024	Thai	1	TTS	Clean, Noisy	7,784
InaSpoof	2024	Indonesian	8	TTS, VC, Replay	Clean, Noisy	~ 300
VSASV	2024	Vietnamese	1	VC, Replay, Adversarial	Clean, Noisy	1,382
UCSYSpooF	2024	Myanmar	1	TTS, VC	Clean	3

3-P-6

3-P-6 聴取実験に基づく 深層電子透かし最適化基準の設計

Design of deep speech watermarking optimization criteria based on listening experiments

○尹道鉉, 安田裕介, 戸田智基(名古屋大学)

- ◆客観指標 (信号対雑音比, ガンマトーングラム歪み) におけるノイズの聴取実験 (DMOS 評価, 動的比較評価) を実施した
- ◆主観的な評点と客観指標間の関係性を調査した
- ◆客観指標を基に DMOS を予測する線形式を導入し、電子透かしの最適化基準として使用できることを示した
- ◆最適化の結果、隠蔽性の高い電子透かしが生成できることを確認した

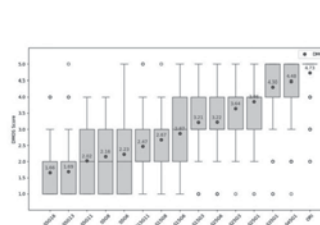


Fig.1: Result of DMOS evaluation for utterances with different noise level

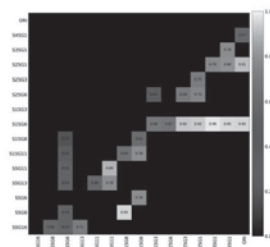


Fig.2: Result of dynamic comparative evaluation

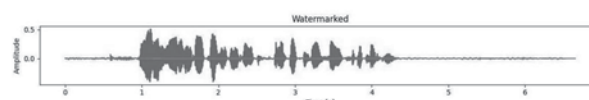


Fig.3: Example of optimized watermark (blue: original, orange: watermark)

3-P-7

3-P-7 大規模言語モデルとクラウドソーシングを活用した感情キャプション収集とその評価

Collecting and Evaluating Emotion Captions Using the Large Language Model and Crowdsourcing

☆永瀬 亮太郎, 福森 隆寛, 山下 洋一(立命館大)

◆音声感情キャプションのためのデータ収集

- 音声感情キャプションニング:
音声を聴いて感じた感情の説明文を入力音声から推論すること
- 音声に対して感情キャプションを手で付与するには
時間と手間がかかる

◆GPT4とクラウドソーシングで半自動的に感情キャプションを収集

- 既存の感情音声コーパスに感情キャプションを付与
- 人が記述した感情キャプションと比較し妥当性を評価

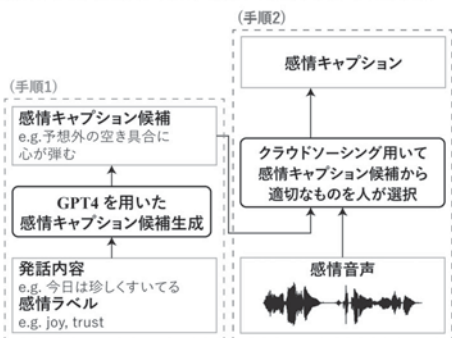


Fig.1: Outline of labeling emotion captions

◆半自動収集した感情キャプションは人が記述した感情キャプションと同等の妥当性があることを確認

3-P-9

3-P-9 Speaking style 空間における対話相手に依存したパラ言語情報分類

Classification of Paralinguistic Information Dependent on Dialogue Partner in Speaking Style Space

☆加藤 隆聖, 大浦 杏奈, 菊池 英明(早大文科)

- ◆Speaking style 空間は明瞭さ・親しさ・社会階層によって構成され、さまざまな発話スタイルを分類するために用いられる。
- ◆本研究では、speaking style 空間において対話相手との関係性を考慮したほうがしない場合よりもパラ言語情報を正確に分類できるかを検証した。
- ◆実験では、20代女性の1秒以上2秒以下の音声の音声を用いた。
- ◆被験者に音声を聴取させ、パラ言語情報(「肯定・同意・相槌」「考え中・戸惑い」「感心・羨望」「聞き返し」「驚き・意外」「疑い」「非難・嫌悪」「不満」「中立」)9種類に分類させた。
- ◆各クラスタリング手法のパラ言語情報識別率の結果を Fig. 1 に示す。どのクラスタリング手法でも、対話相手依存の方が対話相手非依存より精度が高くなることを確認した。

	k-means法	階層型クラスタリング	GMM	DBSCAN	スペクトルクラスタリング
対話非依存	24.12%	21.03%	23.09%	31.52%	18.68%
対話依存の平均	27.61%	28.24%	28.38%	32.39%	24.13%
話者A	25.44%	27.19%	27.19%	35.45%	22.81%
話者B	32.69%	36.54%	34.62%	34.09%	30.77%
話者C	29.21%	29.21%	31.46%	28.75%	24.72%
話者D	29.59%	26.53%	24.49%	23.71%	25.51%
話者E	25.55%	26.28%	24.09%	36.30%	20.44%
話者F	23.16%	23.68%	28.42%	36.02%	20.53%

Fig.1: Comparison of Clustering Results for Paralinguistic Information

3-P-8

3-P-8 Crossmodal Transformer とアテンション重み修正に基づく音声感情認識の検討

Speech Emotion Recognition Based on Crossmodal Transformer and Attention Weight Correction

☆井原 龍成, 山田 武志(筑波大)

- ◆音声感情認識(SER)とは、人の発話に含まれる感情を識別する技術である。
- ◆近年では、音響特徴に加えて音声認識(ASR)テキストから得られたテキスト特徴も用いる SER 手法が主流となっており、さらに Crossmodal Transformer Encoder を用いて音響特徴とテキスト特徴をクロスモーダル統合する手法が提案された。
- ◆しかし、感情を顕著に含む音声は ASR エラーが頻発する。ASR エラーに対処するために、Santoso らは Bidirectional Long Short-Term Memory を用いた SER 手法において単語信頼度に基づくアテンション重み修正を提案した。
- ◆本稿では、SER 精度をさらに向上させるために Crossmodal Transformer Encoder とアテンション重み修正を組み合わせた手法を提案する。
- ◆音声感情認識のベンチマークとして広く用いられる IEMOCAP データセットを用いて実験を行い、提案手法の有効性を確認した。

Table 1 SER performance of the base method and the proposed methods.

Method	UA (%)	WA (%)
Singlemodal	73.4	75.7
Crossmodal (base method)	74.7	76.9
Proposed 1	75.4	77.5
Proposed 2	75.3	77.2
Proposed 3	75.5	77.4

3-P-10

3-P-10 補助情報を同時に認識する日本語音声認識における出力表現形式の比較

Comparison of Output Formats for Japanese Speech Recognition with Auxiliary Information

☆小堀 聡太, 藤江 真也(千葉工大)

- ◆補助情報を検出する際の最適な出力表現形式についての検討
補助情報にはフィラー: F, 言い直し: D, 感動詞: I を選出
- ◆従来手法では、仮名文字と漢字(書字形)をトークンに使用
 - 補助情報を表す特殊トークンとして、フィラーに(F, 感動詞に(I, 言い直しに(D を付与(独立トークン方式)
 - ◇ 同音語の中で補助情報と異なるもの区別が困難
 - ◇ 特殊トークンを前に配置(独立トークン前置方式)
 - 例: (I ウン)(F エー)(D キョ)キョウハ
- ◆提案手法ではカタカナ文字のみ(発音形)をトークンに使用
 - 補助情報の検出方法として、フィラーに+F, 感動詞に+I, 言い直しに+D を付与(複合トークン方式)
 - ◇ 補助情報ではない同音語に対して別のトークンとして認識することが可能
 - ◇ 例: ウ+ン+I エ+F +F キョ+D キョウハ
- ◆提案手法がすべての補助情報に対して高い性能を達成 (Table.1)

Table.1 experimental results

補助情報	手法	書字形			発音形		
		適合率	再現率	F1 値	適合率	再現率	F1 値
フィラー	独立トークン前置方式	82.51	78.33	80.36	81.76	81.26	81.51
	複合トークン方式	81.08	77.65	79.33	83.19	80.92	82.04
感動詞	独立トークン前置方式	86.48	87.71	87.09	86.15	86.12	86.14
	複合トークン方式	86.85	85.92	86.39	87.67	86.66	87.16
言い直し	独立トークン前置方式	41.06	37.01	38.93	41.11	37.14	39.02
	複合トークン方式	40.49	36.17	38.20	42.72	36.97	39.64

3-P-11

3-P-11 HuBERT の中間層別特徴量を用いた 相槌生成タイミングの予測

Prediction of Backchannel Timing Using
HuBERT Intermediate Layer Features

☆杉本飛馬, 西村良太(徳島大), 太田健吾(阿南高専), 北岡教英(豊橋技科大)

- ◆本研究では、対話音声を入力として相槌の生成タイミングを予測する深層学習モデルを提案する。
- ◆従来の手法では言語的特徴が必要とされていたが、提案手法では音響的特徴のみを使用して相槌の生成タイミングを予測する。
- ◆相槌発生から過去数秒間の対話音声を HuBERT に入力し、得られる特徴量を用いて LSTM と全結合層からなる予測モデルを学習する。
- ◆入力音声長や HuBERT の中間層別特徴量による精度比較を行う。
- ◆2 秒間の入力音声で HuBERT の 11 層目の特徴量を使用した結果、F 値 80.3% を達成し、従来手法を大きく上回った。

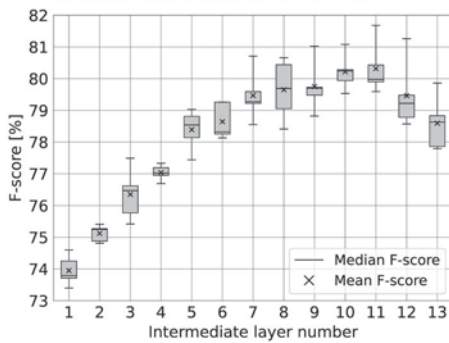


Fig.1: Comparison of F-score with HuBERT intermediate layers using 5-fold cross-validation

3-P-13

3-P-13 話者オーバーラップ音声からの特徴抽出に 向けた自己教師あり学習モデルの検討

Self-supervised learning model for feature extraction from speaker
overlapped speech

☆浅井航平, 齋藤佑樹, 中田亘, 関健太郎, 猿渡洋(東大院・情報理工)

- ◆近年、音声に関する様々なタスクで SSL モデル特徴量が用いられるが、現状の SSL モデルでは話者オーバーラップ音声の特徴をうまく捉えることができない。
- ◆本研究では、話者オーバーラップ音声からの特徴抽出に向けて、話者オーバーラップ音声の特徴量が話者ごとの特徴量に対して加法的になっている SSL モデルである Add-HuBERT を提案する。それに加えて、Add-HuBERT モデルによる特徴量から話者ごとの特徴量を分離する Sep-HuBERT を提案する。
- ◆Add-HuBERT と Sep-HuBERT を用いて話者オーバーラップ音声の分離を行ったところ、発話内容はある程度保持できたものの、話者性は分離前後で変化してしまうという結果が得られた。

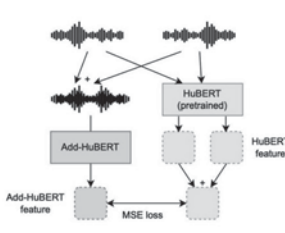


Fig.1: Model structure of Add-HuBERT.

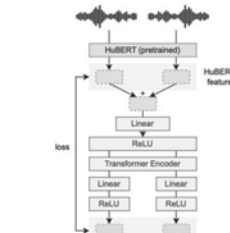


Fig.2: Model structure of Sep-HuBERT.

3-P-12

複数の日本語音声対話コーパスによる Voice Activity Projection の性能比較

Turn-taking prediction model model Performance comparison of Voice
Activity Projection with multiple Japanese spoken dialogue corpus.

○花川主税(千葉工大), 藤江真也(千葉工大/早稲田大)

- ◆複数の日本語コーパスで学習した Voice Activity Projection モデルによるターンテイキング予測性能の比較
- ◆6種類のコーパスをそれぞれ学習データ、評価データに分け、組み合わせを変えて評価
- ◆評価には、ターン保持か交代の2クラスを個別に求めた正解率の平均である balanced Accuracy を用いた
- ◆実験の結果、ターンテイキング予測にはカジュアル/フォーマルといった会話の形式の違いによる影響が大きいことがわかった
 - カジュアルな対話はターンテイキングが多い傾向があり、フォーマルな対話はターンテイキングが少ない傾向
- ◆データ量の大きさが次点のフォーマルな対話コーパスで学習したモデルが平均的に高い評価値を示した
 - データ量が最も大きいコーパスで学習したモデルの評価値が低いのはカジュアルな対話であること多様な会話形式であったため

	CEJC	UADB	PASD	CSJ	RWCP	W	average
CEJC	72.3 (0.35)	73.3 (0.20)	78.0 (0.40)	84.6 (0.50)	78.8 (0.30)	82.4 (0.25)	78.2
UADB	57.2 (0.45)	80.0 (0.55)	55.3 (0.50)	74.1 (0.35)	55.5 (0.50)	51.1 (0.40)	62.2
PASD	59.8 (0.40)	67.7 (0.50)	80.9 (0.30)	70.7 (0.40)	87.5 (0.50)	73.2 (0.55)	73.3
CSJ	58.3 (0.45)	62.2 (0.50)	66.0 (0.65)	92.0 (0.70)	89.9 (0.55)	81.3 (0.40)	74.9
RWCP	57.6 (0.40)	84.4 (0.60)	72.4 (0.25)	82.6 (0.35)	81.2 (0.45)	73.7 (0.30)	75.3
W	54.5 (0.20)	84.4 (0.80)	80.6 (0.15)	80.4 (0.55)	87.8 (0.50)	89.8 (0.15)	79.5
Multi	72.3 (0.35)	90.0 (0.55)	78.9 (0.50)	90.3 (0.50)	93.5 (0.20)	88.3 (0.30)	85.5

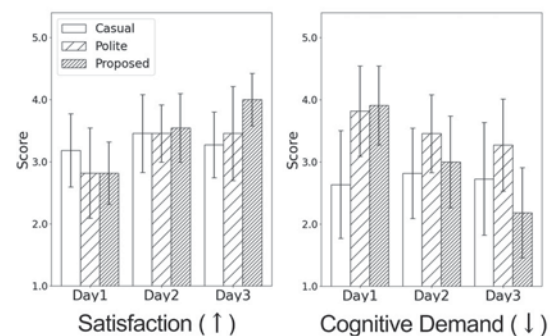
3-P-14

3-P-14 ユーザとの親密さに基づく合成音声制御の 対話実験による検証

Evaluation of Controlling Spoken Response Based on Intimacy with User
through Dialogue Experiments

◎川西翔貴, 伊藤彰則(東北大), 千葉祐弥(NTT), 能勢隆(東北大)

- ◆これまで、我々は与えられた親しみの高低に応じた発話音声を生成できる自然発話音声生成手法を提案した。本稿では、この手法を活用した音声対話システムを構築することで、段階的に親しみを変化させる音声応答制御の効果に対話実験により検証した。
- ◆対話は3日間行い、以下の3条件に被験者を分割した。
 - Casual 条件：常に親しみ高の音声で対話
 - Polite 条件：常に親しみ低の音声で対話
 - Proposed 条件：親しみ低から親しみ高の音声に変化させて対話
- ◆結果、一部の指標で Proposed 条件が他2条件を上回るスコアとなり、段階的に親しみを変化させる音声応答制御の有効性が示された。



3-P-15

3-P-15 会話コーパスに含まれる笑い声を構成する call および吸気音の大規模事前学習モデルによる書き起こし

Transcribing calls and inhalation sounds that constitute laughter in conversational corpora using a large-scale pretrained model

○森 大毅 (宇都宮大)

- ◆笑い声合成では call および吸気音の時間情報付きラベルが必要だが、これを提供しているコーパスは存在しない。アノテーションは極めて大きな労力を必要とするため、笑い声データセット拡充のためには自動アノテーションが望まれる。
- ◆wav2vec 2.0 に基づき笑い声からの phone (call および吸気音) 列の認識モデルを構築した。
- ◆CTC の認識結果の境界は曖昧であるため、境界時刻を同定するための2種類の方法を検討した。
- ◆提案法の評価のため、正解と認識結果とのアラインメントを phone の同一性に依存しない方法で求める方法を考案した。

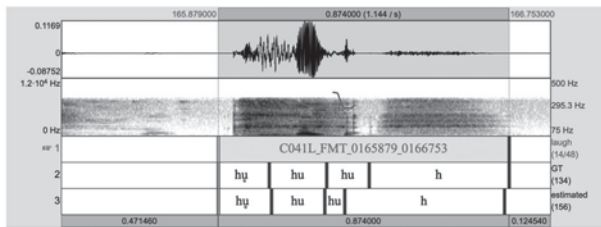


Fig.1: Example of automatic call/inhalation sounds of laughter.

3-P-17

3-P-17 大規模事前学習モデルを用いた 牛の鳴き声の識別

Identification of Cattle Vocalizations Using Large-Scale Pre-trained Models

☆西尾優亜, △坪倉和哉(愛県大・情)

△小島秀介, △佐藤精, △森下忠(愛知県農業総合試験場) 入部百合絵(愛県大・情)

- ◆牛の鳴き声区間検出のため、牛の鳴き声と牛舎内に生じる多種多様な環境音を識別する手法を提案する。
- ◆少数データで識別を行うために、人間の声で学習された大規模事前学習モデルを牛の鳴き声に適用して特徴量を抽出する。さらに事前学習モデルに対し、牛の鳴き声と環境音(エンジン音、金属音、シャベルによる音など)を用いた Fine-Tuning を行い、識別精度の向上を図る。
- ◆ Fine-Tuning を行い、モデルの層毎に抽出した特徴量を用いて SVM により識別を行った結果を Fig1 に示す。事前学習モデルを用いることで高い精度を得ているが、Fine-Tuning により全層で2-5%の精度向上が認められ、1層目では100%の精度を達成した。次に特徴量毎のPCAをFig2に示す。牛舎内の環境音と牛の鳴き声の特徴の違いを捉えられており、Fine-Tuning によって牛の鳴き声と環境音との違いをモデルが適切に学習したためだと考えられる。

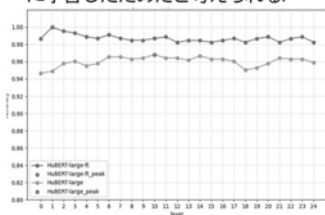


Fig. 1 Identification Results Using Fine-Tuning model

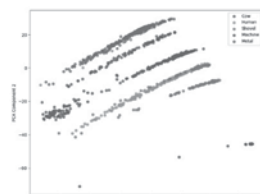


Fig. 2 PCA Plot of Extracted Features

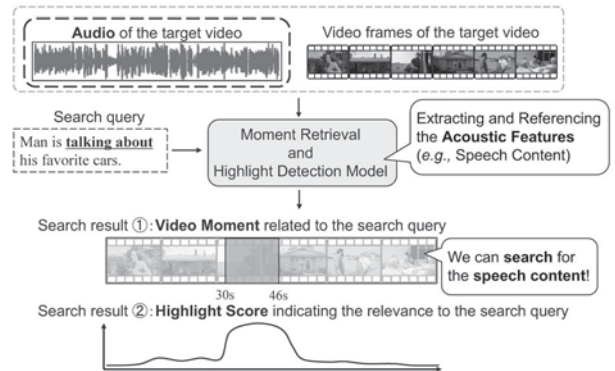
3-P-16

3-P-16 動画内区間検索及びハイライト検出における音響特徴量活用の有効性

The effectiveness of utilizing acoustic features in video moment retrieval and highlight detection.

☆今村剛大¹, △西村太一², 小松達也², 戸田智基¹(¹名大・情報)(²LINE ヤフー)

- ◆動画内区間検索及びハイライト検出とは
動画内区間検索とは：検索文と関連する動画区間を出力
ハイライト検出とは：検索文との関連度を各動画フレーム単位で出力
➢ これらを同時に行うことで、新たな視聴体験を提供可能
- ◆音響特徴量を活用することで、音からしか得られない情報(例：発話内容)を取得することを目指す
- ◆実験的評価により、音響特徴量の活用が有効であること、そしてその有効性は動画の種別に大きく依存することが示唆された



3-P-18

講演取消

3-P-19

3-P-19 変換品質に影響する音響特徴量の DNN 声質変換手法による差異に関する調査

A study on differences in acoustic features affecting conversion quality among DNN-based voice conversion methods

☆松久直生, 坂野秀樹, 旭健作(名城大院・理工学研)

- ◆男女間話者変換の際に声質がどの程度影響を与えるか DNN 声質変換モデル 2 つを用いて調査
- ◆客観評価実験: 男性から女性へ変換する際には, 基本周波数差・ケプストラム距離の影響を強く受ける
 - 基本周波数差 vs 変換後ケプストラム距離および変換前ケプストラム距離 vs 変換後ケプストラム距離において 0.6 以上の相関あり

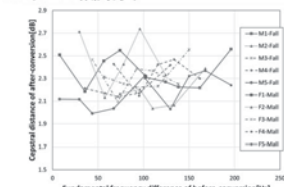
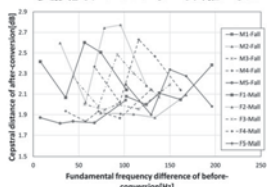


Fig. 1 Relationship between fundamental frequency difference of before-conversion and cepstral distance of after-conversion in StarGANv2-VC

Fig. 2 Relationship between fundamental frequency difference of before-conversion and cepstral distance of after-conversion in RVC

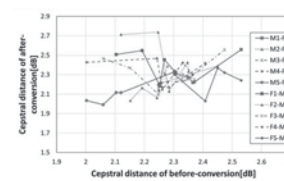
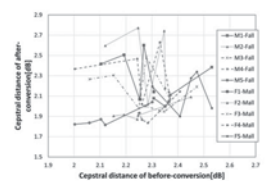


Fig. 3 Relationship between cepstral distance of before-conversion and that of after-conversion in StarGANv2-VC

Fig. 4 Relationship between cepstral distance of before-conversion and that of after-conversion in StarGANv2-VC

3-P-21

3-P-21 感情次元表現に基づいた FaderNet による音声印象変換

Speech impression conversion using FaderNet based on emotional dimensional representation

☆釘本咲, 高橋拓椰, 中庭亘(電通大)

- ◆話者性はそのまま音声の印象のみを変換する印象変換を目的とする
- ◆FaderNetVIC: 声質変換手法である FaderNetVC における話者ラベルを印象ラベルに置き換えることで印象の変換が可能
- ◆Classifier と Discriminator における印象ラベルの識別・推定にベータ分布を用いることで印象変換精度は向上
- ◆印象ラベルが特定の形容詞について定義されているため, 特定の形容詞以外の印象に変換できないという汎用性に面課題
- ◆本稿では印象ラベルに感情次元表現である VAD モデルを取り入れ, 印象ラベルの表現の幅を拡大したモデルを提案

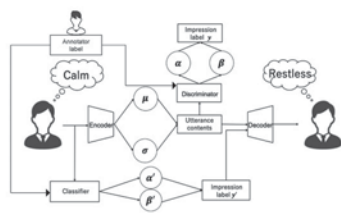


Fig.1.提案手法のモデル構造

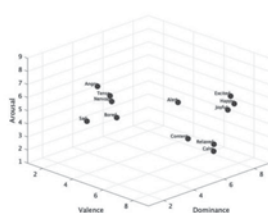


Fig.2.VAD モデル

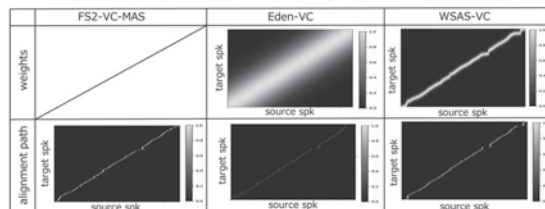
3-P-20

3-P-20 系列変換型声質変換におけるアライメント方式の比較

Comparison of Alignment Methods in Sequence-to-Sequence Voice Conversion.

◎山下陽生(1, 2), 岡本拓磨(2), 高島遼一(1), 大谷大和(2), 滝口哲也(1), 戸田智基(3,2), 河井恒(2)
(1 神戸大学, 2 情報通信研究機構, 3 名古屋大学)

- ◆声質変換モデルにおけるアライメントは Monotonic alignment search (MAS) などの Hard alignment を用いるが, これは安定した学習が出来る一方で Alignment path を間違えたときの修正が効かないなどの柔軟性の面で問題点がある。
- ◆Eden-VC などで行われている Soft alignment は柔軟性の問題を解決しているが, 安定した学習が出来ないという問題がある。
- ◆そこで, 本稿では Eden-VC において Alignment 対角化のために用いられている重みの代わりに, MAS から得られた Hard alignment path から重みを生成して使用する WSAS-VC を提案する。



Learned weights and alignment paths.

3-P-22

3-P-22 音声合成における音響特徴エンコーダの重要性の評価

The Evaluation of the Importance of Acoustic Feature Encoders in Speech Synthesis

◎藁谷友希(埼玉大・工), 杉浦陽介, 島村徹也

- ◆概要
音声合成技術は AI の進展に伴い発展しており, 本研究は構音障害を持つ人々の支援を目的とする。FreeVC モデルを基盤に音響特徴量抽出器を Wav2Vec2.0 に置き換え, 日本語音声の明瞭度向上を目指す。
- ◆研究内容
・ FreeVC の概要(Fig 1): テキスト不要で話者の声質変換が可能な音声合成技術。音声波形から言語特徴量と話者特徴量を抽出し, 別の話者の音声に変換可能。

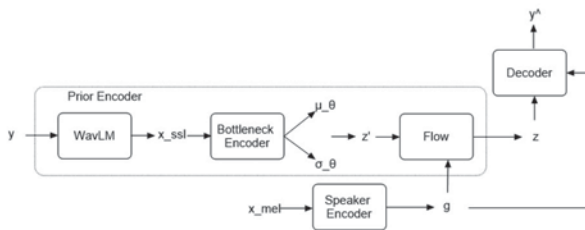


Fig 1. The Architecture of FreeVC

- ・ 問題点: FreeVC で使用される音響特徴量抽出器 WavLM は日本語に最適化されていないため, 日本語音声の生成精度が低い。
- ・ 改良案: 日本語最適化された Wav2Vec2.0 の使用により, 音響特徴量抽出の精度向上を実現し, FreeVC の性能改善を目指す。

3-P-23

3-P-23 健常者の子音との組み合わせを用いた 構音障がい者音声合成

Speech synthesis for a person with dysarthria
using combinations with consonants of physically unimpaired speaker

☆長久保 諒¹, 高島 遼一^{1,2}, 佐々木 千穂³, 滝口 哲也¹

(¹神戸大学, ²JST さきがけ, ³熊本保険科学大学)

- ◆本研究では、**構音障がい者の声質**でかつ**明瞭なテキスト音声合成システム**の作成を目的として、**健常者の子音との組み合わせ**を用いる手法を提案する。
- ◆**構音障がい**とは、言語の理解に問題は見られず発話内容は明確であるが、発声発語器官及びその動作に何らかの異常が見られるために**不明瞭な発話**になってしまう障がいであり、その**音声**を学習した**合成音声も同様に不明瞭なものとなる**。
- ◆調査の結果、**構音障がい者の発話**においては**母音と比較して子音の発音が著しく不明瞭な傾向**にある事がわかった。
- ◆提案手法では、音声の話者性は子音よりも母音に強く現れると仮定し、**不明瞭な子音の音声特徴を健常者の音声特徴で代替**する。

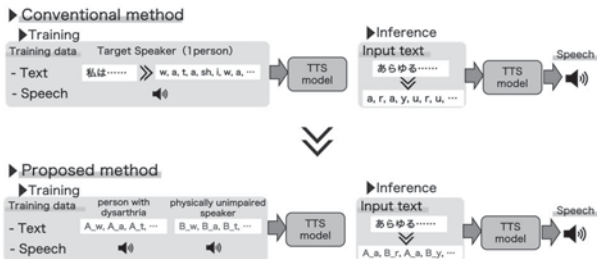


Fig.1 Overview of the proposed method

3-P-25

3-P-25 咬み合わせ治療による発声改善における 母音特徴の分析

An effect of dental edges to the speech production
— Analysis of vowel features —

☆武田陸歩, 勝木愛夏, 大川茂樹(千葉工大), 安藤正之(安藤歯科クリニック)

- ◆**咬み合わせの治療による発声への影響**
 - 咬合不良は、舌の動きを妨げて発声に悪影響を及ぼす
 - 咬み合わせ治療を受けた152名の患者の母音特徴を分析
- ◆**フォルマント分析**
 - F1-F2平面において5母音を結んで形成される面積の変化を調査(Fig.1)
 - 2次元ガウス分布と標準偏差から治療による母音のばらつきを解析
 - 各母音のF1-F2平面における移動方向を割合で表示
- ◆**結果と考察**
患者の7割で面積が増加、標準偏差の減少、母音同士が離れる傾向
 - t検定を実施した結果、5母音フォルマント面積の変化と /a/, /e/ のF1変化、および /i/, /o/ のF2変化において有意差が認められた
 - 治療により舌や唇が正しい位置で発声できるようになり、母音発声の安定性と明瞭性が向上

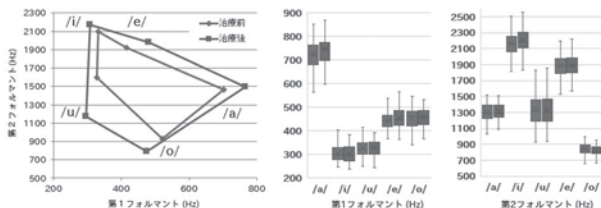


Fig.1: 治療前後のフォルマント変化

(左: 男性患者N.Y.の5母音フォルマント面積, 右: 男性73名のF1, F2変化)

3-P-24

3-P-24 音声のポジティブな感性印象と 音響特徴量との関係

The relationship between positive emotional impressions of voice and
acoustic features

◎林大輔(日本たばこ産業), 森勢将雅(明治大学)

- ◆**音声**を聞いた時に抱く**ポジティブな感性印象**について、その幅広く多面的な要素を適切に表現・評価できる質問紙を作成するため、これまでに調査ならびに実験を行ってきた(2024年秋の音響学会で発表)。
- ◆本研究では、音声の**ポジティブな感性印象**に関する理解を深めるため、先行研究で作成した質問紙の妥当性について検証しつつ、それらの印象と音声の物理的な音響特徴量との関係について検討を行った。
- ◆まず、音声の主観評価実験を行い、多変量解析を行った結果、先行研究で作成した音声の**ポジティブな感性印象**に関する質問紙について、一定程度の妥当性が示された。
- ◆また、音声の好ましさに関する質問紙との関係について分析を行った結果、音声の**ポジティブな感性印象**に関する質問紙の各因子(あたたかさ、明瞭さ、かわいさ、独特さ)と全体としての好ましさととの関係について示唆が得られた。
- ◆続けて、音声の工学分析を行い、声質に関する様々な音響特徴量を算出した。
- ◆音声の主観的な印象と物理的な音響特徴量との関係について重回帰分析を用いて検討した結果、あたたかさやかわいさ、好ましさについて先行研究と整合的な結果が示された一方で、全体に決定係数は十分に高いとは言えず、今後より幅広い音声や音響特徴量を用いた検討を行っていく必要が示唆された。

3-P-26

3-P-26 TV-CAR 分析に基づく声帯音源波モデル パラメータの推定に関する検討

On glottal model parameter estimation based on the TV-CAR speech
analysis

◎舟木慶一(琉球大学)

音声から声帯音源モデルパラメータを推定することは音声研究者の夢であり、声道フィルタ係数とともに同時推定する ARX 分析が1980年代から研究されている。近年、ARX-LF 分析なるものが提案され報告されている。一方、我々は、解析信号に対する TV-CAR(時変複素 AR) 音声分析を用いた ARX-LF 分析の構築を行っている。GCI/GOI 推定において LPC 残差に基づく SEDREAMS 法がよく使われているが、既に TV-CAR 残差に基づく SEDREAMS 法を提案し、声帯音源波の推定方式としてよく用いられている LPC 分析に基づく音源逆フィルタリング(GIF)である IAIF(Iterative Adaptive Inverse Filtering)法を TV-CAR 分析を用いる IAIF 方式に拡張した方式を既に提案した。ARX-LF 分析では予備選択として、IAIF により推定された声帯音源波から、動的プログラミングを用いてフレーム間で最適化された LF モデルのパラメータを推定する DyProg-LF 方式が用いられている。本稿では、解析信号から TVCAR 分析に基づく IAIF により推定された声帯音源波から、DyProg-LF 方式で LF モデルパラメータを推定する方式を構築したので報告する。

3-P-27

3-P-27 変分オートエンコーダに基づく音声-テキストアライメントの汎用性評価

Evaluation of the versatility of speech-text alignment based on variational autoencoder

○郡山知樹 (サイバエージェント)

▶ 背景

▶ VAEに基づくアライメントの評価はこれまで限定的

▶ 主要な結果

- ▶ 英語のTIMITやBuckeyeコーパスでも有効
- ▶ SSL特徴量の選択はアライメント精度に大きく影響しない
- ▶ ドメイン依存性をLoRAによるモデル適応で解消可能
- ▶ 音素ではなくテキスト入力を用いてもアライメントが可能

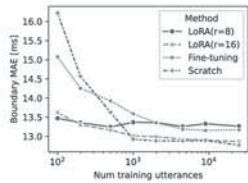


Fig 1: Results of LoRA model adaptation.

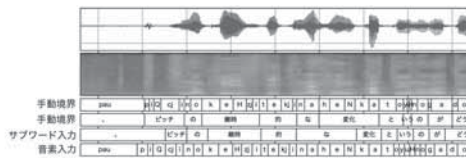


Fig 2: Alignment result using the subword-based input.

3-P-29

3-P-29 学習者本人の自己聴取音の声質で合成されたモデル音声をを用いた発音学習とその効果

Pronunciation Training Using Model Speech Synthesized with Individual Learners' Own Voice Quality

☆山中涼雅, 大佐健人(東大・工), 藤原朱里, 耿浩彰,

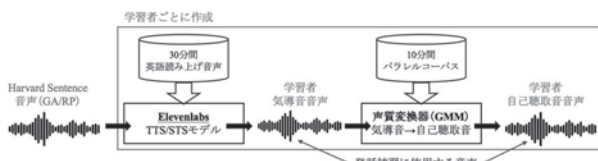
齋藤大輔, 峯松信明(東大院・工学系), 井上雄介(Carriage Inc.)

自分にとっての自分の声(自己聴取音)は一般に録音音声とは異なる

- ・録音音声 = 気導音(空気振動を介して聞く音のみ)
- ・自己聴取音 = 気導音 + 骨導音(体内を通して伝搬する振動)



自己聴取音を含む様々な声質のモデル音源を作成!



自分の声に似ている音声の方が発音を真似しやすい??

用いる音声の声質を変えて発音学習の効果を検証しました!

3-P-28

3-P-28 聴取者の母語による母語・非母語聴取時の話者類似度の違いに関する調査

An investigation of the differences of speaker similarity by the listener's native language in listening native/non-native language

☆両角祐二, 旭健作, 川澄未来子, 坂野秀樹(名城大院・理工学研)

◆日本語母語聴取者・タイ語母語聴取者の話者類似度の違いを調査

▶ 聴取言語: 日本語, タイ語, 英語

◆主観評価実験: タイ語母語聴取者が女性らしい声を女性と認識

◆メルスペクトログラムの比較: 音素の立ち上がりや立ち下がり部分の第二フォルマントの時間的変化が少ない時, 女性らしい声と認識

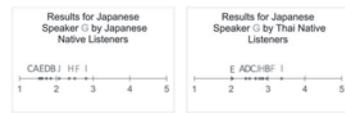


Fig. 1 Japanese comparison results

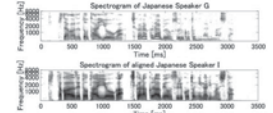


Fig. 4 Spectrograms of Japanese speakers

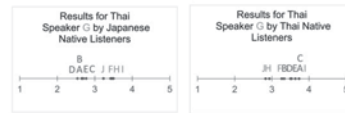


Fig. 2 Thai comparison results

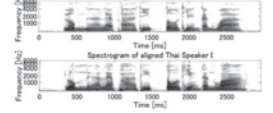


Fig. 5 Spectrograms of Thai speakers

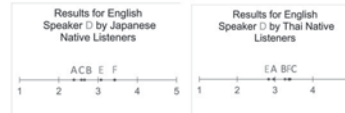


Fig. 3 English comparison results

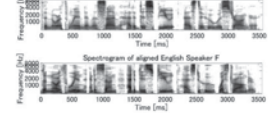


Fig. 6 Spectrograms of English speakers

3-P-30

3-P-30 基本周波数に含まれる微細変動を除去した歌唱音声合成音の作成と自然性の評価

A study on evaluation of naturalness by singing synthetic sounds removing fine fluctuation of fundamental frequency in singing voice.

◎鈴木千文, 坂野秀樹, 旭健作(名城大)

◆本研究ではこれまでノンビブラート音声の基本周波数に含まれる微細変動を観察した

◆今回は基本周波数の微細変動の有無が聴感上与える影響を調査する

▶ 合成音の作成

◇ 基本周波数から回帰直線の傾きである微分信号を算出

◇ 基本周波数の微分信号からハイパスフィルタで抽出した微細変動(Fig.1 左)とローパスフィルタで抽出した大局的変動(Fig.1 右)を抽出

▶ Fig.1 左より細かい変動を持つ成分が抽出できている, Fig.1 右より緩やかな変動を持つ成分が抽出できている

◆作成した合成音と元のノンビブラート音声を聞き比べた結果

▶ 基本周波数の微細変動の有無による影響はほとんどない

▶ 基本周波数の微細変動の有無より大局的変動の有無の方が影響が大きい可能性がある

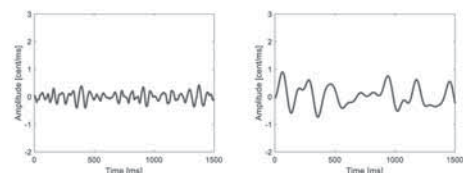


Fig1. Fine fluctuation of Fundamental frequency (Left). Comprehensive fluctuation of Fundamental frequency (Right).

3-P-31

3-P-31 歌声合成を用いた 斉唱の自然性に関する要因調査

Investigating Factors Related to the Naturalness of Synthesized Unison Singing

☆西澤佳飛¹, 山本龍一^{1,2}, Wen-Chin Huang¹, 戸田智基¹
(¹名古屋大学, ²LINEヤフー)

- ◆音響特徴量(タイミング, f_0 , スペクトル特徴)の揺らぎが斉唱の自然性と歌唱人数の知覚に与える影響を調査.
- ◆揺らぎあり歌唱音源の作成には歌声生成モデルNNSVSを使用.
- ◆聴取実験の結果, (1)タイミングと f_0 の揺らぎが斉唱の自然性において重要, (2)複数人歌手の生声斉唱と合成ソロサンプルを使用した斉唱の知覚的な挙動が類似していることが分かった.

	タイミング	f_0	スペクトル特徴	自然性 MOS†
(a)				1.56 ± 0.15
(b)	✓			** 3.94 ± 0.11
(c)		✓		** 3.80 ± 0.11
(d)			✓	1.66 ± 0.16
(e)	✓	✓		* 4.10 ± 0.10
(f)		✓	✓	* 4.04 ± 0.11
(g)	✓		✓	* 4.06 ± 0.11
(h)	✓	✓	✓	* 4.14 ± 0.11

Table1 Naturalness mean opinion score (MOS) results by applying fluctuation to different acoustic features in unison singing synthesis.

デモサイト(聴取実験に使用した斉唱音源)

https://kaito-nishizawa.github.io/presen_demo.github.io/



3-Q-1

3-Q-1 汎用言語音響表現 M2D-CLAP

M2D meets CLAP: A General-purpose Audio-Language Representation

○仁泉大輔, 竹内大起, 大石康智, 原田登, 安田昌弘 (NTT),
椿竣介, 井本桂右 (同志社大学)

- ◆自己教師あり学習 M2D と対照言語音響事前学習(CLAP)を組み合わせる新たな表現学習 M2D-CLAP を提案する.
- ◆従来の CLAP とは異なり, M2D-CLAP は高い性能の音響信号表現とゼロショット分類可能な CLAP 表現を同時に実現する.
- ◆M2D と CLAP のマルチタスク学習は, CLAP 表現の学習に加え, 従来の M2D により学習する表現の性能を更に向上させる.

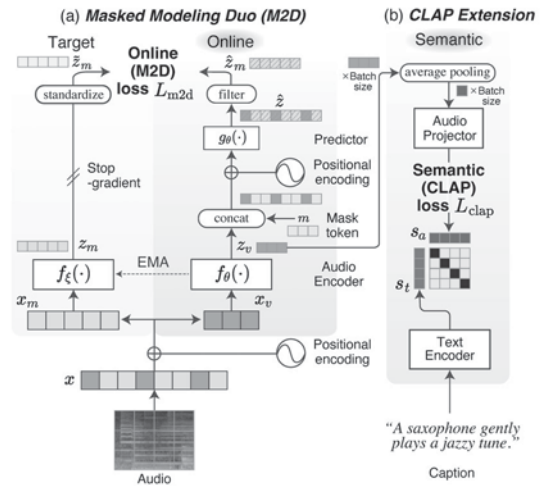


Fig. 1 M2D-CLAP pre-training flow.

3-Q-2

3-Q-2 Classifier chains により タグ間の依存を考慮した音楽タグづけ

Music tagging considering interplay between tags by classifier chains

◎蓮実拓也, 小松達也, 藤田雄介 (LINE ヤフー)

- ◆音楽タグづけは, 音楽からジャンルやムード, 楽器といった属性を推定するタスクであり, 効率的な音楽検索のために重要な技術である.
- ◆従来の音楽タグづけでは, 各音楽タグが条件付き独立であると仮定してデコーダをモデル化する. この手法は, 楽器からジャンルが推定されうるといった, タグのカテゴリー間の関係を明示的に利用できない.
- ◆本稿では, classifier chains を音楽タグづけ問題に適用し, 音楽タグのカテゴリーに基づいたグループ単位での chain である classifier group chains を用いた音楽タグづけ手法を提案する.
- ◆MTG-Jamendo データセットを用いた音楽タグづけ実験により, 音楽の異なる属性間の関係を考慮することの重要性を示す.

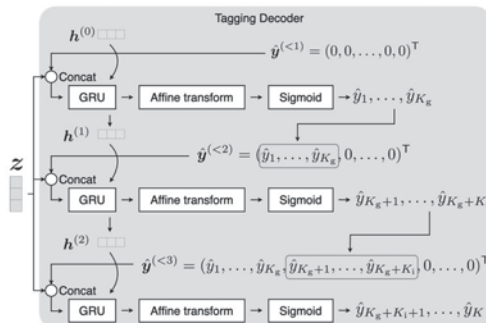


Fig.1: Overview of proposed music tagging with classifier group chains "genre" → "instrument" → "mood/theme".

3-Q-3

3-Q-3 環境音合成における客観評価指標 CLAPScore の時系列評価性能の分析

Analysis of time series evaluation performance of CLAPScore for environmental sound synthesis

☆高野 大成 (東大・工), 岡本 悠希, 齋藤 佑樹 (東大院・情報理工)

- ◆テキストと音の類似度を測る評価指標 CLAPScore を分析
 - テキストから「犬の鳴き声」などの身の回りのあらゆる音を合成する環境音合成技術 (text-to-audio) の評価に利用
- ◆CLAPScore がテキストと音の時系列関係を評価困難であると確認
 - テキストと音の時系列を入れ替えた場合の CLAPScore_{swap} を計算
→ 時系列の入れ替えを行わない場合 (CLAPScore_{original}) と比べて直に変化なし

Table 1: Result of CLAPScore when swapped text-audio temporal alignments

Category	CLAPScore _{original} ↑	CLAPScore _{swap} ↑	変化率 r
Dog - Cat	0.23320158	0.23408227	1.00
Dog - Engine	0.17905925	0.17903274	1.00
Dog - Water drop	0.40429428	0.4089144	1.01

- ◆CLAPScore と主観評価の相関が低いことを確認 (Fig. 1)

- CLAPScore は主観評価を代用可能な客観評価指標ではない

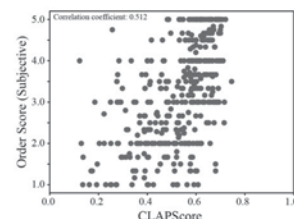


Fig.1: Scatter plot of order score (subjective score) and CLAPScore

3-Q-4

3-Q-4 ホログラフィック縮退表現による音響フィンガープリントのメモリ削減

Memory-efficient Audio Fingerprinting with Holographic Reduced Reps.

○藤田雄介, 小松達也(LINE ヤフー)

- ◆ニューラルネットワークに基づく音響フィンガープリント (NAFP) は、ノイズへの頑健性と高い時間分解能(短い入力音で楽曲が特定できる能力)が得られる一方、大量のメモリが必要となる。
- ◆メモリ削減のため、ホログラフィック縮退表現 (HRR)を用いて複数のNAFPを集約する仕組みを提案する。HRRは、巡回畳み込みと加算を用いて、複数のNAFPを同じ次元数のベクトル(バケット)に集約する。集約されたNAFPの存在は、単純な内積で判定できる。
- ◆FMA データセットを用いて音楽のフィンガープリントの検索実験を行い、HRRの有効性を示す。

Table 1 Top-1 hit rate (%) with different aggregation methods. The fingerprint dimension is 512. M is the block size for aggregation.

Method	M	Query length (s)				
		1	2	3	5	10
No-aggregation	1	71.1	90.4	95.1	97.9	99.4
Summation	2	31.2	74.2	84.9	92.7	96.4
Decimation	2	39.0	74.0	86.0	93.4	95.9
HRR (proposed)	2	58.8	83.8	90.9	95.1	97.8
Summation	4	3.0	6.6	39.4	44.3	89.8
Decimation	4	19.3	45.8	37.7	71.3	92.8
HRR (proposed)	4	31.0	58.3	45.0	79.0	96.0

3-Q-6

3-Q-6 位相干渉に基づく音響測距法を用いた合成開口レーダーによるイメージング~素子間隔による空間エイリアシング軽減の試み~

Sound field imaging with synthetic aperture radar using acoustic distance measurement based on phase interference. ~ Consideration of spatial aliasing reduction by microphone element interval. ~

☆田邊広大, 上保徹志(近畿大), 中山雅人(阪産大), 中迫昇(近畿大)

- ◆本稿では、位相干渉に基づく音響測距(Acoustic Distance Measurement: ADM)法に着目し、イメージングを行う。シミュレーションを通じて、対象物の位置・概形を推定する際のマイクロホン素子間隔による空間エイリアシング軽減の試みる。
- ◆1chの音響測距を用いて、合成開口レーダーの原理に基づくイメージング法に対して、シミュレーションを行う。スピーカーを1つ(位置は固定)、合成開口長を約1.0m、素子間隔を0.33mまたは0.03mとして、直線的に走査を行う。それぞれのマイク素子位置で得た対象物までの距離に関する振幅と位相の情報を用いて、イメージングを行う。
- ◆Fig.1, 2 それぞれ素子間隔を0.33m, 0.03mの時に得られた位相情報と振幅情報を用いてイメージングを行った結果を示す。今回想定していた位置と推定結果は一致しており、素子間隔が狭くなると空間エイリアシングが軽減されることが確認できた。

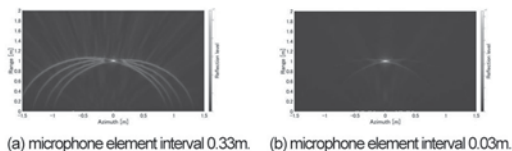


Fig.1: Imaging with use of ADM method.

3-Q-5

3-Q-5 複数ビームフォーマと空間・時間アテンション機構を用いた音響シーン識別の検討

A study on acoustic scene classification using multiple beamformers and spatial-temporal attention mechanism

☆松倉 弘樹, 山田 武志(筑波大)

- ◆個々の音響シーンにおいて目的音(識別に有用な音)がどの音であり、どの方向から到来するのかが自明ではないため、ビームフォーマを前処理として単純に適用することは難しい。
- ◆この問題に対して我々は、複数ビームフォーマと空間アテンション機構を用いて目的音を自動的に発見する手法を提案したが、目的音が時間的に断続・局在している場合の処理に課題があった。
- ◆本稿では、従来手法に時間アテンション機構を適用して各時間ブロックを特徴量レベルで統合する手法を提案する。
- ◆劣悪な雑音環境を模したデータセットを用いて実験を行った結果、提案手法により識別精度が向上することを確認した。

Table1: Experimental results (F-Score [%])

RT [s]	method	5 blocks	10 blocks
0.00	Single channel	49.99	42.66
	Conventional	85.53	77.06
	Proposed	95.57	95.14
0.31	Single channel	43.91	37.18
	Conventional	72.34	72.08
	Proposed	82.21	79.29

3-Q-7

3-Q-7 畳み込みニューラルネットワーク(CNN)を用いた超音波エコー画像の分類によるしじみの良否判別

Classification of ultrasonic echo images using Convolutional Neural Network(CNN) to discriminate good or bad freshwater clams

☆古川尚幸(近畿大院), 篠原寿広, 上保徹志, 中迫昇(近畿大)

- ◆しじみは出荷前に漁師の手作業によって出荷可能な貝(良しじみ)か、出荷不可能な貝(不良しじみ)かに判別されている。しかし、この作業は集中力を必要とする長時間の労働である。
- ◆本研究では、とくに良しじみと判別されたしじみ貝の中に不良しじみが含まれないよう高精度に判別できる手法について検討する。具体的には、CNNを用いて超音波エコー画像の分類を行うことでしじみの良否を判別する。
- ◆Fig.1(a), (b)はそれぞれ良しじみ, 不良しじみの超音波エコー画像の一例である。本研究では、Fig.2のCNNのネットワーク構造を用いてこれらの超音波エコー画像の分類を行うことで、しじみの高精度な判別が可能であることを示した。

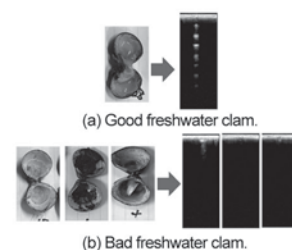


Fig.1: Sample ultrasonic echo images of freshwater clam.

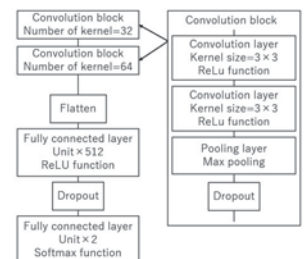


Fig.2: Network structure of CNN.

3-Q-8

3-Q-8 雑音下における空気伝導音への RLS 法の適用
～骨伝導音を入力とした雑音除去と音質改善～

Application of the RLS method to air-conduction sound in noisy environments.

～ Bone-conduction sound as input for noise reduction and sound quality improvement ～

☆竹内彩華 (近畿大院), 上保徹志 (近畿大), 中山雅人 (阪産大), 折本寿子, 生田頭 (県立広島大学), 中迫昇 (近畿大)

- ◆情報機器を安全に操作する手段として、音声認識に代表される音声の利用への関心が高まっている。音声の認識率は、静寂環境ではほぼ完璧に近いが、雑音が存在すると認識率が著しく劣化することが知られている。
- ◆そこで雑音対策として、原理的に周囲の騒音が入り難い骨伝導マイクロホンの利用が考えられる。骨伝導音は空気伝導音に比べ音質が劣っていることから、先行研究では、深層学習を用いた音質改善システムの提案やコードブックマッピングによる骨伝導の音質改善が提案されている。
- ◆本研究では、前回の報告に引き続き、空気伝導音の雑音低減の手法を、骨伝導音を入力、雑音下の空気伝導音を出力として、適応信号処理の基礎的な観点(具体的には、RLS(Recursive Least Squares)法)から検討し、同時に骨伝導音の音質改善を試みる。
- ◆Fig.1.2は空気伝導音と骨伝導音の時間波形、Fig.3はRLS法を適用した出力信号を示す。Fig.3より雑音が低減されたことが確認できる。

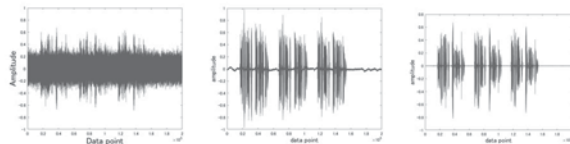


Fig.1: Air conduction sound. Fig.2: Bone conduction sound. Fig.3: Output signal.

3-Q-10

3-Q-10 SPレコードの音溝写真からの音声再生

Audio reproduction from sound groove photographs of SP records

☆松田龍介, 前山有紗 (早大理工), 来住伸子(津田塾大学), 及川靖広 (早大理工)

- ◆SPレコードは貴重な文化遺産であり、非接触再生技術が求められている。本研究では、音溝画像から音声を再構成する方法を提案する。
- ◆撮影した音溝画像を極座標変換とエッジ処理で解析し、波形データの生成を実現した。
- ◆撮影結果からある程度の音は確認できたが、ノイズや高周波成分の減衰が課題として残った。精度向上によりさらなる改善が期待される。

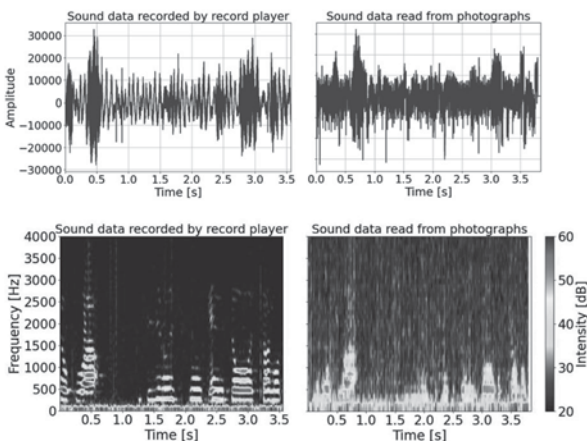


Fig.1: Waveforms and spectra of sound data recorded from records (left) and read from photographs (right)

3-Q-9

3-Q-9 先行音効果を用いた避難誘導システムにおけるスピーカの設置台数の影響

Effect of number of loudspeakers in an evacuation guidance system using the precedence effect

☆遠藤優太, 青木茂明, 島内末廣(金沢工大)

- ◆建物内で火災などの災害が発生した場合、人々を安全な場所に避難する方法として、先行音効果を利用した避難誘導システムがある。
- ◆本研究では、スピーカ16個を避難経路の左右両側に等間隔にかつ並列に配置した場合の最適な遅延時間および音声長による音像定位感への影響について検討を、音像定位の距離(定位距離)、標準偏差、定位割合(前方)、定位割合(前方分離)の4つを基に行った。
- ◆聴取に用いたのは長い音声長(8秒)、中程度の音声長(4秒)、短い音声長(2秒)の3種類である。また、スピーカ間の遅延時間の差を1~50msの12種類とした。
- ◆聴取試験結果 Fig.1よりスピーカ16個では5ms、先行研究よりスピーカ8個では25msが最適な遅延時間であるといえる。
- ◆音声長ごとに比較すると、Fig.2よりスピーカ16個では長い音声長、先行研究よりスピーカ8個では短い音声長が望ましいといえる。
- ◆以上より、スピーカの設置台数を増やすと、最適な遅延時間は短くなり、望ましい音声長は長くなる傾向にあると考えられる。

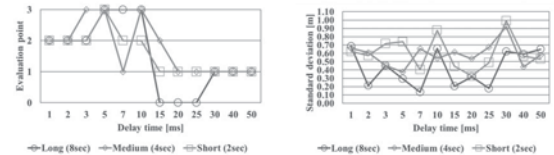


Fig.1 Evaluation point of sound localization Fig.2 Standard deviation of sound localization

3-Q-11

3-Q-11 MIDS法を用いたサーモホンアレイの紙への実装

Thermophone array mounted on paper using Multi-Input Delta Sigma Modulator

☆田中瑞樹, 武岡成人(静理工大)

- ◆提案する多入力型 $\Delta\Sigma$ 変調(MIDS)を用いたマトリクス駆動法とサーモホンを組み合わせることで、“共通配線とサーモホン素材のみ”でありながら個別駆動できるスピーカアレイを提案している
- ◆配線部分に導電塗料、サーモホン素子にCNT分散液を用いて、紙に塗布するといった印刷工程のみでのスピーカアレイ作成を試みた
- ◆本法の応用により、プリンタを用いた微細な超音波アレイなど形状の制約を緩和したスピーカアレイの実現が期待される

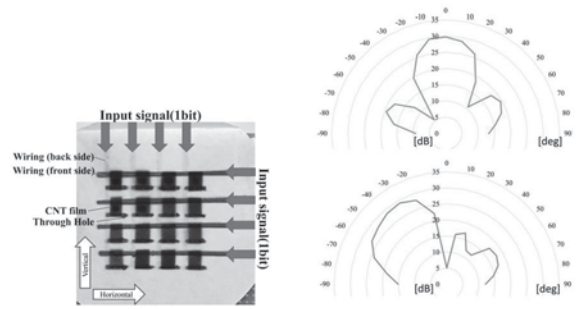


Fig.1 Printed thermophone array Fig.2 Measurement results

3-Q-12

3-Q-12 連続測定法による曲げ波スピーカの指向周波数特性測定

Bending Wave Loudspeaker Characteristics Measurement by using Continuous Measurement Method

○河原一彦, 岩田水晶, △小林由佳(九州大・芸工), 立花杜斗, 前田和昭(TOA)

- ◆ 著者の河原は、聴感上の“いよゆる放射の拡散的性質”を重視した曲げ波スピーカ(Lab-made BWL)の設計法を提案している。
- ◆ また著者らにより、スピーカの指向特性を高速に測定する、連続測定法が提案されている。
- ◆ 連続測定法により Lab-made BWL の方位角による周波数特性の微細な変化を測定できることができた。
- ◆ その際の連続測定法の測定パラメータと、方向別の振幅周波数特性の概要を報告する。

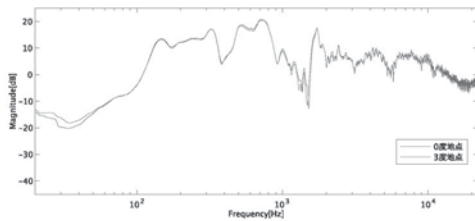


Fig.1: Frequency Response of the Lab-made Bending Wave Loudspeaker (0° and 3° direction).

3-Q-14

3-Q-14

高速度カメラとカラー構造化照明を用いた面振動の計測

Measurement of surface vibration using high-speed camera and color structured light.

☆ 安藤元暉, 新美智也, 矢田部浩平(農工大)

前回 構造化照明を投影し高速度カメラで撮影
▶ 膜の振動に影響を与えず振動を解析

課題 カラーフィルタの特性により撮影後の色特性やノイズを補正する必要がある

手法 色成分のリークを混合行列を用いて、ノイズをフィルタを用いて補正

結果 RGB各成分を均一に使用でき、より滑らかな振動の可視化を実現

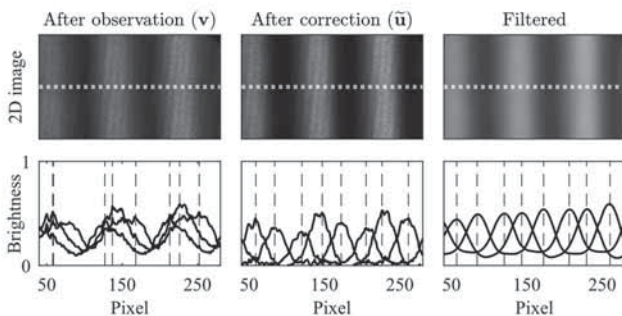


Fig. 1 The observed image (left), image with color leakage corrected (middle), and image with color leakage and noise corrected (right). The bottom row indicates the corresponding brightness of images above.

3-Q-13

3-Q-13 AR技術を用いた音像定位評価システムの開発

Development of a sound image localization evaluation system using AR technology

○藤井優美, 村山好孝(シーイヤー), 渡邊祐子(東京電大・デザ工)

- ◆ Meta Quest3 を用いた AR 音像定位評価システムを開発し、対象物の位置を正確に回答できるか、評価した。
- A: 従来からのスケール読み上げの方法 (STD)
- B: AR 音像定位評価システム (AR)
- ◆ 結果: 本システム (AR) は回答のばらつきが少ないことが分かった。

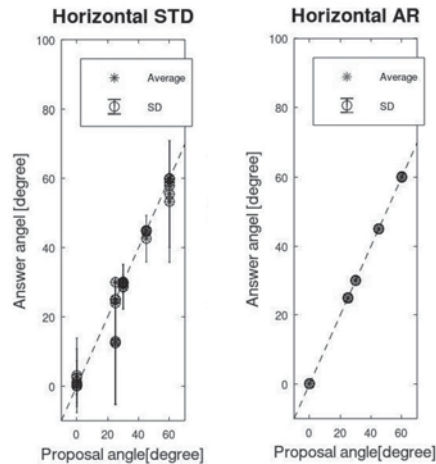


Fig.1: Relationship with presentation direction in horizontal angle. (*: Average of 10 subjects Error bar: Standard deviation)

3-Q-15

3-Q-15

デフォーカスを用いた弦振動変位測定の実験的検証

Experimental validation of string vibration displacement measurement by defocusing

☆ 新美智也, 安藤元暉, 矢田部浩平(農工大)

前回 デフォーカスを利用した高速度カメラによるエレキギター弦の二次元振動の測定手法を提案

- ピクセルの変位と実際の移動量の対応関係が不明瞭
- カメラと弦の角度(仰角)や弦の種類などの条件が性能に与える影響の調査が必要

今回 弦上の測定点の検出性能や信頼性を実験的に検証 (Fig. 1 はプレーン弦, 仰角 5° の場合)

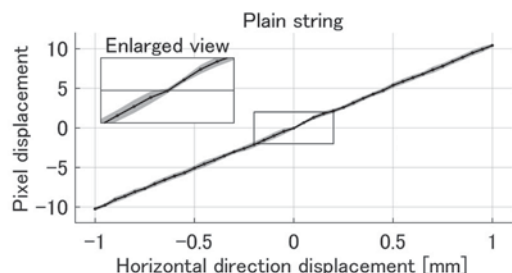


Fig. 1 Mean value and 95% confidence interval of pixel displacement corresponding to physical displacement

3-Q-16

3-Q-16

Neural ODEを用いたギター弦振動のモデル化 Modeling of guitar string vibration using Neural ODE

☆ 犬飼朋樹, 新美智也, 矢田部浩平 (農工大)

背景 エレキギターの弦振動のモデル化には、非線形性に対応できるモデルの構築が必要

今回 複雑な力学系をモデル化することが可能な Neural ODE を用いた弦振動の学習

結果 ある程度の精度で弦振動のモデル化を実現
▶ 非線形性のモデル化への寄与が示唆された

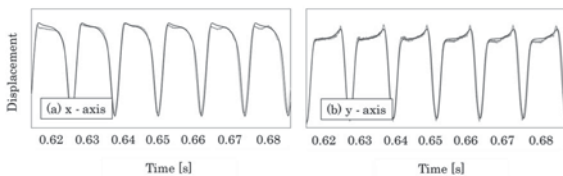


Fig. 1 An example of a guitar string vibration in the direction of x-axis (a) and y-axis (b). The true signal and predicted signal are illustrated with the light and dark curve, respectively.

3-Q-18

複数マイクロホンアレイで推定した 空間エネルギー分布に基づく 音響イベント検出と音源定位

Acoustic event detection and sound source localization based on spatial energy distribution estimation using multiple microphone arrays

☆ 上林大輝, 松本真太郎, 中山雅人, 高橋徹 (阪大院・工学研)

【背景】

- 室内の対話活動可視化システム(Fig. 1)の開発を目指している。
- システムの一部である音源定位課題に取り組んできた。
 - > 広範囲に適用可能な複数の非同相分散マイクロホンアレイで音響イベント検出と音源定位する手法を提案
 - > 空間の音圧エネルギー分布に基づく音響イベント検出と音源定位
 - > 空間の音圧エネルギー分布の推定方法の改良が鍵

【目的】 エネルギー分布推定に使用するマイクロホンアレイ選択

- Before 近傍の3マイクロホンアレイ
- After 他にも利用可能なマイクロホンアレイの存在
- <<最適なマイクロホンアレイの選択方法を調査>>

【選択方法】

- [Neighbor 基準] 音圧エネルギーを推定する点に近い n 個のマイクロホンアレイを使う方法。
- [Energy 基準] 観測エネルギーが大きい n 個のマイクロホンアレイを使う方法。

【結果】 同一許容定位誤差に対する音源検出数を比較(Fig. 2)。許容誤差 0.5m 未満 Neighbor 基準の $n=3, 4$ の手法が有効。許容誤差 0.5m 以上 Energy 基準の $n=3, 4$ の手法が有効。

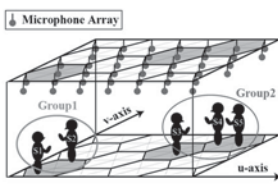


Fig 1 : dialogue monitoring system.

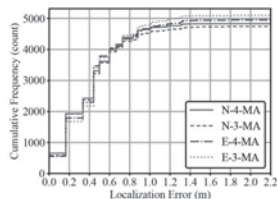


Fig 2 : Comparison between neighbor criterion and energy criterion.

3-Q-17

球状マイクロホンアレイにより取得した 3次元音源放射特性のグラフ信号処理を 用いた空間アップサンプリング

Spatial Upsampling of 3D Sound Source Radiation Characteristics Captured by a Spherical Microphone Array Using Graph Signal Processing

◎ 大久保翔太, 堀内俊浩, △渡邊良亮 (KDDI 総合研)

◆音源放射特性において、半球面上の限られた位置で収録された音響信号から、収録していない位置の音響信号を補間することを目的とし、グラフ信号処理を用いた補間手法を導入した。

◆振幅値に基づくグラフ構築手法を提案し、相関係数を用いて補間した音響信号の再現精度を評価した。

◆グラフ信号処理において汎用的に使われるk近傍法によるグラフ構築手法と比較し、再現精度改善が確認された。

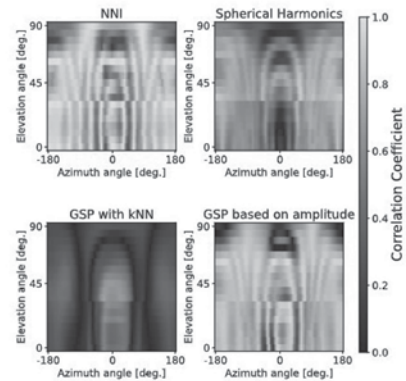


Fig.1: Correlation coefficient values of interpolation methods

3-Q-19

3-Q-19

時間周波数領域での厳密な畳み込みを 応用するための斜交射影

Oblique projection for error-free convolution in time-frequency domain

◎ 松山直哉, 矢田部浩平 (農工大)

- 時間周波数領域での近似畳み込みは、インパルス応答が長いほど精度が悪くなってしまふ
- これまでに、時間周波数領域での厳密な畳み込みを導出した
- 厳密な計算に必要なインパルス応答の時間周波数表現を得るため、順変換と逆変換を定式化した

音響信号処理に時間周波数領域での厳密な畳み込みを用いるための直交射影と斜交射影を導出

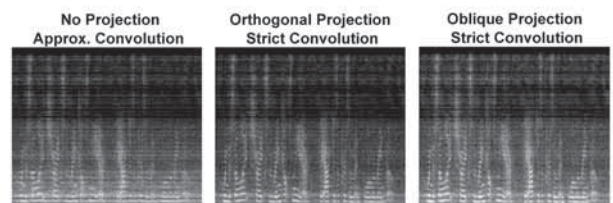


Fig. 1 Convolution results with and without the projections of the time-frequency filter

3-Q-20

3-Q-20 時間伸長圧縮処理を用いた 音声情報ハイディング法の検討

Study on speech watermarking method with time-stretching and compression process

◎磯山拓都(都立産技高専), 鶴木祐史(北陸先端大)

【目的】時間伸長圧縮処理を用いた音声情報ハイディング法の提案

【提案法】

- 透かしの埋め込み: 音声信号に対して時間伸長および圧縮処理 (Fig. 1) を適用し、異なる時間スケールでの埋め込み
- 透かしの検出: 時間伸長圧縮処理後の複素スペクトル変化特性を利用した検出

【結果】音質評価 (MOS 値) において従来法 (LSB 法) と同等の性能を達成 (Fig. 2(a))

BER で従来法 (DSS 法) と同等の耐性を保持 (Fig. 2(b))

【結論】時間伸長圧縮処理を用いた音声情報ハイディング法を実現した

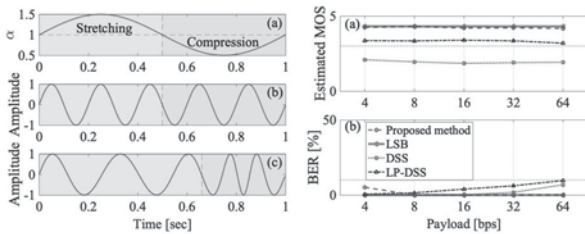


Fig. 1: Example of time-stretching and compression process: (a) time-stretching and compression ratio, (b) estimated input signal, (c) processed input signal.

Fig. 2: Results of sound quality evaluation and BER: (a) MOS values of sound quality estimated by NISQA and (b) BER.

3-Q-22

3-Q-22 重み付き更新量を用いた 適応アルゴリズムによるインパルス応答推定

Impulse response estimation using adaptive algorithm with weighted update

☆村田佳斗, 深田壮真, 矢田部浩平(農工大), 泉悠斗, 高橋祐, 近藤多伸(ヤマハ株式会社)

研究背景

適応アルゴリズムのステップサイズに対する重み付けがインパルス応答推定精度の向上に寄与

- ▶ より高精度な推定を実現する重みを検討

比較手法

定数重み, 指数減衰型重み, 包絡重みの3種類

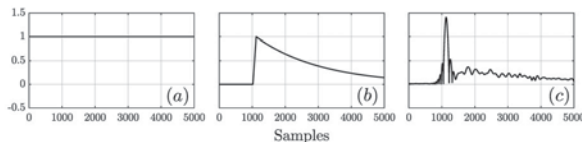


Fig. 1: (a) Constant, (b) Exponential decay, and (c) Envelope weight

結果 包絡重みが最も高い推定精度

- ▶ 真のインパルス応答の概形に基づく重みが有効

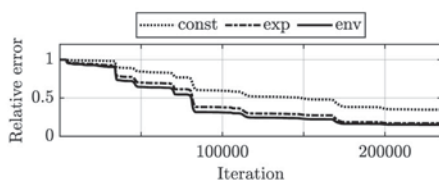


Fig. 2: Results of the estimation relative error by the number of iteration

3-Q-21

3-Q-21 重み共有型分類半制限ボルツマンマシンによる 全音程関係と時系列を考慮した和音認識

Chord recognition considering all interval relations and time series using weight-sharing classification semi-restricted Boltzmann machines

☆石川峻弥, 中鹿亘(電通大)

◆目的

- ▶ 以前に提案した分類半制限ボルツマンマシン (CSRBM) を、時系列を効率的に考慮できるように改良する

◆提案手法

- ▶ 考慮するフレーム数だけ CSRBM を並列に接続し、一部の重みを共有した 重み共有型分類半制限ボルツマンマシン を提案
- ▶ 2音の共起性を学習する重みに 倍音構造の事前知識を導入 し、より効率的な倍音の考慮を実現

◆結果

- ▶ 認識精度向上、および一部の重みを共有することの有効性を確認
- ▶ 多様なデータセットでの実験、操作したい音程の共起性のみを的確に操作できる手法の検討が今後の課題

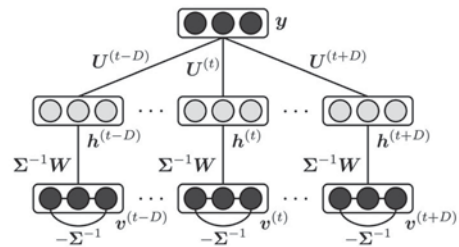


Fig. Weight-sharing CSRBM

3-Q-23

3-Q-23 音光変換デバイス「ブリンクキー」を用いた 音響シーン分類のためのデータセット構築の検討

Development of prototype "Blinky" dataset for acoustic scene classification.

☆木次駿, 河村隆生, 小野順貴(都立大)

◆目的: ブリンクキーを用いた音響シーン分類

- ▶ ビデオカメラによって同期された広範囲の音情報を伝送できる

◆収録: ブリンクキーを部屋に分散配置 (Fig. 1)

- ▶ シーン分類に有用なブリンクキーの光らせ方を検討するために、実際にブリンクキーで観測される音響信号を取得した
- ▶ "Vacuum cleaner", "Watching TV", "Chatting", "Absence" の4つのシーンを収録

◆実験: ブリンクキーのパワーの時間変化を観察

- ▶ 各ブリンクキーで得られたパワーの時間変化を観察することで、シーンごとに特徴があることを確認
- ▶ パワーの時間変化を入力としたシーン認識を行った結果、"Chatting"を除き、F1-score が7割を超えることを確認



Fig.1: Recording environment, Blinky arrangement, and approximate positions of sound sources

3-Q-24

3-Q-24 犯罪抑止を目的とした悲鳴検知システムの街頭における多地点実装事例

The case study of scream detection system across a city to prevent crime

○磯部 忠明, △河目 瞬, △船橋 勇那,
△桂田 紗希, △木村 卓哉(総合警備保障)

- ◆兵庫県の加古川市では、「加古川市スマートシティ構想」の一環として、市内における悲鳴検知システムの設置を進めている。
- ◆本稿では、悲鳴検知システムの機能である悲鳴検知と、自動で音声警告をする機能について紹介し、屋外かつ多地点における悲鳴検知システムを運用する上で発生した技術的課題等について報告する。
- ◆Fig.1 に悲鳴検知システムの外観を示す。ボックス外部に警告灯、スピーカー、カメラ、マイクロフォンを装着し、内部には悲鳴検知、警告灯制御、スピーカー制御などの処理を担うエッジPC が格納されている。悲鳴もしくは怒声を検知した場合には、警告灯を点灯し、スピーカーから警告音声を再生することで、犯罪に対する抑止効果を狙う。
- ◆本システムを屋外の多地点で運用する上で、現地音声の多様性による誤検知や、マイクロフォンの故障覚知が困難といった課題があった。

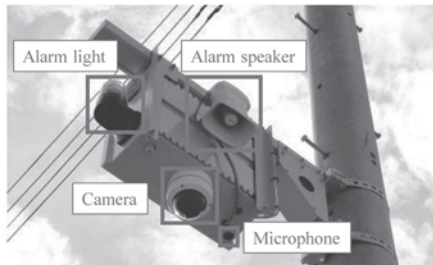


Fig.1 The exterior of the detection system

3-Q-26

3-Q-26 深層学習を用いた フィードバック型 ANC の軽量化

Lightweight Feedback Active Noise Control Using Deep Learning

○人見一也, 杉浦陽介, 島村徹也 (埼玉大)

目的

従来法である DNoiseNet の高い騒音抑制性能を維持しつつ、小型デバイスへの実装に向けた軽量化なフィードバック型 ANC アルゴリズムを提案する。

提案法

提案法では、計算負荷の大きい畳み込み層を除去し、GRU2 層を SRU1 層と GRU1 層に置き換えることで計算量を大幅に削減した。また、全結合層間に SeLU を導入することで、勾配爆発を抑えつつ安定した学習を実現している。

実験および結果

実験には NoiseX-92 の車内ノイズ、コックピットノイズ、工事現場ノイズを使用し、ノイズ減衰量、パラメータ数、計算時間を評価指標とした。提案手法は従来法に比べ、パラメータ数を約4分の1、計算時間を約3分の1に削減しつつ、ノイズ抑制性能を維持できることを確認した。

3-Q-25

3-Q-25 スペクトログラムの再構成に基づく 多重音高推定と楽器推定に関する検討

A study on multi-pitch estimation and instrument identification based on reconstruction of spectrograms

☆小倉真了, 坂野秀樹, 旭健作(名城大院・理工学研)

- ◆先行研究：音色情報の抽出とスペクトログラムの再構成に基づく多重音高推定モデル (Fig.1 赤破線で囲んだ部分)
- ◆提案手法：音色情報を利用して楽器推定を行うよう拡張 (Fig.1)

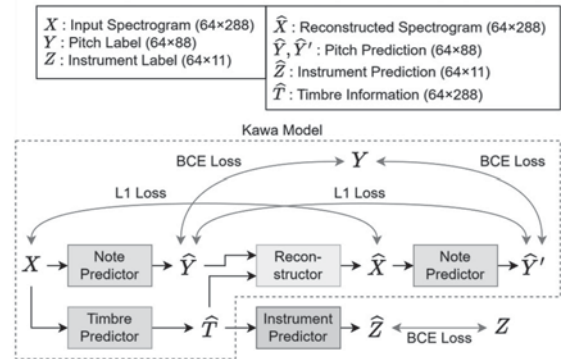


Fig. 1: Architecture of the proposed model

◆結果

- 楽器推定は可能だが、スペクトログラムから直接推定する場合と比べて Average Precision は向上しない (91.8% → 最高 90.9%)
- 音高推定の AP も向上しない (79.9% → 最高 79.8%)
- 音高推定と楽器推定の精度がトレードオフの関係にある可能性
 - ◇ 一方に最適化された音色情報は他方にあまり適さない?

3-Q-27

3-Q-27 疑似船舶航行音に対する 2 点観察信号を 誤差信号とした適応振動制御

Active window vibration control system using the single error signal added at two observation points signals for pseudo boat navigation noise.

☆中島尚哉, 武藤憲司(芝工大)

- ◆我々のこれまでの研究では、1つの振動観察点と1つの誤差信号を用いた ANC で、単一周波数を対象に制御した。
- ◆本研究は窓からの透過音低減のために、2か所の振動信号を加算し1つの誤差信号とする適応振動制御システムを検討した。特にタグボートの航行音に似た周波数を対象とした。
- ◆音圧レベル低減量は3周波数の信号を加算した時、窓全体の平均2.1 dB となった。音圧レベル測定点40地点中36地点の音圧レベルを低減できた。これより、以前の単一周波数を対象とした実験に加えて実際の航行音に近づけた場合でも効果を発揮した。

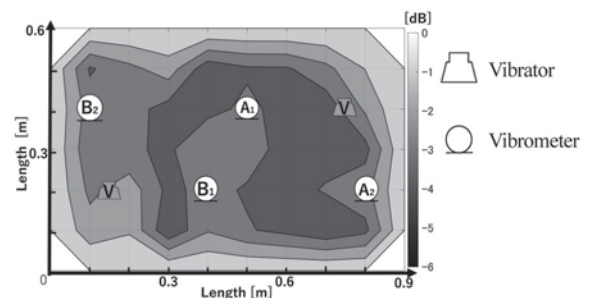


Fig.1: Contour plot of the attenuation level of the sound pressure. It was included the positions of vibrator and vibrometers A and B.

3-Q-28

3-Q-28

閉ループ構成の深層学習に基づく補聴器のための音響フィードバックキャンセラーの性能評価

Performance evaluation of a deep learning-based acoustic feedback canceller for hearing aids trained in a closed-loop configuration

◎湯野悠希, 佐藤僚, 屋間信彦(リオン)

- ◆補聴器のための音響フィードバックキャンセラー(AFC)として、DNNを用いたNSM-AFCを提案し、基礎的な性能評価を行った。
 - 適応フィルタを用いた従来法では、周期性の高い信号の入力によって推定誤差が大きくなり、収束速度と推定精度にトレードオフの関係が存在する。
 - DNNのみで時間周波数マスク推定を行うAFCとすることで、周期信号による推定誤差の低減やトレードオフの解消を試みた。
 - また、AFCは閉ループの問題であるため、NSM-AFCは学習内に閉ループを組み込んで学習を行った。
- ◆性能評価の結果、NSM-AFCはフィードバック経路の変化に追従でき、周期信号の入力に対してもロバストであることが確かめられた。

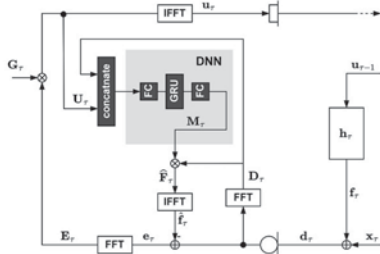


Fig.1: Block diagram of NSM-AFC

3-Q-30

機械学習を用いたMFCCによる心不全診断システムの基礎研究

Fundamental study on a heart failure diagnosis evaluation system with using MFCC and machine learning

☆柳原直也, 藤村秀弥, 大村美結, 大島風雅,

中山仁史(広島市大院), 田村雄一(国際医療福祉大)

- ◆音響特徴量としてMFCCを採用し、MLPを用いた心不全診断システムの検討を行う。
- ◆対象として入力を20次元のMFCC、出力をHF(Before, During and After)の3群及びCOの1群の計4クラス分類のMLPを構築する。
- ◆ランダムなシード値においても95%以上の認識結果を確認した。
- ◆実験結果より、テスト用データにおいて98.3%の認識性能を確認した。

Table.1: Loss Value and Accuracy in 4-class classification.

	Loss	Accuracy
Training	0.026	0.991
Validation	0.096	0.970
Test (Average)	-	0.960
Test (Best)	-	0.983

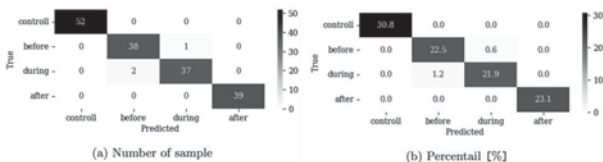


Fig.1: Recognition result of MLP on confusion matrix.

3-Q-29

3-Q-29 腸蠕動音による排便前後の識別の検討

Classification of pre- and post-defecation by intestinal peristalsis sounds

☆川畑秀斗, 大塚 翔, 中川誠司(千葉大)

- ◆腸蠕動音を連続的に観測することで排便時間の予測が試みられている。排泄ケアは介護現場において最も負荷の高い作業の一つであり、排便時間の予測は介護の効率化に大きく寄与する可能性が高い。
- ◆健康成人男性9名を対象に、排便前後の蠕動音を聴診器マイクロフォンで計測した。腸内容物の移動を反映する大きな腸蠕動音ピークを切り出したData1と、腸蠕動音ピークに先立つ腸運動そのものを反映する区間を切り出したData2の2つのデータセットを作成した。
- ◆各データセットに対し、RMS、スペクトル重心、振幅閾値を超えた音の持続時間、尖度、歪度、角周波数帯域のエネルギーの各特徴量を抽出した。次元削減を行ったのち、Random Forestを用いた排便前後の分類を試みた結果、Data1では60.8%の正解率、Data2では62.8%の正解率が得られた。200-300 HzのBand energy、尖度、RMSが高い重要度を示した(Fig.1, Fig.2)。
- ◆排便前後の蠕動音をある程度識別できることが示されたが、十分な分類精度が得られたとは言えない。多様な特徴量の選択やデータの拡充、分類アルゴリズムの改善を行い、精度向上を図る必要がある。

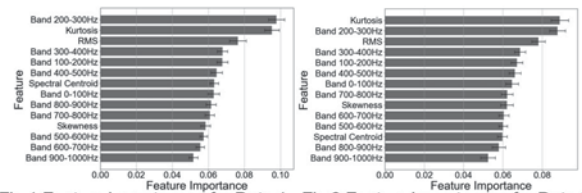


Fig.1 Feature Importances for Data 1. Fig.2 Feature Importances for Data 2.

3-Q-31

3-Q-31 LLDを用いた心不全診断システムのための音響特徴量の評価

Evaluation of acoustic features for a heart failure diagnosis system using LLD.

☆大村美結, 藤村秀弥, 大島風雅, 中山仁史(広島市大院), 田村雄一(国際医療福祉大)

- ◆心不全診断システムのための音声入力を用いた音声バイオマーカーを検討している。
- ◆本稿では感情認識に関する研究で用いられている音響特徴量を対象に心不全の診断に有用な探索的な評価を行う。
- ◆実験結果より音響特徴量セットに存在する88特徴量及び6,373特徴量を対象に探索したところ、心不全の識別に有用な音響特徴量を複数個得られることが期待できることを確認した。

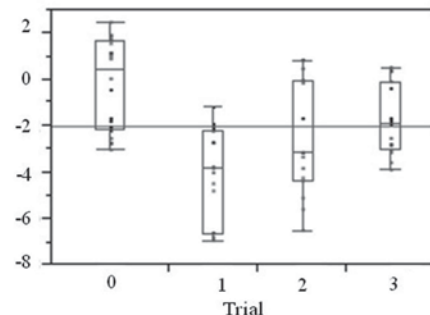


Fig.1 : Distribution of a feature at eGeMAPSv2.

3-Q-32

3-Q-32 音声想起分類のための電流源推定に基づく脳磁図データ拡張の検討

Data augmentation based on source estimation for imagined speech classification

☆能勢幸樹, 矢野肇, 高島遼一, 滝口哲也(神戸大), 中川誠司(千葉大)

- ◆脳活動データのデータ拡張では、一般的に、頭外のセンサで観測された信号に対して処理が行われるが、脳内の活動源における信号に対して直接処理を行うことで効果的なデータ拡張が行える可能性がある。
- ◆本研究では、音声想起時の脳磁図に対し電流源推定を行った後、推定された脳内の電流源の時系列に対して、その振幅や周波数成分を変化させる6通りのデータ拡張を適用した。
- ◆データ拡張を適用した電流源から合成した脳磁図データ (Fig. 1) を用いて、想起された音声を脳磁図データから識別する分類モデルを訓練した。
- ◆実験の結果、電流源空間での Threshold dropout と Fourier transform surrogate を用いたデータ拡張が分類精度の向上に特に寄与した。
- ◆電流源空間におけるデータ拡張は、音声想起に関連する電流源を直接操作することで、より頑健な分類モデルの構築に寄与する可能性が示された。

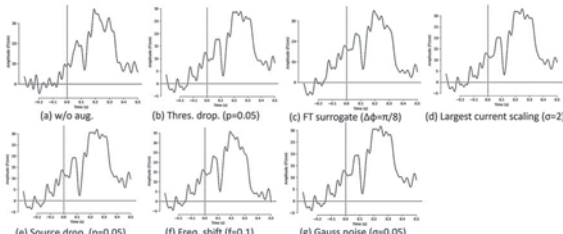


Fig.1: (a) Observed and (b-g) reconstructed MEG waveforms at an MEG channel around the right temporal area.

3-Q-34

3-Q-34 500Hz 以下の低い周波数を対象とした板振動利用複数周波数吸音音響メタマテリアル

The evaluation of the developed four peak sound absorber using the plate vibration for under 500Hz noise

◎後藤 達彦, 江波戸 明彦, 西村 修 ((株) 東芝)

- ◆本稿では、製作容易な吊り下げ部材2種を、板に適切に配置することで、4周波数吸音を実現する構成を提案した。音響管評価にて、4周波数吸音の妥当性、並びに、副産物として追加1周波数の吸音効果が生じることを示した ⇒ 合計5周波数吸音

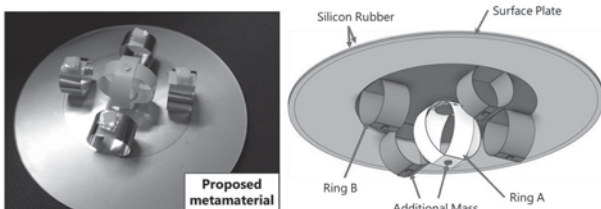


Fig.1: The structure of the proposed sound metamaterial.

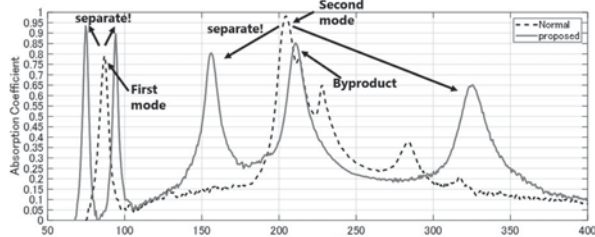


Fig.2: Measured absorption coefficient.

3-Q-33

3-Q-33 オートエンコーダによる豚呼吸器病の早期異常検知

Early detection of abnormalities in swine respiratory disease by autoencoder

☆神谷 太星(広島市立大), 的場 悠基, 石光 俊介,

△湯本 誠司(広島市立大院), △三上 修, △小川 洋介, △井上 寛暁,

△石田 三佳(農研機構), △田中 大介(新居浜高専)

- ◆養豚の大規模化により豚の健康管理の効率化が問題となっている。本研究では豚呼吸器病の早期発見を目的とし、時系列特徴量を入力とした異常検知モデルを構築し、その有効性を確認した。
- ◆本研究では豚の体内伝導音を収録するために、耳標センサを開発した。このセンサはバッテリー、データ送信機、ピエゾセンサなどで構成され、豚の体動や脈拍を収集し、Wi-Fi 経由で PC にデータを送信することができる。
- ◆呼吸器病罹患豚の体内伝導音を収集するため、豚胸膜肺炎を引き起こす細菌である Actinobacillus pleuropneumoniae (App) を使用した。耳標センサを装着した5週齢の豚に対し、菌接種前後の状態を収録してデータを収集した。
- ◆獣医師が重度と判断した2頭の豚(豚Aと豚B)のうち、特に症状が多く見られた豚Aを対象に、クリアランス係数など4つの時間領域パラメータの解析を行った。この解析に基づき前処理を行ったデータを入力とし、オートエンコーダによる異常検知モデルを構築した。
- ◆今回の個体では高い精度での異常検知が可能であった。また、時間領域パラメータの前処理が識別結果に影響を与えることを確認した。

3-Q-35

3-Q-35 アブラゼミ1匹の主鳴音における騒音レベルの変動幅分析 #2024年豊洲

Fluctuating range analysis of A-weighted sound pressure level in courtship song of a single large brown cicada #Toyosu 2024

◎御園玲央, 武藤憲司(芝浦工大)

- ◆我々のこれまでの研究で、28 kHz まで自由音場とみなせる範囲内で、ミンゼミ1匹が発した主鳴音1回当たりの騒音レベルとその変動幅を明らかにした。
- ◆本研究では、70 kHz まで自由音場とみなせる範囲内で、アブラゼミ1匹が発した主鳴音1回当たりの騒音レベルとその変動幅を明らかにした。また、主鳴音1回の持続時間の基本統計量を算出した。
- ◆測定したアブラゼミは60分で主鳴音を57回発した。
- ◆主鳴音1回当たりの等価騒音レベル L_{Aeq} の平均値は80.0 dB、変動幅は9.2 dBであった。
- ◆主鳴音1回当たりの持続時間の平均値は24.0秒であった。

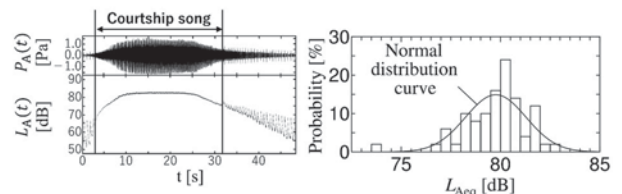


Fig. 1 Time waveform of the courtship song produced by a large brown cicada ($r_2 = 1$ m)

Fig. 2 Histogram of the L_{Aeq} at 1 m reference of the courtship song produced by a large brown cicada. (Number of samples: 57)

3-Q-36

3-Q-36 共鳴型音響メタマテリアル構造による吸音性能の検討

Acoustic metamaterial structure with helical resonator
○佐竹正義, 松浦充保(SOKEN), Δ久保俊逸(デンソー)

- ◆車両の電動化に伴い、従来よりも高い静粛性が要求されている。その実現に向けて、コストアップにつながる特殊な吸音材を用いず、限られた空間容積で高い吸音性を実現することを目指している。
- ◆そこで、樹脂(ポリプロピレン材など)の板材内部に **i 屈曲させた共鳴管** と、 **ii 共鳴管に積層した空洞部** を設けるメタマテリアル構造を構成した(Fig. 1)。この構造により薄型小体格で所望周波数(700~850 Hz)での共鳴現象による吸音効果を実現した。
- ◆Fig. 2は、既存のウレタン吸音材(カームフレックスF2, t10)と考案したメタマテリアル構造を周期配置したt10板材のそれぞれの吸音特性を示している。考案メタマテリアル構造は700~880 Hzにおいて、同厚さの既存ウレタン吸音材と比較して高い吸音特性を有することを明らかにした。

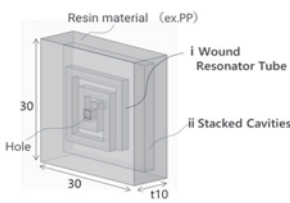


Fig.1: Proposed metamaterial.

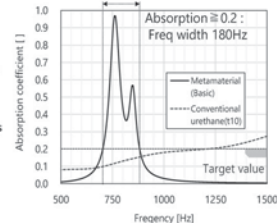


Fig.2: Sound absorption properties.

3-Q-38

3-Q-38 短時間騒音が睡眠に及ぼす即時的影響に関する基礎的研究

Experimental study on the immediate effects of short duration noise on sleep
☆松原千青, Δ太田周彰, 平栗靖浩(近畿大)

- ◆背景・目的
音圧レベルや周波数などの音の特徴による睡眠深度への影響の違いや、騒音曝露後の睡眠深度への即時的な影響は明らかになっていない。そこで、睡眠中に異なる複数種類の音を提示する被験者実験を行い、睡眠深度への影響を検討した。
- ◆検討項目と結果
 - ・音の提示の有無による睡眠深度への影響 (Fig. 1)
 - 音の提示によって中途覚醒の回数が有意に大きくなった。
 - ・提示音の特徴による音提示前後の睡眠深度への影響の違い (Fig. 2)
 - 音圧レベルが大きいと睡眠深度に即時的影響がある。
 - 周波数の違いによる差は見られなかった。

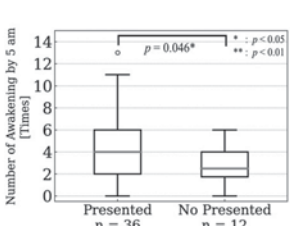


Fig.1: The difference in the number of awakenings with or without noise presentation.

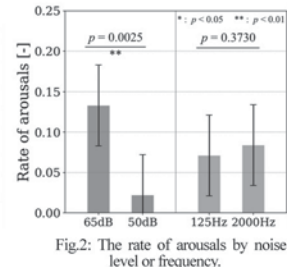


Fig.2: The rate of arousals by noise level or frequency.

3-Q-37

3-Q-37 検出閾を下回る骨導超音波によるマスキングの可能性

Characteristics of masking by bone-conducted ultrasound at levels below the threshold
☆出山嘉俊, 大塚 翔, 中川誠司(千葉大)

- ◆ヒトの聴覚が超音波にもある程度の感度を持つことが示されている。
- ◆しかしながら、超音波の聞こえは概してか細く、その検出閾やピッチにはバラツキが多い。また、明瞭なピッチ感が生じる検出閾を下回るレベルの超音波に対しても何らかの“気配”が感じられるという報告も存在し、受容器に有意に作用している可能性がある。
- ◆検出閾を下回るレベルの骨導超音波が気導可聴音の聴取に及ぼすマスキング効果を調べた。
- ◆30 kHz の骨導トーンバーストに対する検出閾を基準とし、(1) -1 dB SL, (2) 0 dB SL, (3) 5 dB SL の骨導超音波マスカ存在下、および (4) マスカなしの条件下で 12-19 kHz の気導トーンバーストの検出閾を推定した。
- ◆5 dB SL 骨導超音波マスカの提示下では、気導音の検出閾は上昇する傾向が観察された一方、-1 および 0 dB SL の骨導超音波マスカ提示下では気導音の検出閾が低下する傾向が認められた。
- ◆これらの結果は、検出閾を下回るレベルの超音波であっても内耳に作用する可能性があることを示す。また、検出閾下の骨導超音波マスカによって、内耳で“確率共鳴”が生じている可能性が示唆される。

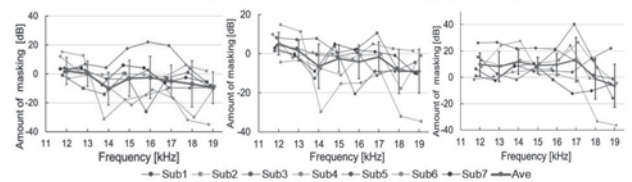


Fig.1: Amounts of masking for each masker. (Left: -1 dB SL, Center: 0 dB SL, Right: 5 dB SL)