1-1-1

1-1-1　GST-BERT-TTS：アクセントラベル不要な複数話者日本語 TTS

GST-BERT-TTS：Japanese Multi-speaker TTS without accentual labels

◎小椋忠志，岡本拓磨，大谷大和，Erica Cooper(NICT)，
戸田智基(名大/NICT)，河井恒(NICT)

- ◆辞書依存のアクセントラベルを必要としないデータ駆動型の複数話者日本語 TTS 用韻律予測手法の提案
- ◆単一話者向け従来法（f_0-BERT）にグローバルスタイルトークン（GST）を統合し、話者特性を考慮した韻律予測を実現
- ◆テキストとカタカナ列から韻律を予測し、複雑な日本語の韻律に対応
- ◆複数話者 TTS コーパスでの評価実験により、アクセント正確性と韻律予測精度の両面で従来手法を上回る性能を確認
- ◆アクセントラベル不要でありながら実用的な合成音声品質を実現

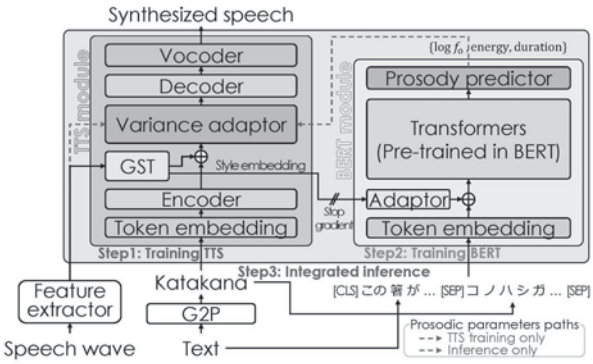


Fig. Proposed GST-BERT-TTS framework. Step 1: Train the TTS module. Step 2: Train the BERT module using features extracted from the TTS module. Step 3: Perform inference using both modules.

1-1-2

音声からの音素・韻律ラベルの獲得とその応用

Phonemic and prosodic labeling of speech and its applications

◎白旗悠真，朴炳宣，山本龍一(LINE ヤフー)

- ◆テキスト音声合成において、高品質な音声とテキストの対データは必要不可欠である。
- ◆音素・韻律アノテーションは専門家の知識が必要であるため大量に用意することが困難である。
- ◆本論文では、既存の音声認識モデル (whisper) をラベルありデータで適応することで、音声から自動で音素・韻律のアノテーションを行えるモデルを提案する。
- ◆適応に用いるデータの量を増やすため、補助テキスト音声合成モデルを用いたデータ拡張を合わせて提案する。
- ◆実験の結果、テキスト音声合成のタスクにおいて提案法を用いて生成したラベルが、正解ラベルを使った条件と同程度の品質を達成した。

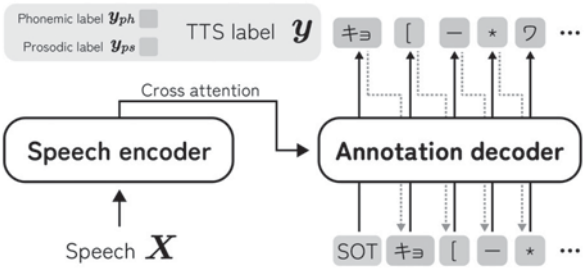


Fig. 1 音声から音素・韻律のアノテーションモデル

1-1-3

1-1-3 アクセント制御音声合成と主観比較評価最適化による専門家非依存アクセントアノテーション法

Expert-independent accent annotation method using accent-controlled speech synthesis and optimized subjective preference evaluation.

◎安田裕介（NII, 産総研），井本桂右 (京都大学)，深山覚（産総研），戸田智基 (名古屋大学)

本研究は，アクセント制御音声合成とトーナメントに基づく動的比較評価を組み合わせることで，専門家非依存のアクセントアノテーション法を提案します．提案法は，高精度かつ安価なアクセントアノテーションの実現を可能にします．

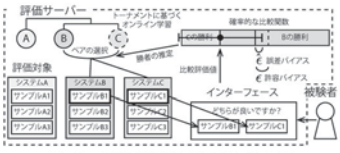


図1: トーナメントに基づく動的比較評価の概要図.

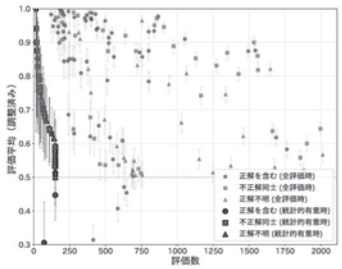


図2: 2つのアクセント型の比較評価平均と評価数の関係.

1-1-4

1-1-4 音素BERTと音声基盤モデルを用いた自動韻律アノテーションの検討

A Study on Automatic Prosody Annotation Using Phoneme BERT and Speech Foundation Models

○郡山知樹 (サイバーエージェント)

- ▸ **背景**
  - ▸ 韻律制御可能なTTSの実現には韻律アノテーションが必要
- ▸ **提案手法**
  - ▸ 音声の韻律ラベル予測に基盤モデルを使用
    - ▸ 音声基盤モデル (HuBERT, Whisper)
    - ▸ 言語基盤モデル (PnG BERT, PL-BERT)
- ▸ **実験**
  - ▸ アクセント記号、高低、境界強度(BI)、ポーズ有無のマルチラベル予測
  - ▸ 結果から、音声と言語の基盤モデルの組合せが有効

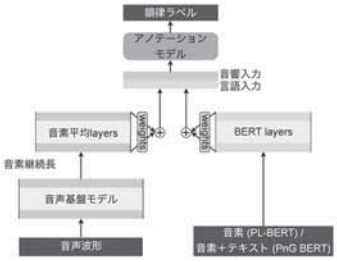


Fig 1: Proposed model architecture.

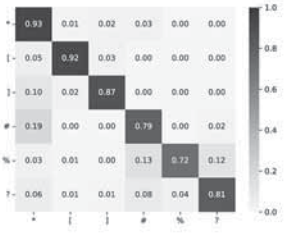


Fig 2: Confusuin matrix on accent labels.

1-1-5

1-1-5　日本語テキスト音声合成のための高低アクセントの直接推定に関する検討

An Investigation into the Direct Estimation of Pitch Accent for Japanese Text-to-Speech Synthesis

☆苅谷楓，藤本崇人，橋本佳，南角吉彦，徳田恵一（名工大）

◆背景
- 日本語 TTS におけるルールベースなアクセント推定の限界
- 統計的手法が提案されているが，データ確保の困難性や推定精度，アクセントの制御性に課題が残る

◆提案手法
- 標準的な日本語 TTS ソフト Open JTalk から得られる情報や単語埋め込みなどを活用し，同等の性能を持つモデルを学習する
- その後文章に対応する音響情報を用いて強化学習を行い，性能の向上を図る

◆実験結果
- 学習コーパスや入力特徴量を適切に選ぶことで，Open JTalk とほぼ同等の性能を持つ推定モデルが作成できることが示唆された

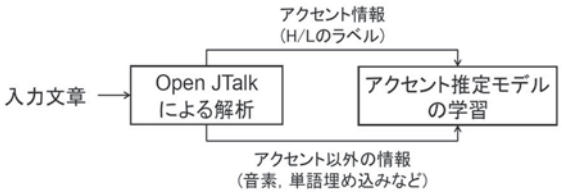


Fig.1 : Structure of the proposed model

1-1-6

1-1-6　大規模弱教師あり日本語音声認識におけるフィルタリングとデータ選択の検討

Filtering and Selection for Large-scale Weakly-supervised Japanese ASR

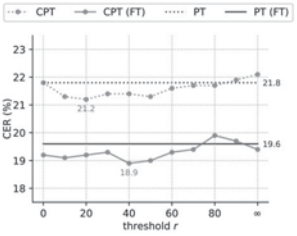
◎松浦 孝平(NTT，京大)，芦原 孝典，三村 正人(NTT)，河原 達也(京大)

■概要: 大規模な弱教師ありデータセットを用いた日本語音声認識モデルの学習における，フィルタリングとデータ選択を検証

■フィルタリング: 文字誤り率（CER）をもとにラベル品質の低いサンプルを除外し事前学習を継続．データの品質向上と多様性の減少による認識精度のトレードオフを確認（Fig. 1）

■ドメイン適応のためのデータ選択: 評価用ドメインに近い音響特徴のサンプルを，自己教師ありモデル表現を用いて選択しファインチューニングすることで，認識精度の向上を確認（Table. 1）

■提案手法の有効性と一般性: シンプルな CTC ベースの ASR モデルでも既存の大規模モデル Whisper を上回る認識精度を達成．以上の手法は日本語以外の言語にも適用可能



		proposed	random
threshold r	0	37.9±0.3	38.4±0.2
	10	36.3±0.4	37.3±0.3
	20	35.7±0.3	36.7±0.2
	30	**35.5**±0.1	36.4±0.3
w/o FT ($r = 30$)		37.0	

Fig.1 (left): CERs (%) with various filtering thresholds on the CommonVoice Japanese split. (C)PT: (continued) pre-training/ FT: fine-tuning/ threshold r: CERs between model outputs and original transcription on the large-scale weakly-supervised dataset.

Table 1 (right): CERs (%) after fine-tuning on samples similar to the target domain, i.e., the Noisy-KU dataset, selected from filtered sub-datasets.

1-1-7

1-1-7　目的話者と非目的話者を同時に考慮したシングルチャネルマルチトーカー音声認識

Target and Non-Target Speakers Single Channel Multi-Talker ASR

○増村 亮，牧島 直輝，山根 大河，庵 愛，佐藤 宏，田中 智大，
高島 瑛彦，鈴木 聡志，森谷 崇史，安藤 厚志（NTT）

◆本稿では，目的話者音声の事前登録の上で，目的話者と非目的話者を見分けながら両者の音声を認識する目的・非目的話者同時音声認識 (TS-NTS-ASR;Target and Non-Target Speakers ASR) を提案する．

◆提案手法は，条件付き自己回帰モデルの枠組みで複数話者の重畳音声を同時に音声認識する手法に対して，目的話者と非目的話者を見分ける機構を導入することで実現される．具体的には，各話者の発話テキストを構成するためのトークンと，目的話者と非目的話者を見分けるためのトークンの両者を自己回帰モデルの範囲で同時に扱うことで，これを実現する．モデル構造は図１のとおりである．

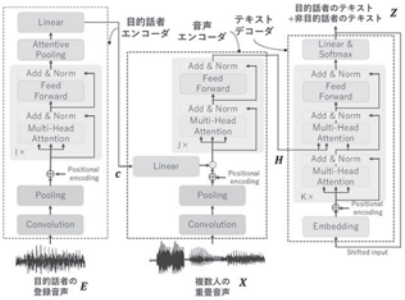


図１. TS-NTS-ASR のモデル構造.

1-1-8

1-1-8　目的話者登録の有無を切り替え可能なシングルチャネルマルチトーカー音声認識

Single Channel Multi-Talker ASR with Switchable Target Speaker Enrollment

○増村 亮，牧島 直輝，田中 智大，庵 愛，河田 尚孝，
折橋 翔太，山根 大河，鈴木 聡志，森谷 崇史，安藤 厚志（NTT）

◆本稿では，目的話者登録駆動型MT-ASR と目的話者非依存型MT-ASR の両者を統一的なモデリングで表現し，目的話者の有無を切り替えて動作可能な統合型 MT-ASR を提案する．

◆提案手法では，取り外し可能な目的話者エンコーダを備えた条件付き自己回帰モデルを構成し，モデルパラメータを両システム形態で共有する．モデル構造は図１のとおりである．

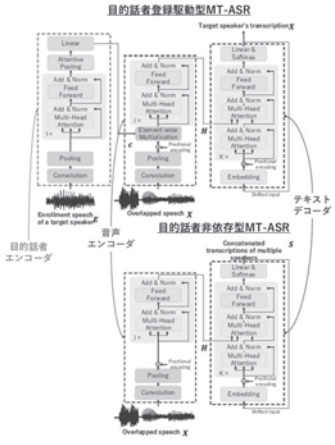


図１. 統合型 MT-ASR のモデル構造.

1-1-9

1-1-9　音声要約における抽出要約と抽象要約の相乗効果の調査

Integrating Extractive and Abstractive Speech Summarization

◎叶高朋，福田りょう，小川厚徳，デルクロア・マーク（NTT, inc.），¥¥チェン・ウィリアム，渡部晋治（CMU）

◆End-to-end 音声要約は、非常に長い音声特徴量系列を入力とし、直接長期の文脈を学習し、重要情報を抽出するため困難なタスクである。

◆この研究では、長い音声から重要な発話のみを書き起こす抽象音声要約と、重要な内容を自由形式で生成する抽象要約を組み合わせて要約精度を改善した。

◆提案法は、抽象要約と抽象要約を１つのモデルで学習し、１度のデコーディングで生成することができる。はじめに抽出要約を生成することで、抽象要約は抽出要約結果に条件づけられる。

◆下記の要約例では、従来の End-to-end 要約は全体の内容を捉えているが、詳細（専門用語）を要約に含むことができていない。一方で、提案法では、抽出要約に専門用語を含む発話を抽出し、抽出要約は抽出要約結果を参照することで、専門用語を含む情報欠落が少ない要約を生成できる。

Table 1　要約例

正解抽象要約	Are you a giver or a taker? In every workplace, there are three basic kinds of people: givers, takers and matchers. Organizational psychologist Adam Grant breaks down these personalities and offers simple strategies to promote a culture of generosity and keep self-serving employees from taking more than their share.
提案法：抽出要約	... if you're more of a giver or a taker, and you can take it right now. So just out of curiosity, how many of you self-identify more as givers than takers or matchers? I want to talk about what it takes to build cultures ...
提案法：抽象要約	How to find the giver person in your organization & How do you catch a taker before it's too late? In this funny, funny talk, social entrepreneur Adam Rifkin shows how to find the most paranoid people in your organization – and how you can create a culture of successful givers.
End-to-end	How to build a culture of successful givers & How do you build a culture of successful and generous givers? In this funny talk, TED Fellow Adam Rifkin shows how to build a culture of giving and taking – and how to build a culture.

1-1-10

1-1-10　TV-CAR 分析による正弦波分析

Sin wave analysis using TV-CAR analysis

○舟木慶一（琉球大学）

線形予測分析 (Linear Predictive Coding: LPC) は 1966 年、当時電電公社の板倉、斉藤、AT &T Bell 研の B.S.Atal により独立に提案された手法であり、前者の手法は自己相関法 [1]、後者の手法は共分散法[2] と呼ばれている。LPC 分析は音声信号から自己相関、または共分散を介して、正規方程式の解法により自己回帰 (Auto-Regressive: AR) モデル係数を抽出する音声分析法 [3][4] であり、CELP 音声符号化[5][6][7][8][9][10] では必要不可欠な技術となっている。本稿では、LPC 分析の教科書的な説明を行うとともに、石山 [11] の報告のように正弦波で自己相関法より共分散法の方が良好な推定を行うのかの検証と筆者らによる TV-CAR 分析 [12] との比較を行う。

1-1-11

1-1-11

周波数領域でのスパース性に基づく線形予測分析

Linear predictive analysis based on sparsity in frequency domain

☆ 鳥居美月，矢田部浩平（農工大）

◆ **線形予測分析**（LPC）
過去の信号から現在の信号を予測し，音声のスペクトル包絡を推定する音声分析手法

従来　予測残差の L1 ノルムを最小化するスパース LPC (Conv)
・・・時間領域でのスパース性を仮定している
▶ 位相のずれにより**非スパースとなる場合がある**
★ **周波数領域では位相がずれてもスパース性を維持**

提案　1. **周波数領域**でのスパース性に基づく手法（Prop1）
2. **両領域**でのスパース性に基づく手法（Prop2）

結果　スパース LPC よりも推定誤差が低くなる可能性がある

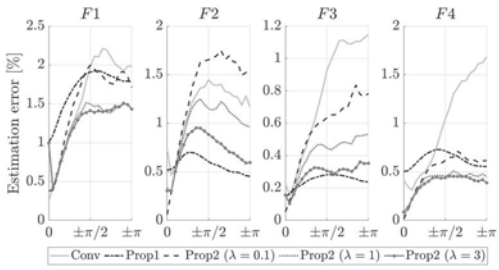


Fig. 1: Estimation error of formant frequencies

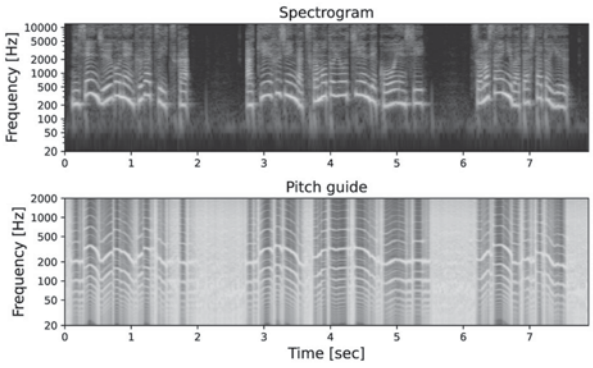
1-1-12

1-1-12　SLASH: 信号処理と自己教師あり学習を組み合わせた基本周波数推定法

SLASH: A Fundamental Frequency Estimation Method Combining Signal Processing and Self-Supervised Learning

○寺島涼，白旗悠真，川村真也 (LINE ヤフー株式会社)

本稿では，信号処理と自己教師あり学習を組み合わせた基本周波数推定法 SLASH を提案する．従来の信号処理および自己教師あり学習による基本周波数推定の精度を上回り，有声・無声判定においても優れた性能を示した．



Model	Trained on	Pitch		V/UV
		RPA (RCA) ↑	log-F_0 RMSE ↓	V/UV ER ↓
DIO	-	0.943 (0.943)	0.030	0.052
Harvest	-	0.945 (0.947)	0.045	0.083
SLASH	LibriTTS-R	**0.969 (0.969)**	0.018	**0.033**
	MIR-1K	0.967 (0.967)	**0.017**	**0.033**
PESTO	MIR-1K	0.962 (0.966)	0.057	0.098

1-1-13

1-1-13　器質性・機能性構音障害者音声を用いた小規模データセット開発と検討

Development and Evaluation of a Small-Scale Speech Corpus for Organic and Functional Dysarthria

☆小笠原朝陽，談宜育（岩手大学大学院）

◆構音障害による AI 格差是正のため、器質性１名・機能性２名から新規音声を収録しデータセットを開発、その発話傾向を分析した。

◆コーパスの質は良好。分析結果は、機能性と比べ器質性患者音声は安定した発話特徴を見せた。これを布石にし、大規模コーパス開発を目指していく。

話者	障害タイプ	品質SNR(dB)	F0平均	F0変動性	スペクトル重心	調和比
			変動係数(CV%)			
話者1	器質性	30.8	**8.74**	**19.84**	**21.67**	**26.84**
話者2	機能性1	28.7	12.76	44.05	15.60	20.97
話者3	機能性2	27.8	12.18	26.61	20.60	29.30

1-1-14

1-1-14　Common Crawl を用いた大規模音声音響データセットの構築

Construction of large audio-acoustic datasets using Common Crawl

☆淺井航平(東大/NII LLMC)，杉浦一瑳(京大/NII LLMC)，中田亘,(東大) 栗田修平(NII LLMC)，高道慎之介(慶大/東大)，小川哲司,(早大) 東中竜一郎(名大/NII LLMC)

◆近年の大規模音声モデルの開発には数百万時間単位のデータが必要であるが、特に日本語のオープンな音声音響データセットは不足している 。この研究では、新たなデータソースとして Common Crawl に着目し、継続的に更新可能な大規模データセットの構築を目指した 。

◆ウェブページの HTML ファイルのうち、lang=ja のものから音 URL から収集した音データは、全体が 38000 時間程度と推定され、そのうち日本語の音声データは 32000 時間程度であった。

◆ポッドキャストなどで使われる RSS フィードから抽出するアプローチでは、language field を絞っていないため、全体で約 450,000 時間と、より多くのデータを収集できたが 、その中で日本語の音声データは 18000 時間程度にとどまった。

◆大規模音声モデルの開発に利用できるようなより大きなデータセットを構築するためには、Common Crawl 以外のデータソースも活用していく必要がある。

1-1-15

1-1-15　音声収録に用いる機材の種類と設定が音響特性に与える影響について

The Effects of Speech Recording Equipment Types and Settings on Acoustic Parameters

○河原英紀(和歌山大/電気通信大)，榊原健一(北海道医療大)，北村達也(甲南大)，松井淑恵(豊橋技科大)，饗庭絵里子(電気通信大)，矢田部浩平(東京農工大)

◆広範にわたる品質の膨大な音声資源が生み出されている。これらの音声資源の価値を向上させるために，用途や品質に応じたクラス分けを提案し，それぞれのクラスに対する支援策の研究を進めている。

◆ここでは，様々なクラスの音声資料を収録する際の機材選択と設定を支援する測定方法とツールを紹介する。これらは，流通している高品質メディアから日常会話まで，品質に関わる音響特性などの要因を客観的に評価するために開発された。。

◆実際に使用される講演音声や音楽などを，音響測定用の信号に変換する方法である「信号安全化: signal safeguarding」を，giant-FFT SRC (Sampling Rate Conversion)の考え方に基づいて実装し直すことにより，応用範囲と使用法の柔軟性を拡大している。

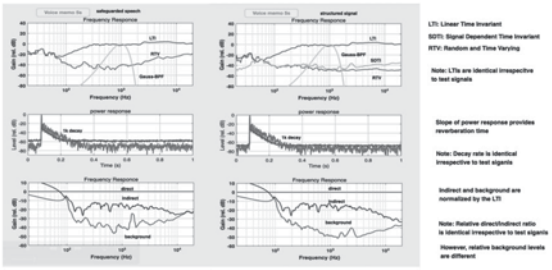


Fig.1 : Acoustic parameters derived from impulse response using a structured signal (right) and safeguarded signal (left).

1-1-16

1-1-16　The AudioMOS Challenge 2025

◎Huang Wen-Chin[1], Wang Hui[2], Liu Cheng[2], Wu Yi-Chiao[3], Tjandra Andros[3], Hsu Wei-Ning[3], Cooper Erica[4], Yong Qin[2], 戸田智基[1]
([1]名古屋大学, [2]Nankai Univ., [3]Meta, [4]NICT)

本稿では，音メディア品質推定技術に関する国際チャレンジ「AudioMOS Challenge 2025」の開催結果をまとめた。本チャレンジは 3 つのトラックで構成され，アカデミア及び企業から計 24 チームが参加した。音メディア自動評価技術の性能を明らかにするとともに，トップシステムが共通する技術を整理した。2025 年 12 月ハワイに開催される国際会議 ASRU にて特別セッションを開催しる予定。

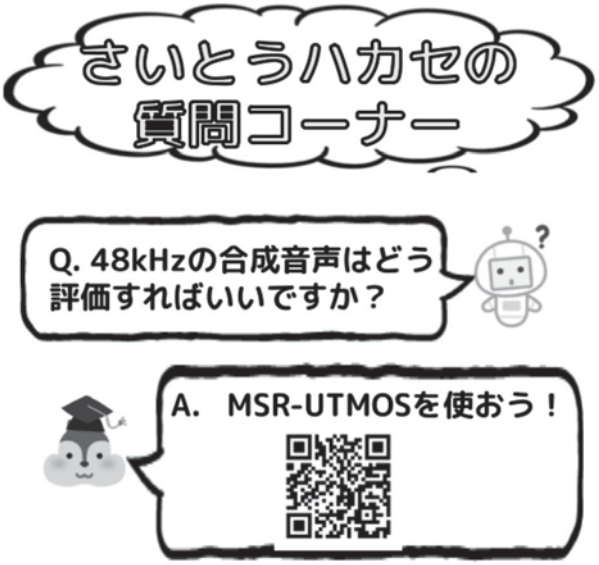
https://sites.google.com/view/voicemos-challenge/audiomos-challenge-2025

1-1-17

1-1-17　周波数非依存畳み込み層を導入した自己教師あり学習モデルによるマルチサンプリングレート自然性 MOS 予測

Multi-sampling-rate naturalness MOS prediction using self-supervised learning model incorporating sampling-frequency-independent convolutional layer

西川 剛, 中田 亘, ○齋藤 佑樹, 今村 奏海, 猿渡 洋(東大)



1-2-1

1-2-1　発話・ピアノ演奏の同時実行における遅延聴覚フィードバックの影響

Effects of delayed auditory feedback on simultaneous speech production and piano performance

○廣谷定男, 持田岳美（NTT CS 研）

◆これまでに我々は、単音節の発話および単音のピアノ演奏に対して遅延聴覚フィードバック（DAF）を適用し、発話における聴覚フィードバックの影響がピアノ演奏よりも有意に大きいことを明らかにしてきた。

◆さらに、発話と演奏を同時に行う「弾き語り条件」において、ピアノ演奏のみに遅延を加えた場合、遅延を加えていない発話の方が強く影響を受けるという予想外の結果が得られた。

◆本研究では、この結果を定量的に評価するために、状態空間モデルを用いてオンセットデータを発話および演奏ごとにモデル化した。

◆その結果、発話とピアノ演奏という異なるモダリティを同時に制御する状況において、それぞれの誤差修正に対する重みづけを動的に調整している可能性が示唆された。

	平均	標準偏差	t 値	p 値
β_S	0.169	0.078	5.707	$^{***}p<0.001$
γ_S	0.182	0.127	3.788	$^{**}p<0.01$
β_P	0.113	0.117	2.551	$^{*}p<0.05$
γ_P	0.048	0.078	1.614	$p=0.08$

Table.1 : Estimated parameters.

1-2-2

1-2-2　無声子音・有声子音の知覚：声帯振動が存在しない声による検証

Perception of unvoiced and voiced consonants: Verification using speech sounds without vocal fold vibration

○内田照久（大学入試センター・研究開発部）

目的　無声・有声子音の知覚上の弁別の手がかりや異聴傾向を検討
白色雑音から生成する疑似ささやき声系の声種を用いて検証

方法　語頭 / 語中の子音が有声・無声で対立する単語対を利用
（例: 単発 [5.5-4.0] — 断髪 [5.5-4.0]　t - d　[単語親密度]）
1. 単純再合成音声，2. 雑音駆動音声，3. 疑似ささやき声，
4. 多層動的バンドノイズ音声 の 4 条件を設定 (Fig.1)

結果　首都圏の国公立大学 (12 大学) の 1 年生 310 名 が評価
☆ **語頭子音**：雑音駆動音声は有声子音に，
疑似ささやき声は無声子音に知覚される異聴傾向
低域エネルギーが弁別手がかりになっている可能性
☆ **語中子音**：異聴傾向が消失，より正確な知覚判断に
低域エネルギーによる知覚判断への寄与が低下
☆彡 他の音響特徴が無声有声の知覚判断のドメインに
☆彡 閉鎖無音時間の時間長などが手がかりになっているか

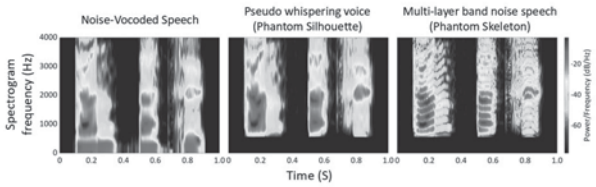


Fig. 1 Spectral envelope and structure to generate whispering voices for the experiment. (単発: tampatsu)

1-2-3

1-2-3　Study on the Contribution of Temporally Stretched and Compressed Instantaneous Modulation Components of the Temporal Amplitude Envelope to Vocal Emotion Perception

☆Taiyang GUO (JAIST), Takuto ISOYAMA (TMCIT), Shunsuke KIDANI, Masashi UNOKI (JAIST)

◆Instantaneous modulation components (IMCs) of temporal amplitude envelope contribute to explaining vocal-emotion perception.

◆Aim: To clarify whether temporal stretching and compression of IMCs contribute to the restoration of vocal-emotion perception.

◆Method: Temporal stretching and compression were applied to the IMCs, the emotion recognition rates of time-reversed noise-vocoded speech (NVS) were compared before and after temporal manipulation.

◆Results: Time-reversed NVS achieved higher emotion recognition rates after temporal manipulation than before temporal manipulation.

◆Conclusion: IMCs contribute to restore vocal-emotion perception.

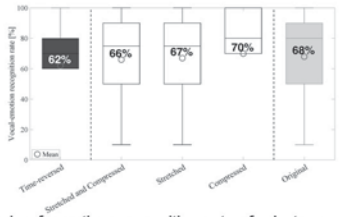


Fig.1 : Example of emotion recognition rates for hot anger: time-reversed NVS before and after temporal manipulation.

1-2-4

1-2-4　感情を考慮した宣伝音声の合成手法および主観的評価

Emotion-constrained advertising speech synthesis and subjective evaluation

◎長野瑞生，藤田健一，井島勇祐，廣谷定男（NTT）

◆音声などの外部刺激と購買意欲の関係は非線形 [Chattopadhyay+, 2003]
- 音声と購買意欲の関係は感情媒介モデルにより説明可能 [Nagano+, 2023]
- 音声と購買意欲だけでなく、感情を考慮した深層学習モデルによって購買意欲を高める音声を生成できるかは未検証

◆感情の制約あり/なしの深層学習モデルで生成した音声の比較
- 中間層に聴取者が知覚した感情（快感情・覚醒）の制約を加えたConstrained モデルと、制約を持たないUnconstrained モデルを構築
- 2 つのモデルで生成した宣伝音声を用いた大規模な主観評価実験を実施し、知覚される購買意欲や感情が異なるか検証

◆Thurstone の一対比較法とカイ二乗検定による分析
- Constrained 音声はUnconstrained 音声より購買意欲と快感情が有意に高い
- Unconstrained モデルは話速が速くなると購買意欲が高まりやすい等の単純なパターンのみを学習していた可能性がある
- 感情を考慮した宣伝音声が購買意欲の向上に有効であることを示唆

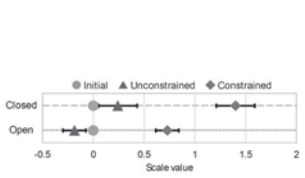


Fig.1 : Thurstone scale values for purchase intention.

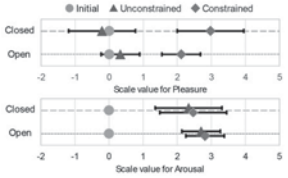


Fig.2: Thurstone scale values for perceived Pleasure and Arousal.

1-2-5

1-2-5　アナウンス音声の印象が実店舗における顧客行動に与える影響

The relationship between subjective impressions of announce voices and customer's behaviors in a store

◎林大輔（日本たばこ産業），△伊藤貴彦，△上妻志的，△稲畑伸一郎（TOA）

◆店舗における顧客行動には，聴覚情報を含めた様々な情報が影響する

◆聴覚情報のうち，BGM が主観的な感情と顧客行動に与える影響については，実店舗における研究が行われている

◆しかし，アナウンス音声の声質が主観的な感情と購買意欲に与える影響については，実験室実験における検討にとどまっている

◆そこで本研究では，声を聞いた際の主観的な感情と，実店舗における顧客行動との関係に着目して検討を行った

◆同一の文章を用いて，心理的覚醒度（低い―高い）と感情価（ネガティブ―ポジティブ）の２軸で，聞き手において異なる主観的な感情が喚起されるよう企図したアナウンス音声を複数収録した

◆収録した音声を用いて実験室実験を行った結果，音声を聞いた際の「行きたい気持ち」と主観的な感情との間に正の相関関係が示され，これは先行研究と整合的な結果であった

◆続けて，収録した音声を実店舗で流し，顧客行動について調べた結果，相対的に「心理的覚醒度が高く，感情価がネガティブ」になるアナウンス音声を用いた際に最も顧客行動が促進される可能性が示された

◆ただし，この実店舗における条件間の差は統計的に有意でなく，様々な交絡要因の影響も検討できていないため，話し方を含めた声質の違いによる顧客行動への影響についてはさらなる検討が必要である

1-2-6

1-2-6　模擬難聴とモデルによる聴覚特性理解へのアプローチ

Application of hearing loss simulation and auditory modeling for exploring auditory perception

○ 入野 俊夫 （和歌山大/電通大）

◆模擬難聴システムWHIS を用いて、健聴者が模擬難聴音を聞く聴覚心理実験を実施してきた。解説記事も出しており、音響学会誌 2022 年 12 月号では基本的な定式化と実践的な応用、2025 年 5 月号で加齢性難聴特性の理解へのアプローチの一方法としての利用を紹介した。

◆本講演では、その後の展開と技術的な課題に関して述べる。

◆WHIS の基盤となる聴覚フィルタにおいて、全体の聴力損失 HL_{total} が能動的な損失 HL_{act} と受動的な損失 HL_{pas} のdB 上での和として定式化されている。ところが、オージオグラムからは両者の比がわからず、Fig.1 の蝸牛の入出力関数（あるいは圧縮特性の健全度α）の推定が必要であるが、簡易測定法は存在せず、課題として残っている。

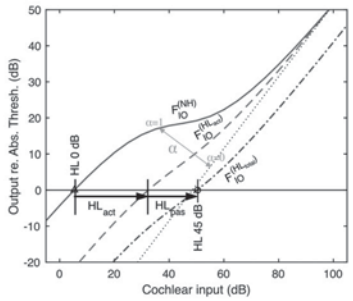


Fig.1 : Cochlear input-output functions of normal hearing $F_{IO}^{(NH)}$, degraded active process $F_{IO}^{(HLact)}$, and total hearing loss $F_{IO}^{(HLtotal)}$. Compression health is represented with a parameter α .

1-2-7

1-2-7　（招待講演）聴覚時間分解能検査の作成

Development of tests of auditory temporal resolution

○森周司（九州大）

◆著者と共同研究者による、聴覚時間分解能検査の作成に向けた取り組みを概説する。

◆検査では無音検出と振幅変調検出の閾値測定を行うこととし、ZEST を用いることで所要時間を短くすることを試みている。

◆ZEST はベイズの定理に基づき、呈示刺激に対する参加者の反応に応じて閾値の確率密度関数を尤度関数で更新し、これを繰り返すことで測定値が真の閾値に近づくようにする。

◆私達はシミュレーションとヒトでの測定実験を併用し、15 試行程度で閾値測定が完了するようにした（Fig.1）。

◆測定結果から閾値分布を推定し、無音検出閾値では分布の 99%点（4.47 msec）が時間分解能の何らかの障害の目安となることが示唆された。

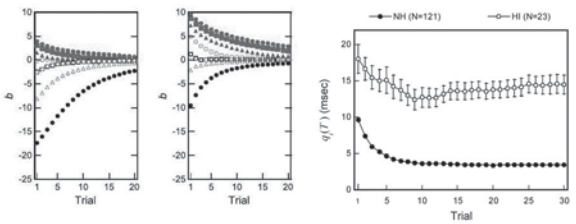


Fig.1 : Threshold estimates of simulations (left) and human measurements (right) of gap detection by ZEST, as a function of number of trials.

1-2-8

1-2-8　(招待講演)聴覚障害から聞き取り困難

Hearing impairment and Listening difficulty

○岡本 康秀(済生会中央病院・耳鼻咽喉/慶應大・医)

◆難聴では純音聴力検査で閾値の上昇は起きるが、語音聴力検査では明瞭度の低下は一律に悪化するわけではない。
◆感音難聴のうち内耳性の場合、周波数選択性の低下による聞き取りの悪化が見られるが、後迷路性の場合、時間包絡情報や時間微細構造の劣化による聞き取りの悪化が見られる。
◆聴覚補償は内耳性では有効であるが、後迷路性ではあまり補償がうまくいかない。
◆難聴のないにもかかわらず、特に雑音下での聞き取りの困難、聴覚弁別の困難(言葉の聞き間違い)、時間処理の困難、両耳処理の困難が伴うものを、聞き取り困難症の概念でとらえる。
◆その背景的な要因として、注意欠陥・多動性障害、学習障害、自閉症スペクトラムなどの発達の問題や認知機能の問題が共存していることが指摘されている。
◆特に認知機能のうち注意(attention)が大きな要素となっている可能性が指摘されている。
◆今回は難聴に関して簡単に触れ、聞き取り困難症/聴覚情報処理障害についてその聞こえの問題点を聴覚的にどのように評価していくのかを考えたい。

1-2-9

1-2-9　子供の気導純音聴力の発達的変化

Developmental changes in air-conduction pure-tone hearing thresholds in children

○倉片憲治(早稲田大・人間科学)

◆既存論文に記載された多数の測定データに基づいて子供の平均的な気導純音聴力を推計し，その発達的変化について考察した。
◆その結果，以下の点が明らかとなった：
(1) 聴力は 10～15 歳にかけて次第に向上し，その後は加齢に伴って低下していく（Fig. 1）。
(2) 性別及び左右耳の違いによる顕著な聴力差は認められない。
(3) 聴力の個人差（標準偏差）は 12～15 歳で最も小さく，周波数が高くなるほど，その年齢は低下する傾向にある。
◆これらのデータ及び検討結果は，ISO 7029（聴覚閾値の年齢別・男女別統計的分布）の次期改訂版において，対象年齢を 18 歳未満に拡張するための基礎資料として有効に活用できると考えられる。

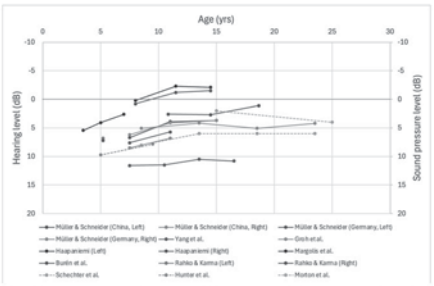


Fig. 1: Auditory thresholds at 1 kHz reported in the analyzed studies. Solid lines: hearing level (dB); dashed lines: sound pressure level (dB)

1-2-10

1-2-10　音楽練習における強大音暴露による蝸牛シナプス障害の可能性 ―Paired-click 刺激法を用いた聴性脳幹反応による検証―

Potential cochlear synaptopathy induced by high-level sound exposure during musical practice: Evidence from auditory brainstem responses using a paired-click paradigm

○藤平晴奈，東莉奈子(九州大学)，山岸慎平(NTT CS 研)，古川茂人(静岡社会健康医学大院大・静岡県立総合病院・NTT CS 研)，柏野牧夫(NTT CS 研)，森周司(九州大学)

◆蝸牛シナプス障害（cochlear synaptopathy: CS）は，内有毛細胞と蝸牛神経との間のシナプスに生じる障害である.
◆CS は，強大音暴露や加齢によって生じる蝸牛における最初の障害であることが示されている.
◆我々はヒトにおいて CS を捉える手法として，paired-click 刺激に対する聴性脳幹反応（auditory brainstem response: ABR）を提案した (Fujihira et al., 2024).
◆本研究では，paired-click 刺激に対する ABR を指標として，強大音暴露による CS が検出可能かどうかを検証した.
◆実験では，音楽練習を頻繁に行っている群（暴露群）と音楽経験のない群（対照群）において，paired-click 刺激に対する ABR を比較した.
◆その結果，聴力レベルや歪成分耳音響放射では群間に差が見られなかったが，paired-click 刺激に対する ABR において暴露群が対照群に比べて有意に低い値を示した ($p = 0.009$).
◆Paired-click 刺激に対する ABR は，ヒトにおいて CS を評価する有用な指標となる可能性がある.

1-3-1

1-3-1　空間正則化付き ILRMA と雑音事前分布ありランク制約付き空間共分散行列推定法を用いたドローン搭載マイクロホンアレーによる音声抽出

Speech extraction using microphone array on UAV by spatially regularized ILRMA and rank-constrained spatial covariance matrix estimation with noise prior

☆錦織 広尚，高宗 典玄，猿渡 洋(東大)，△Jin Xuan Teh(The University of Auckland)，△Benjamin Yen(科学大)，日岡 裕輔(The University of Auckland)

◆近年，ドローンに搭載したマイクロホンアレーによる音声抽出技術が災害救助等の場面において注目されており，空間正則化付き ILRMA (SR-ILRMA) および RCSCME に基づく手法を適用した研究がある.
◆本稿では RCSCME に対して，事前に計測したエゴノイズを活用して設計した雑音事前分布を導入し，観測信号の生成モデルを多変量複素一般化 Gaussian 分布に拡張した手法を適用する.
◆シミュレーション実験により，入力 SNR が −20 dB 以下の過酷な環境における提案手法の有効性を確認した.

Fig.1：UAV (drone) system with seven microphones. This picture was provided by University of Auckland, CAL.

1-3-2

1-3-2　コンテンツ制作に向けた残響制御を可能にする生成的音声強調

Generative Speech restoration with Reverberation Control for Content Creation

◎中田　亘(東大/Google), 小泉　悠馬, 苅田　成樹, Scheibler　Robin, 石川　治子, Guevara-Rukoz　Adriana, 全　炳河, Bacchiani　Michiel (Google)

残響のある音声って
空間を感じられて

僕は好きです。

1-3-3

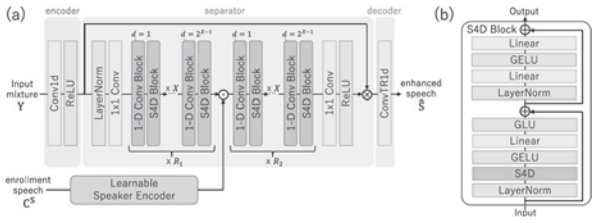
1-3-3　SpeakerBeam-SS: 状態空間モデルを用いた軽量な目的話者抽出

SpeakerBeam-SS: Lightweight Target Speaker Extraction with State Space Modeling

◎佐藤宏, 森谷崇史, 三村正人, 堀口翔太, 落合翼, 芦原孝典, 安藤厚志, 品山健太朗, デルクロアマーク (NTT)

◆概要

・リアルタイム目的話者抽出を実現するには低演算量化が不可欠
　➢ 実用上 CPU 1 コアで演算時間の実時間比を 1 に収めたい
・モデル軽量化と性能の両立を目的として、長コンテキストのモデル化において有効性が示されている状態空間モデル (S4D)を導入



◆評価実験 (Average Precision [%])

・性能を維持しつつ実時間比 (RTF) 78%の削減に成功

System	#Block repetition X	Complexity #param ↓	Complexity RTF ↓	Complexity Latency [ms] ↓	SE task SDR [dB]↑	SE task DNSMOS OVRL↑
Mixture	-	-	-	-	-0.45	2.44
Baseline-noncausal	8	8.69 M	-	∞	13.63	3.40
(a) Baseline	8	8.69 M	1.67	1.25	11.10	2.82
(b) + modify 1D Conv param.	8	10.91 M	0.54	20	11.41	2.91
(c) + S4D (SpeakerBeam-SS)	2	7.93M	**0.36**	20	**11.58**	**2.95**

1-3-4

1-3-4　GRU-Based Low-Latency Noise Reduction for Indoor Hearing Induction Loop Systems

☆Ruiteng LI(Waseda Univ.), ΔKanyu CHEN(Keio Univ.), Yasuhiro OIKAWA(Waseda Univ.)

◆Background

Hearing-induction-loop systems (HILS) send audio as a magnetic field to hearing-aid telecoils, but real signals are influenced by coil non-linearities, electromagnetic hum, and HVAC noise.

◆Methods

We captured 3 h of telecoil speech in a 3.5 m × 3.5 m multiloop room and retrained four enhancers—STFT spectral subtraction, Wiener filtering, Conv-TasNet, DTLN, and an 18 k-parameter bidirectional GRU.

◆Results

Running 8 × faster than real time (RTF = 0.12), the GRU lifts PESQ from 1.56 to 2.68, SDR from 0.8 dB to 7.6 dB, and STOI from 0.74 to 0.91—outperforming classical filters and coming close to the heavier DTLN.

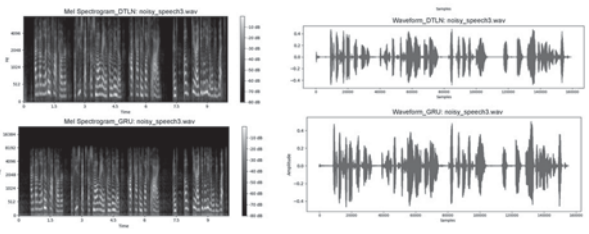


Fig.1 : Waveform and log-mel spectrogram of denoising methods.

1-3-5

1-3-5　可変ステップサイズと並列 ALE による軟骨伝導音声強調

Cartilage Conduction Speech Enhancement using Variable Step Size and Parallel ALE

☆植田英亮, 下倉良太, 飯國洋二(大阪大院・基礎工学研), 笹岡直人(鳥取大・工)

◆キーワード　軟骨伝導, ALE, 可変ステップサイズ, 並列処理, 音声強調

◆あらまし

近年, 周囲の人の話し声や警報音等の, 日常生活に必要な音を認識しながら利用できるオープンイヤー型の音響機器が普及しており, そのひとつである「軟骨伝導」を利用したデバイスは Fig.1 に示すように耳を塞がない構造である. このような構造上, 耳近傍に設置されたマイクでは音声と騒音が混在した信号が集音されるため, この信号を補聴器のようにそのまま増幅するだけでは, 音声と騒音のどちらも増幅するため, 信号対雑音比（SN 比）は原理的に改善されない.

そこで本研究では, 騒音重畳音声から周期性を有する音声信号を Adaptive Line Enhancer（ALE）により選択的に強調することで, 入力 SN 比が高い白色雑音環境下での SN 比を改善し, 音声の明瞭性向上を図る. また, ALE に可変ステップサイズや, Fig.2 に示すような並列処理を導入することで, 従来の ALE を用いた場合よりも SN 比改善量向上を目指す.

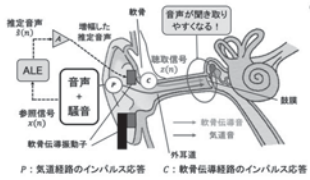


Fig.1 : The mechanism of Cartilage Conduction

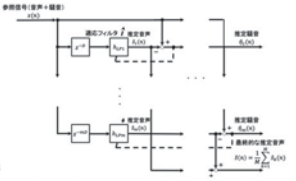


Fig.2 : The block diagram of Parallel ALE

1-3-6

1-3-6　鉄道車両を対象とした空制部品における エア漏れ音検出のための音源分離

Blind source separation for leak sound in air-controlled equipment for railway vehicle

○中山仁史，大島風雅，大村美結（広島市大院），村田剛基（JR 東海）

◆新幹線車両の仕業検査では，外観検査および異音検査が項目数としては多い．異音において最も多く発見されるものはエア漏れである．

◆本研究では低 SNR 環境下を対象とした音源収録と移動体上を観測点とした音源収録を行った．実験対象の編成は同社所属の N700 系とし，空制部品の詳細は保安上の理由により示さないものとする．

◆erSS-NMF では車両所内騒音のアクティベーション行列，模擬エア漏れ音の基底行列及びそのアクティベーション行列を推定する．数値はエア漏れ音の再生機における音量を示しており，最も小さい音を出力する音量 1 のときは SNR が約-25 dB であった．

◆各結果より暗騒音に埋没した模擬エア漏れ音の音源分離が可能なことを確認した．このことから，低 SNR 条件下や移動体上であっても erSS-NMF を用いた音源分離が期待できる．

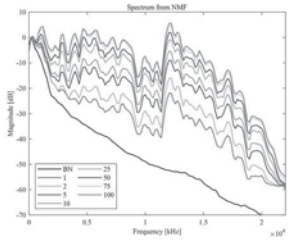


Fig. 1: Leak sound with back ground noise.

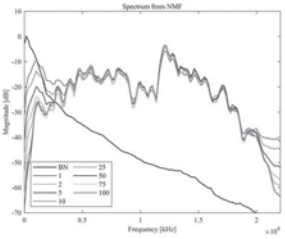


Fig. 2: Estimated by erSS-NMF.

1-3-7

1-3-7　MWSC による 2 マイクロホンでの 3 音源分離

Three Sound Source Separation with 2 microphones using MWSC.

☆多田歩，志賀芳則，△陶山健仁（東京電機大院）

◆音源分離手法として，複数の複素重み付け加算回路を用いた MWSC による指向性形成手法が提案されている。従来の MWSC によって形成される指向性では 2 音源の分離に留まり，より多くの音源信号が到来する際に音源分離を行うことが不可能であった。

◆MWSC は，異なる方向から到来する各音源信号を受音した際に，マイク間に生じる到達時間差をもとに複素重みを算出し指向性を形成する回路である。MWSC によって形成される指向性では一定の区間に対して低感度な抑圧区間を持つという特徴がある。

◆提案法では，到来する音源信号それぞれに対し異なる到達時間差が存在することに着目する。そして，音源方向ごとに各複素重みが独立しているという仮定に基づき，MWSC を改良し 2 マイクロホンによる 3 音源分離を行う。

◆提案する MWSC で形成する指向性を Fig.1 に示す。図に示されるような高指向性を用いて，ある一方向のみの音源信号の抽出を行う。

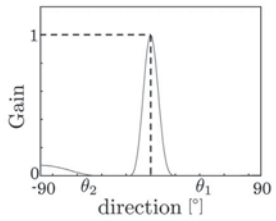


Fig.1 MWSC directionality used for 3-Sound Source Separation

1-3-8

1-3-8　インパルス応答の MTF を向上させる最適化

Optimization for Improving the MTF of Impulse Response

☆奥村純平（大阪大院），下倉良太，飯國洋二

音の反射によって生じる残響は，音声明瞭度を低下させる要因である．従来の MTF-RIR-reshaping 手法では Modulation Transfer Function（MTF）を指標とした残響除去が提案されているが，行列演算を伴うため，高サンプリングレート環境では実用的でなかった．本研究では，MTF の分子項に対する周波数領域での効率的な畳み込み計算を導入することで，この問題を解決した．さらに，勾配情報に基づく最適化手法を適用し，MTF が最大となるようにフィルタ係数を逐次更新する方法を提案する．実験では，従来法では高いサンプリングレート(本論文では 48000[Hz])では動作しなかったが,このようなサンプリングレートにおいても MTF が向上し，最大で 0.0867 の改善が確認された．また，残響時間が 1.8 秒程度のインパルス応答に提案手法の適用実験を行ったところ，残響時間が 0.1 秒ほど短縮されており，高サンプリング環境下でも安定して動作することを示した．

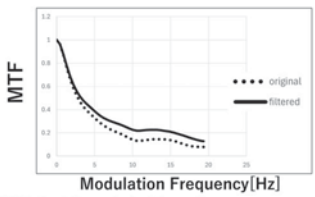


Fig.1 : MTF (0–20 Hz) of the original impulse response and the impulse response filtered by the proposed method

1-3-9

1-3-9

振幅のクリッピング操作により Lipschitz 連続性を保証した DNN

Provably Lipschitz continuous deep neural network via clipping of amplitude spectrogram

◎ 松本和樹，内田蓮，矢田部浩平（農工大）

研究背景：Plug-and-Play (PnP)

アルゴリズム中の近接作用素を DNN で置き換える枠組み

- DNN が Lipschitz 連続でない場合 ▶ **収束性を損なう**
- 音響信号処理では，スペクトログラムの振幅値のみを処理する構造 (Fig.1, 左) がよく用いられる
- ▶ 振幅推定器 $\mathrm{DNN}_{\mathbb{R}_+}$ を Lipschitz 連続なものに制約しても **全体としては Lipschitz 連続でない**

本研究：振幅処理型 DNN に振幅のクリッピング操作を導入

- 提案する構造 (Fig.1, 右) と 1-Lipschitz 連続に制約した $\mathrm{DNN}_{\mathbb{R}_+}$ を用いることで **$\sqrt{2}$-Lipschitz 連続になる**
- クリッピング操作が PnP 雑音・残響除去アルゴリズムの **収束性を改善** することを実験的に確認

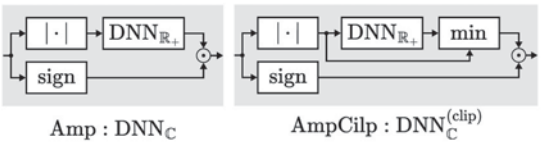


Fig. 1: Conventional (left) and proposed (right) DNN architectures.

1-3-10

1-3-10　液柱型音圧発生装置によるインフラサウンドセンサの一次校正－装置系の高精度化および位相校正の実現－

Primary calibration of infrasound sensors by the liquid-column-type sound pressure generator－device improvement to uncertainty reduction and realization of phase calibration－

☆平野琴（産総研・東北大通研/院情科研）、山田桂輔、高橋弘宜、
△野里英明（産総研）、坂本修一（東北大通研/院情科研）

◆背景：0.5 Hz 以下の周波数域で音圧センサの一次校正を可能とするため、液柱型音圧発生装置を開発中。測定再現性の低さに起因する大きな不確かさと位相校正への未対応が課題。

◆本研究：S/N 改善のため、加振部分を改良し、遮音箱を導入。位相校正も試験的に実施。

◆結果：標準マイクロホン BK4160 の音圧感度を振幅・位相ともに下限 0.01 Hz まで評価。振幅感度の拡張不確かさは改良前の最大 2.6 dB から 0.3 dB まで縮小 (Fig. 1)。0.1 Hz 以上で従来法（レーザピストンホン法）との整合を確認。

◆今後：低周波数域における感度特性の検証、位相の不確かさ評価。

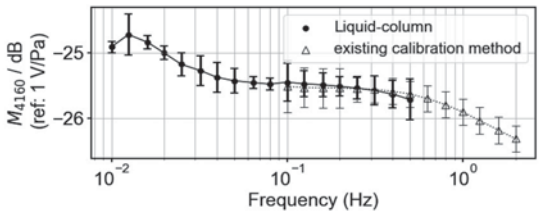


Fig.1：Pressure sensitivity modulus of BK4160. The error bar shows the expanded uncertainty ($k=2$)

1-3-11

1-3-11　ヘッドホンの騒音制御のための等価回路と有限要素法を用いた特性解析

Analysis of a headphone through an equivalent circuit and finite element method for external noise control

☆土岐壮海，鮫島俊哉（九大・芸工）

◆本研究では，**イヤーシェルに有限要素法を導入**した物理モデル（Fig.1）を用いて，ヘッドホンの特性解析及び構造に関する考察を行う。

◆イヤーシェルの有限要素モデルとその他の 1 自由度要素は，それぞれ**等価回路によって表現することで連成解析**を行う。

◆本物理モデルを用いた解析によって，**イヤーパッドの位置による特性の違い**，**イヤーシェルの形状による振動姿態の違い**といった構造に関する考察を行う。これにより，ヘッドホンの動作原理についての理解を深め，更なるモデルの精度向上を目指す。

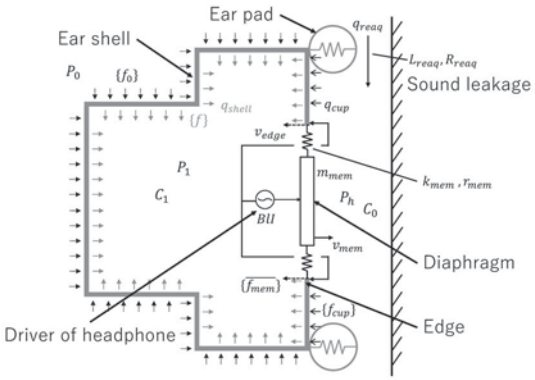


Fig.1：Physical model of a closed headphone introducing finite element method to the ear shell

1-3-12

1-3-12　電力フィードバック進行波型熱音響発電機を四端子回路とみなした発振余裕解析

Oscillation margin analysis of electricity-feedback traveling-wave type thermoacoustic generator as a four-terminal network

○小林泰秀（長岡技科大）

◆最終目的：電力フィードバック進行波型熱音響発電機（Fig.1）の小型化、そのための発振余裕を最大化する帰還回路の設計ツール開発

◆本研究：現有の発電機と与えられた帰還回路に対する発振余裕解析

◆従来手法と課題：(i)ナイキスト安定判別による従来の解析手法…音響計測のコスト大; (ii)発振回路の設計手法…受動部品のみの構成は困難

◆アプローチ：(i)帰還回路からみた発電機全体を四端子回路とみなした周波数応答計測；(ii)切り開かれた回路両端の端子電圧が共通となる解を用いて算出された電力差が正⇒発振、その大きさ＝発振余裕

◆解析結果：(i)発振可否が実験に整合する；(ii)帰還回路にコイルを追加すると発振余裕が大きくなる効果がある …実験及び従来手法に整合

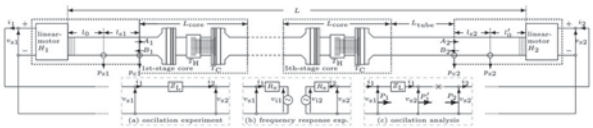


Fig. 1　Experimental apparatus of electricity-feedback thermoacoustic electric generator

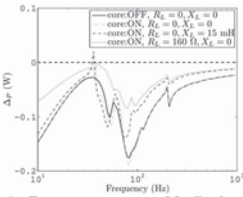


Fig. 2　Frequency response of feedback power

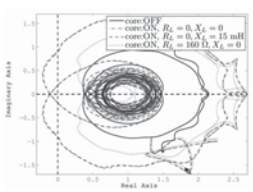


Fig. 4　Nyquist plot by existing method

1-3-13

1-3-13　Temporal super-resolution of optically-measurement sound field based on latent diffusion model

◎Hao Di (Waseda Univ.), Kenji Ishikawa (NTT), Risako Tanigawa (NTT/Waseda Univ.) and Yasuhiro Oikawa (Waseda Univ.)

◆Proposal:

- We propose a latent-diffusion-model-based method to achieve temporal super-resolution of 2D sound field images obtained by optical measurements.

◆Method:

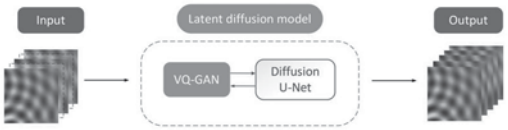


Fig.1：Inference process. The input temporal sound-field data are processed by the latent diffusion model, wherein VQ-GAN as autoencorder transforms the data from the pixel space to the latent space, followed by processing with a diffusion U-Net, and subsequently mapping back to the pixel space. After T time steps, the desired intermediate frames are generated.

◆Results:

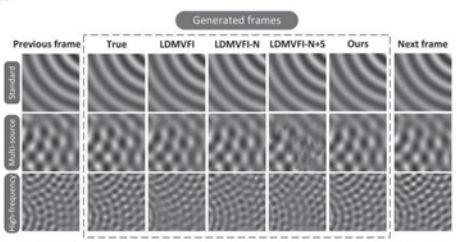


Fig.2：Visual examples of generated images for sound field data by comparison methods and the proposed method. Our method successfully generates intermediate frames without noticeable degradation and demonstrates superior performance compared to the comparison methods.

1-3-14

1-3-14

構造化照明を用いた振動計測のための位相の凸最適化

Convex optimization of phase for vibration measurement using structured light

☆古平真理子，赤石夏輝，矢田部浩平（農工大）

Fourier Transform Profilometry（FTP）

- 構造化照明（縞）を用いた三次元形状計測手法
- 測定対象の凹凸に応じた位相の変化から高さを計算
- 観測画像に縞がない部分があると誤差が生じる

FTP における位相推定を最適化問題として解釈
➡ 観測画像のうち観測したい領域のみに FTP

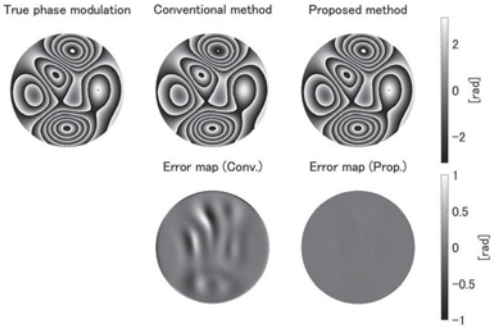


Fig.1 Error between estimated and true phase modulation

1-4-1

1-4-1　ダンパ構造を有する超音波プローブ用ヘッドセットの試作

A prototype of headset-style ultrasound probe holder with damper

○北村 達也（甲南大），能田 由紀子（国語研），正木 信夫，
孫 静（神戸大/甲南大），△岸 竜也(キシテック)

◆超音波画像法を用いた舌運動の計測においては，超音波プローブの安定した固定が極めて重要である．

◆プローブ固定具の世界的デファクトスタンダードである UltraFit は，固定されたプローブが下顎の動きを阻害する可能性がある．

◆これらの問題を軽減することを目指し，プローブ保持部にダンパ構造を持たせたヘッドセットを開発している．

◆本研究では，試作１号 (Kitamura & Nota, 2024) に改良を加えた試作２号を紹介し，予備的な評価実験の結果を示す．

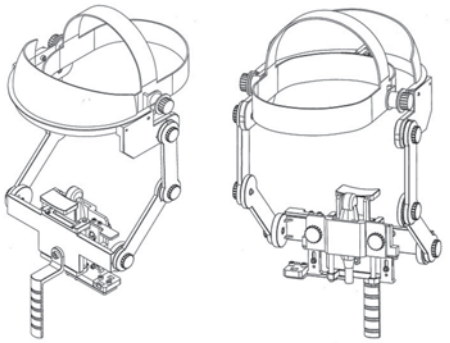


図1 超音波プローブ固定用ヘッドセット試作2号．（左）表，（右）裏．

1-4-2

1-4-2　舌背変位指数（DEI）の周波数解析による発話評価方法の検討

Exploration of a speech evaluation method based on frequency analysis of the Dorsum Excursion Index (DEI)

☆清水 美聖（上智大/JAIST），荒井 隆行（上智大），溝口 愛（前橋工科大），
守本 真帆（中央大），李 瑋昱（上智大），山根 典子（広島大）

◆健常な成人話者７名を対象に，高速で発話した際の舌運動を，超音波機器で観察した．

◆舌形状から舌背変異指数（DEI）を算出し，周期性の時間変化をみるために連続ウェーブレット変換を行った．

◆F02 は一貫して周期性があったが，周期性が一部なくなった話者（F01，M01，M02）や，中盤以降周期性がなかった話者（F03，F04，M03）がいた．

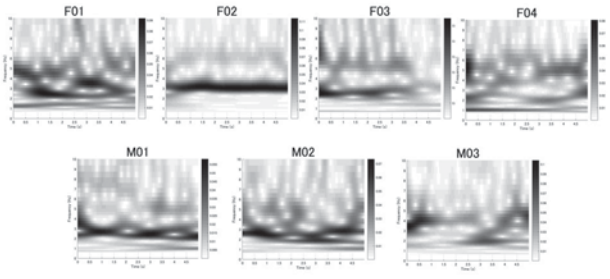


Fig.1 : Time-frequency analysis of Dorsum Excursion Index using the continuous wavelet transform. The colormap shows wavelet coefficient magnitudes, with darker colors indicating stronger signal components in the corresponding time-frequency regions.

1-4-3

1-4-3　上顎洞の開放範囲の違いが鼻音に与える影響

Effects of the extent of maxillary sinus opening on nasal acoustic characteristics

☆海野那月，△岩崎光汰，竹本浩典（千葉工大），
小黒亮史，△大村和弘（慈恵医大），北村達也（甲南大）

◆上顎洞の開放手術では，鉤状突起を切除する場合と，その後方の膜様部まで切除する場合がある。この切除範囲の違いが鼻音にどのような影響を与えるか模擬手術と音響シミュレーションで検討した。

◆健常者 10 名の CT データから鼻副鼻腔形状を抽出したモデル（ORG），鉤状突起を切除したモデル（UP），鉤状突起と膜様部を切除したモデル（MS）を作成し，FDTD 法で伝達関数を求めて音声を生成した。

◆すべての被験者の UP と MS の伝達関数には，800 Hz 付近に大きなピーク，1~3 kHz に大きなディップが生成された。

◆すべての被験者の ORG と UP, ORG と MS の音声は有意に異なって知覚されたが, UP と MS の音声は有意に異なって知覚されなかった。

◆よって，上顎洞の開放手術では，切除範囲によらず音声の知覚に大きな影響を与えるが，切除範囲の違いは知覚に影響を与えないことが示された。

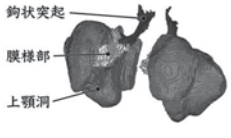


Fig.1 : 上顎洞周辺の構造

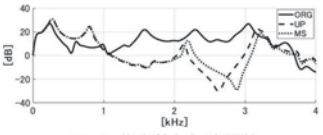


Fig.2 : 代表的な伝達関数

1-4-4

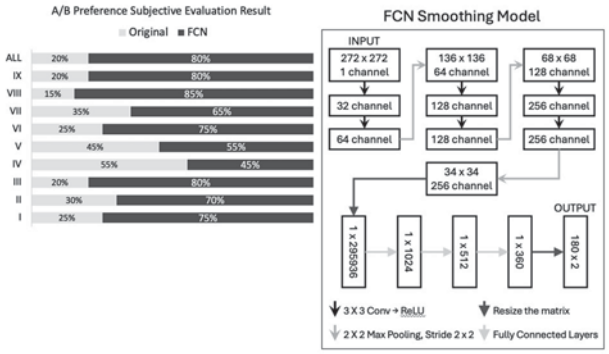
1-4-4 FCN-based Smoothing for rtMRI Landmark-Based Vocal Tract Contour Data

○ Mushaffa Rasyid Ridha, Faisal Mehmood, Sakriani Sakti

◆This paper employs outlier removal and a fully convolutional network (FCN)-based smoothing for the publicly available dataset to refine it as it contains error and inaccuracies.

◆The smoothed result is available publicly: https://github.com/ha3ci-lab/USC-TIMIT_rtMRI_Landmarks

◆Subjective evaluation result: Recruited subjects judged that FCN outputs labels are better than the original labels in almost all specified articulatory areas (except IV). Specifically, for the uvular area (VII), 85% of the subjects chose FCN labels. In conclusion, 80% of the recruited subjects preferred FCN labels for overall areas (ALL).



1-4-5

1-4-5 逆推定に基づく調音運動の評価 −/agV/と/akV/音形における声道体積の特徴解析−

Evaluation of articulatory behavior based on inverse estimation: Feature analysis of vocal tract volume in /agV/ and /akV/ sequences

☆牛島稜太, 緒方公一（熊本大）

◆声道音響管マッピングインタフェースに基づく声道形状逆推定の研究を進めている。先の報告では，有声子音を含む/agV/と無声子音を含む/akV/の発話における咽頭部の声道体積推移を評価してきた。

◆本稿では，子音閉鎖時の体積数値に注目し，各後続母音の/agV/と/akV/の違いをレーダーチャートにより可視化することで分かる特徴について考察する。

◆Fig. 1 に，男性発話者１名における発話の子音閉鎖直前（左）および終了直後（右）における声道の咽頭部体積数値をレーダーチャートで示している。グラフの頂点は各後続母音を配置している。グラフより，いずれの後続母音の発話でも/agV/の方が/akV/よりも体積数値が大きい傾向であることが確認された。

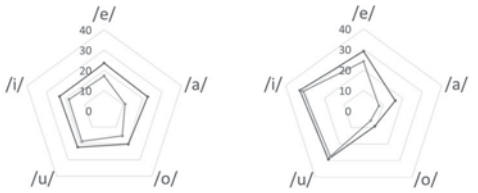


Fig.1 Radar chart for pharyngeal cavity volume immediately before consonant closure (left) and release (right) (Unit cm^3)

1-6-1

1-6-1 粒状炭の吸音特性に関する実験的検討

An Experimental Investigation of the Sound Absorption Characteristics of Granular Charcoal

◎原澤悠, 中島美百合（フジタ）

粒状炭の室内用吸音材への適用可能性を検討することを目的とし、音響管による垂直入射吸音率測定、および残響室法によるランダム入射吸音率測定を実施し、以下の知見を得た。

◆粒状炭の吸音率は、粒径が小さくなるにつれピーク周波数は低域側にシフトし、値が上昇する。また、粒径が極端に小さい場合、材質によっては吸音率が低下する。

◆同粒径においては、粒子形状が不成形である程、吸音率のピーク周波数は低域側にシフトする。

◆ランダム入射条件下では、多孔質型吸音に類似した吸音特性を示す(Fig.1)。また、表面材を付加することで、低域の吸音率が向上し、高域の吸音率が低下する。

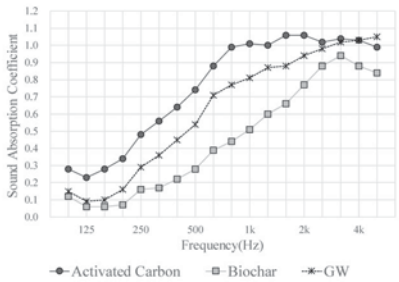


Fig.1 : Measured Sound Absorption Coefficients Using the Reverberation Room Method.

1-6-2

講演取消

1-6-3

1-6-3　伝達マトリクス法による一般化マクスウェルモデルの波動解析と制御

Wave Analysis and Control of the Generalized Maxwell Model Using the Transfer Matrix Method

○沈奥(成蹊大院)，岩本宏之(成蹊大)，久野翔太郎(北九州高専)

本研究では，粘弾性体の振動抑制を目的に，クリープや緩和特性を表現できる一般化マクスウェルモデルを導入し，その波動制御手法を検討した．三要素モデルを用いて伝達関数法と伝達マトリクス法の解析結果を比較し，伝達マトリクス法の有効性を確認した上で，一般化マクスウェルモデルの伝達マトリクスを体系的に導出した．さらに波動解析を行い，外力による応答を正負方向の波に分解し，負方向成分をゼロと仮定することで制御則を構築した．数値解析により，提案手法の振動抑制効果を確認した．

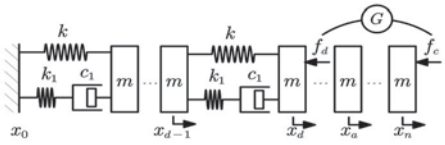


Fig.1 Control law for the Three-element Generalized Maxwell Model

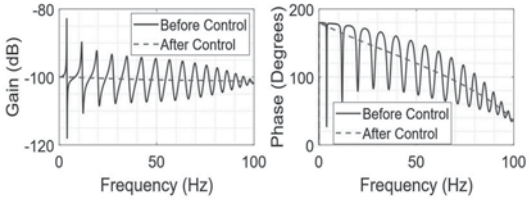


Fig.2 Frequency response comparison before and after control

1-6-4

1-6-4　ノッチ周波数の変化に着目した1 ch マイク音源方向推定法の開発

A single-channel-based estimation of sound source direction using directional characteristics of notch frequency

◎永田瑚詩，後藤達彦，江波戸明彦((株)東芝)

◆音入射方向と行路差によるノッチ周波数との相関性を活用した1 ch マイク音源方向推定方式の提案（Fig. 1, Fig. 2）

◆音声データを用いた簡易評価にて３度未満の推定誤差を実現

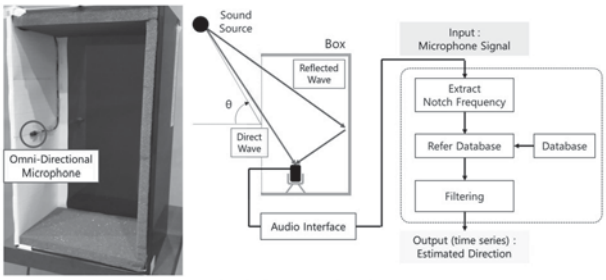


Fig. 1: Proposed system

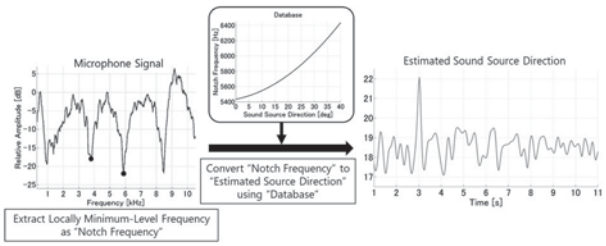


Fig. 2: Estimation method of sound source direction using notch frequency

1-6-5

1-6-5　DSP を活用した軟骨伝導技術による狭帯域ノイズのリアルタイム選択的消音：二次適応バンドパスフィルタを用いたアプローチ

Real-time Selective Noise Cancellation or Narrowband Noise Using Cartilage Conduction and Second-order Adaptive Bandpass Filter

☆加藤良樹，下倉良太(大阪大)，笹岡直人(鳥取大)，飯國洋二(大阪大)

◆オープンイヤーでのノイズキャンセリングを目指しており，オープンイヤーで音出力が可能な軟骨伝導に着目した．

◆音声と狭帯域ノイズが重畳した環境において，二次音源として軟骨伝導から信号を出力することで，鼓膜での狭帯域ノイズの除去を目標としており，本発表では，リアルタイム実験に関して発表する．

◆消音を可能とするFiltered-x least means square (FxLMS) アルゴリズムに，ノイズの周波数を逐次推定する適応バンドパスフィルタを組み込んだABF-FxLMS アルゴリズムを提案した．

◆リアルタイム消音を実装するために，DSP を使用し，ABF-FxLMS のフィルタを二次にして処理負荷を軽減した．

◆提案アルゴリズムを用いた結果，ノイズを狭帯域ノイズ 11dB，消音でき選択的消音性能が向上したことを確認した．

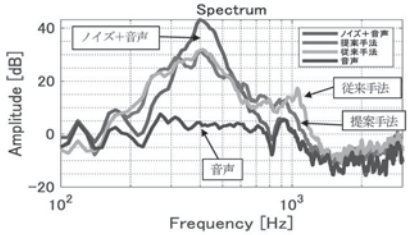


Fig.1：Spectra of four signals (narrowband noise at 400 Hz)

1-6-6

1-6-6　音声に重畳する周期性及び広帯域騒音のための軟骨伝導アクティブノイズコントロール

Cartilage conduction active noise control for periodic and broadband noise superimposed on speech

☆西村悠希，笹岡直人(鳥取大・工)，下倉良太(大阪大・基礎工)

◆耳を塞ぐことなく外耳道内に音を発生させることのできる軟骨伝導にANC を搭載して耳を塞がない消音を目指す。

◆騒音の対象は周期性騒音(サイレン音)と広帯域騒音(白色ノイズ)としており，騒音と音声の分離には 2 種類の適応線スペクトル強調器(ALE : Adaptive Line Enhancer)と，騒音再合成フィルタ(NRF : Noise Reconstruction Filter)を並べたフィルタを使用している(Fig. 1)。また，これらをAAN(ALE1，ALE2 & NRF)と呼称する。

◆x(n)は騒音重畳音声，s(n)は音声，$\eta(n)$は周期性騒音(サイレン音)，$\xi(n)$は広帯域騒音(白色ノイズ)を示している。

◆AAN で分離された推定騒音を利用し，軟骨伝導を適用したFxLMS アルゴリズムを使用して騒音低減のシミュレーションを行った。

◆結果として，周期性騒音の低減は確認でき，広帯域騒音はわずかに低減された。

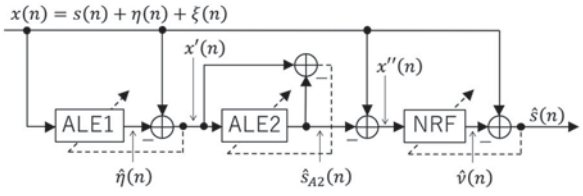


Fig. 1：Block diagram of AAN

1-6-7

1-6-7　室内低周波騒音のアクティブ制御

—小型加振器で振動させた壁面を制御音源として用いた検討—

Active control of indoor low-frequency noise:
Study using a wall vibrated by an inertial exciter as a control sound source

◎岩永景一郎，土肥哲也（小林理研）

◆低周波騒音 100 Hz を室内でアクティブ制御（Fig.1, 2）

◆加振器で内壁を振動させて制御スピーカとして利用（Fig.2）

◆制御点（頭部想定）周辺 0.8 m×0.8 m に低減効果，最大 11 dB（Fig.3）

◆低周波騒音が気になり眠れない場合などへの適用が期待される

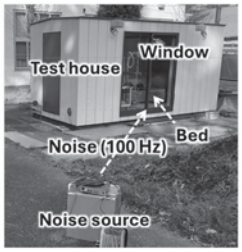


Fig.1 Outside the house
(100 Hz noise enters the house)

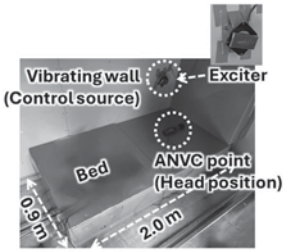
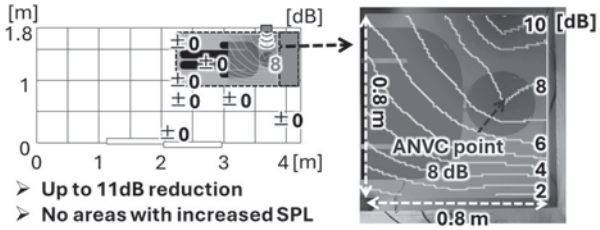


Fig.2 Inside the house
(ANVC by vibrating wall)



➢ Up to 11dB reduction
➢ No areas with increased SPL

Fig.3 Reduction effect on the bed and in the house

1-6-8

1-6-8　風車騒音を対象とした深層学習による音源分類の検討

Sound source classification using deep learning:
focusing on wind turbine noise

☆中村壮志，田鎖順太（北大院・工）

◆低コストでの風車騒音の長期モニタリング実現に向けて，畳み込みニューラルネットワークモデルによる音源分類を試みた。

◆音源分類の正解率は，3 秒間を対象とした場合は約 70%，10 分間を対象とした場合は約 80%であった。

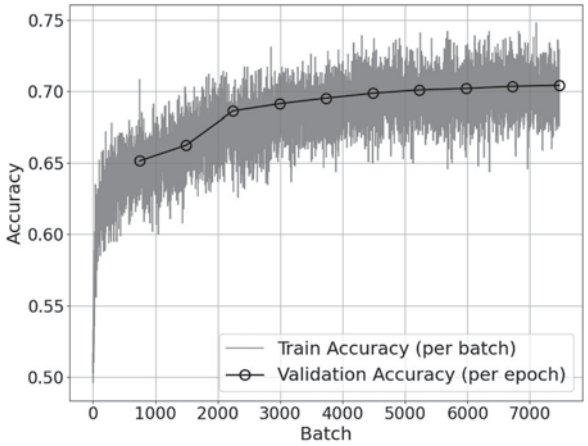


Fig. 1: Accuracy of 3-seconds sound source classification using the convolutional neural network model during training.

1-6-9

1-6-9　自動音声認識を用いたオフィスのスピーチプライバシー評価の試み

An attempt to evaluate speech privacy
in workplace using automatic speech recognition

◎丸山直也，Ainun Nadiroh，丸山弥優，川井敬二（熊本大院）

◆オープンプランオフィスのスピーチプライバシーを定量的かつ経時的に評価することを目的として，自動音声認識（ASR）の平均対数尤度（ALP）を評価に用いることを試みた。

◆音素バランス文を対象として SNR と残響時間を変化させた音源を対象に予備検討を行い，ALP の閾値を-0.6 と決定した。（Fig. 1 に SNR の例を示す）

◆実際のオフィスでの執務中の環境音を用いて、音声の漏洩の推定を試みた結果、ミーティングスペースの近くで多く検出されるなど、実際の環境に応じた結果が推定された。

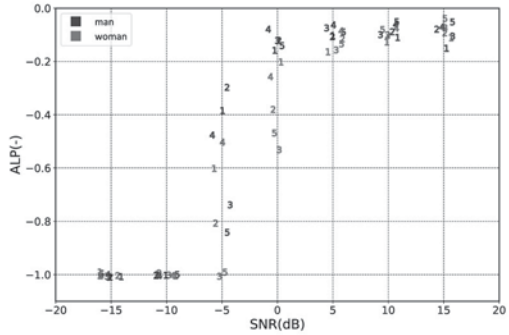


Fig.1 : SNR (Signal-Noise ratio) v.s. ALP (Average log-probability) of ASR

1-6-10

1-6-10　マイクロホンアレイと環境音による室内状態変化検知の試み

－ 検知手法の比較および異常特定要因の分析－

An attempt to detect indoor abnormal states
using environmental sounds and a microphone array.
- detection methods comparison and anomaly identification factors analysis -

○池野修一，△池田賢二（セコム・IS 研）

◆「環境音を用いた聴覚による障害物知覚」から着想を得た，人物侵入や扉・窓開放などの室内状態変化検知

◆マイクロホンアレイで観測した環境音から得られる空間スペクトルを画像化し、そこに表現される室内状態変化に伴う微小変化を外観検査向け異常検知手法により検知するアプローチ (Fig. 1)

➢ 収集容易な無人時データのみからモデル構築可能な学習済深層学習モデルを特徴量抽出器として用いる距離ベース手法に着目

◆異常検知手法として PatchCore と PaDiM を検知精度 / 異常要因方向の表現力 (anomaly map 上での異常特定部位の明瞭性) の観点で比較

➢ 検知精度: ほぼ同等 (AUROC: 0.848～0.993)
異常要因方向の表現力: PaDiM が優れる (Fig. 2)

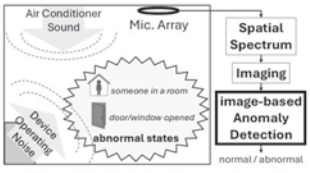


Fig.1 Overview of the anomaly detection

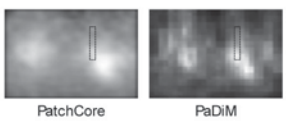


Fig.2 Samples of anomaly maps

1-6-11

1-6-11　マルチセンサシステムによる独居高齢者の生活様式変化検知の評価

Evaluation of changes in the lifestyles of elderly people living alone using a multi-sensor system

○韓 毅弘，坂口晃啓(FEAT)，△鈴木慎之介(IC)，○有光哲彦(FEAT)

◆**研究目的**: 独居高齢者の生活習慣の変化をマルチセンサーシステムを用いて定量的に評価し、生活の質（QOL）向上を目指す。

◆**データ処理手法**: センサーデータを正規化し、時間的パターン分析、空間的エントロピー計算、軌跡長計算を組み合わせた特徴量を抽出して統合的に評価する。

◆**オンラインクラスタリング手法**: 日々のデータを既存クラスターとの類似度で判定し、閾値を超える場合は新クラスター作成。品質指標 Q で閾値を動的調整する。

◆**実証実験**: 8 人の被験者で実験を行い、生活様式を認識・推定できることを確認(Fig.1)。特に生活様式変化の検出に有効性を示した(Fig.2)。

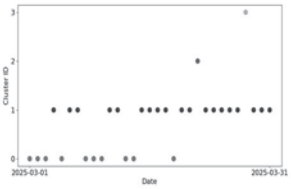


Fig.1 :Lifestyles clustering results of one of the participants in March 2025.

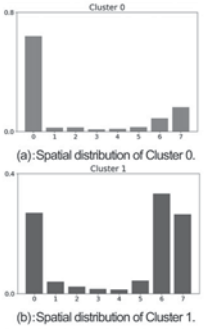


Fig 2: Spatial distribution between Cluster 0 and Cluster 1.

1-6-12

1-6-12　MIMII-Gen: 異常音検知ｱﾙｺﾞﾘｽﾞﾑの性能を評価するための生成ﾓﾃﾞﾙﾍﾞｰｽｱﾌﾟﾛｰﾁ

Generative Modeling Approach for Simulated Evaluation of Anomalous Sound Detection　System

○遠藤隆，△プロヒト・ハーシュ，西田智哉，土肥宏太，川口洋平(日立)

異常音検知（Anomalous Sound Detection, ASD）は、産業機械の予知保全において重要な技術である。しかし実環境での異常音の取得は困難であり、ASD モデルの性能評価が制限される。本研究では、実環境に近い異常機械音を条件付きで生成する手法「MIMIIGen」を提案する。本手法は、キャプション生成と拡散モデルを用いて、高品質かつ多様な異常音を生成することで、ASD システムの汎用性と頑健性を異常音を収集することなく検証可能にした。生成された異常音を用いた ASD ベースラインの AUC の推定では、実データを用いた場合との差が平均 4.8%に収まることが確認された。

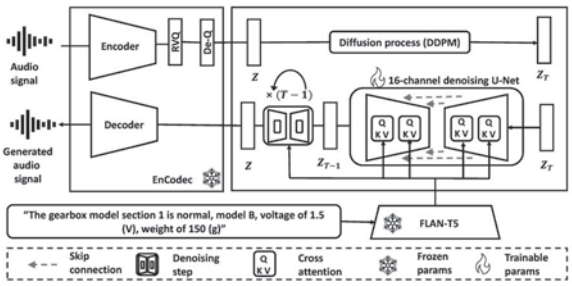


Fig.1 Block diagram of proposed approach for machine sound generation

1-6-13

1-6-13　Physics-informed Neural Operator に基づく 2 次元音響解析

Two-dimensional acoustic analysis based on a physics-informed neural operator

◎横田和哉，△坂爪孝熙，△塩田朔矢，△小倉正崇（長岡技術科学大学）

◆機械学習を用いた数値シミュレーションの代替モデル（サロゲートモデル）に関する研究が近年活発化している。

◆Physics-informed Neural Operator は、ニューラルネットワークを用いた数値シミュレーション手法である。損失関数に支配方程式に関する拘束を導入することで、事前のデータ準備を不要としている。

◆本研究では、時間領域での 2 次元音波伝搬シミュレーションを行う Neural Operator を構築した。

◆提案法は FDTD 法と同等の数値解析結果を高速に推論可能であった。

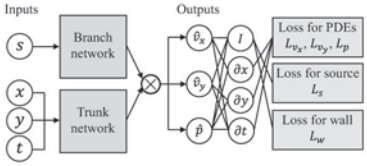


Fig.1： Proposed Physics-informed Neural Operator.

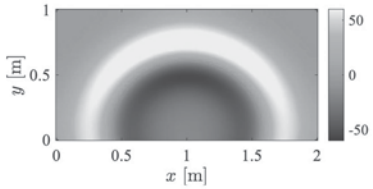


Fig. 2： Analyzed sound pressure using proposed neural operator.

1-6-14

1-6-14　音響イベント認識と音源方向推定に基づく PTZ カメラ制御技術の開発 －研究概要と先行研究レビュー－

Development of PTZ camera control technology based on acoustic event recognition and sound source direction estimation – Study overview and review of previous studies –

○森淳一(神奈川大)，石井要次(積水化学)，森長誠(大同大)

◆本研究は、人工知能に基づく音響イベント認識、到来時間差（TDOA）による音源方向推定、PTZ カメラ制御を統合し、突発的な異常音に即応可能な都市監視システムの構築を目指すものである。

◆映像監視の限界を補完し、リアルタイムかつ分散型のエッジ処理により、実装性と実用性の両立を図る点が特徴で、今後の IoT 監視基盤としての発展が期待される。

◆今後は、実環境における応答性能や検出精度の検証も進めていく。

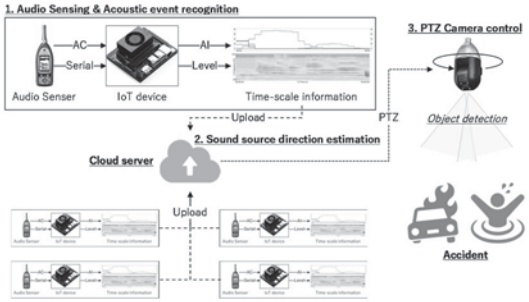


Fig.1:System diagram of surveillance system based on acoustic event recognition

1-6-15

1-6-15　音響イベント認識と音源方向推定に基づくPTZ カメラ制御技術の開発 －屋内外試験による音源方向推定とAI 識別精度の検証－

Development of PTZ camera control technology based on acoustic event recognition and sound source direction estimation – Verification of sound source direction estimation and AI identification accuracy in indoor and outdoor test –

○石井要次，東総介（積水化学），髙橋莉紗（Sound One），渡部光（小野測器），森長誠（大同大），風間亮介（リオン），山内勝也（九州大），森淳一（神奈川大）

◆著者らは，都市空間におけるリアルタイム監視システムの実現に向けたシステムを提案している。本報では，事故等を想定した衝撃性の音源に対し，TDOA(Time Difference of Arrival)による音源位置推定，PANNs(Pretrained Audio Neural Networks, 527 category)による音響イベント認識に関する基礎実験を実施。

◆無響室実験での結果は以下の通り：

・音源位置推定(Fig.1)：(a)音源がマイクロホンアレイ内にある場合は十分な精度だったが，(b)マイクロホンアレイ外の場合は推定誤差が顕著に増大。

・音響イベント認識：暗騒音がない場合は“爆発”などの事故に近いカテゴリが推定されたが，暗騒音(Urban Noise)がある場合はその影響が支配的

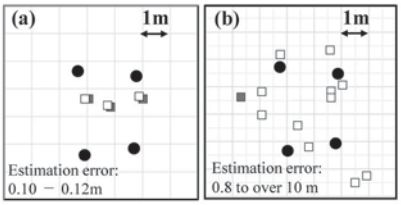


Fig.1：Results of TDOA applied to the crash sound (car) under background noise. (●Mic.，■：Actual sound source，□：Estimated sound source)

1-6-16

1-6-16　音響イベント認識と音源方向推定に基づくPTZ カメラ制御技術の開発 -PTZ カメラ制御の試み-

◎東総介，石井要次（積水化学），山内勝也（九州大），森淳一（神奈川）

1. はじめに

本研究は，街頭の PTZ（Pan-Tilt-Zoom）カメラに音響情報を連携させ，画角外で発生した突発事象を撮影できる自動追従システムの基礎技術を検討したものである。

2. PTZ カメラ制御のための技術検討

まず、RTSP／ONVIF プロトコルにより市販 PTZ カメラを Python で制御可能な基盤を構築した。次に MobileNet-SSD による人物検出を用いて、対象が常に画面中央に収まるような自動トラッキングを実現し、PTZ カメラの自動制御を検証した。さらに、将来的な音源発生位置の受け渡しとして地理情報を活用した方法を想定し、PTZ カメラ・ターゲット双方に GPS を設置し、球面座標間の方位角を活用し PTZ カメラの自動制御を検証した。上記取り組みより音響イベントに基づく PTZ カメラ自動制御技術に向けた基礎的な技術検証を実施する事ができた。

Fig.1：The target was brought into the camera's field of view using feedback control based on the GPS heading angle

1-6-17

1-6-17　航空機騒音モニタリングにおけるAI 活用：航空機騒音と妨害音の音源分離の試み

A study on sound source separation between aircraft noise and disturbance sounds

◎中山紬，佐藤僚，昼間信彦，△江副泰亮（リオン），横田考俊，△甲田俊介（小林理研）

◆航空機騒音モニタリングへのAI 活用を検討
- ➢ これまでにCNN による妨害音有無の自動判定手法を開発
- ➢ 妨害音が評価に与える影響を定量的に推定することは依然として困難

◆本研究では，深層学習による音源分離技術を適用し，航空機騒音と妨害音の分離を試みた

◆Conv-TasNet をアーキテクチャとして採用し，航空機騒音と２種類の妨害音（建設作業に伴う金属音・道路交通騒音）との分離を行った (Fig.1)

◆妨害音が金属音の場合は高い分離精度を示したが，道路交通騒音は航空機騒音と音源間の特徴が類似しているため低下した (Table.1)

◆今後は，マスク推定の精度向上や分離後信号による定量的評価の有効性への検討が求められる

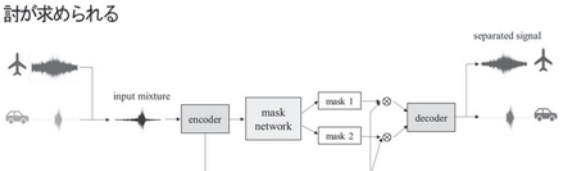


Fig.1：Overview of sound source separation

Table 1: Results (partial). The left column shows the case where the disturbance sound is metal, and the right column shows the case where the disturbance sound is road traffic noise.

SNR	$\Delta L_{EA,aircraft}$	$\Delta L_{EA,metal}$	SNR	SNR	$\Delta L_{EA,aircraft}$	$\Delta L_{EA,road}$	SNR
-5	-0.15	-0.32	-4.83	-5	-1.96	-1.44	-5.52
-2	-0.17	-0.35	-1.81	-2	-1.69	-1.57	-2.13
0	-0.26	-0.44	0.17	0	-1.31	-1.73	0.42
2	-0.21	-0.48	2.28	2	-1.47	-1.60	2.13
5	-0.22	-0.58	5.20	5	-1.50	-1.38	4.64
10	-0.30	-0.75	10.45	10	-1.06	-1.33	10.27

$\Delta L_{EA} = \hat{L}_{EA} - L_{EA}$
$\hat{L}_{EA}$：L_{EA} of the separated signal
SNR：The SNR calculated from the L_{EA} of the separated aircraft noise and disturbance sounds

1-6-18

道路交通騒音における単体車両交通騒音の抽出*

○鄭　萬溶（沼津高専）

1　はじめに

本研究では、道路交通騒音の実態把握を目的とした交通モニタリングシステムの開発を目指し、その要素技術として深層学習を用いた音源分離と特徴抽出について検討を行う。特に、単体車両騒音の分離性能向上、パワートレイン系騒音とタイヤ路面騒音の分離、および環境条件に依存しない交通騒音の特徴抽出に焦点を当てる。

2　深層距離学習による交通騒音の潜在表現学習

カメラベースの交通監視システムが低照度環境下での観測困難や季節による観測可能時間の変動といった課題を抱える中、音響情報を用いた交通監視システムは、低コスト、電力効率、設置の容易さ、悪天候や視界不良に対する堅牢性などの点で有利である。しかし、交通騒音はノイズが多く、気圧・気温・湿度・マイク位置などの環境条件に大きく依存するため、特徴抽出が困難である。本研究では、このような環境条件に依存しない特徴抽出手法として、深層距離学習を用いた潜在空間の取得を提案している。深層距離学習は、高次元の入力データを低次元の潜在空間へと圧縮し、入力空間で類似しているデータ同士が潜在空間上でも近接するようにマッピングする手法である。

2.1 Circle Loss による潜在空間の取得

離散的な分類ラベル（車種・進行方向）に対しては、Circle Loss を採用した。Circle Loss は、クラス内類似度を最大化し、クラス間類似度を最小化するように損失を与える手法である。DCASE Challenge 2024 で提供されたシミュレートされた車両走行音データを用いて、車種（car, cv）と進行方向（right, left）による 4 種類のラベルを対象に実験を行った。

2.2 Log-ratio Loss による潜在空間の取得

Log-ratio Loss を用いて可視化された潜在空間では、全体として１つのクラスタを形成し、左から右にかけて速度ラベルが大きくなるよう分布していることが確認された。

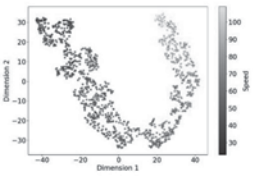


Fig. 7 Latent Space Visualization Using Log-ratio Loss

2.3 速度予測モデルの構築

深層距離学習で得られた速度の潜在空間に DNN を接続し、速度予測モデルを構築した。事前学習された CNN エンコーダを用いることで、学習に使用していないデータに対して平均 ±5.4 km/h の誤差で速度を予測できた。これは、非接触かつ音響情報のみを用いるという制約下では実用的な範囲と考えられ、従来のより大きな誤差が生じる手法と比較しても安定した性能を示している。

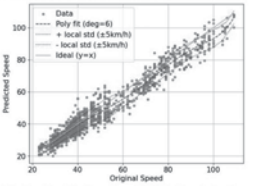


Fig. 8　Predicted vs. True Vehicle Speeds by the Proposed Speed Estimation Model

3　結論

本研究では、深層距離学習を用いた特徴抽出では、Circle Loss により車種や進行方向ごとの潜在空間での分離に成功し、Log-ratio Loss により速度情報を保持した潜在空間の取得が可能であることを示した。さらに、これらの特徴量を用いた速度予測モデルは、実用的な精度を達成した。

1-6-19

1-6-19　点群データおよび Segment Anything Model を用いた騒音予測のための障害物推定手法の開発

Estimation of Obstacles for Sound level Prediction Using Point Cloud Data and Segment Anything Model

○田鎖順太(北大院・工)

◆正確な騒音予測のために，正確な障害物（建物）の位置や形状の把握が重要である。既存の建物形状データは誤りも多い。

◆3 次元点群データおよび Visual Transformer である Segment Anything Model を利用することによって，建物形状が高精度に示された。

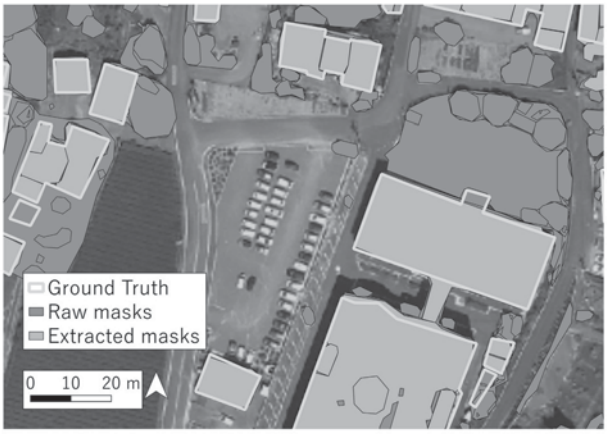


Fig. 1: Example of the detection of the building masks using Segment Anything Model

1-7-1

1-7-1　球バッフルマイクを用いた鋭指向性実現の試み

Attempt to realize narrow directivity with a single spherical baffle microphone

☆中島佑樹，岩見貴弘，尾本章(九大・芸工)，今西勝之，石井健史((株)SOKEN)

◆球状マイクアレイに配置できるマイク数の物理的制約による球面調和展開次数の制限や，複数のアレイを用いたサンプリング手法の実測比較におけるコスト等の問題を解決するため，任意のマイク数 / 配置に対応した逐次測定を可能とする球バッフルマイクを構築

◆IR の逐次測定と球面調和領域におけるビームフォーミングの適用

◆異なるサンプリング手法に基づくアレイの性能比較から，理論的に優位な Fibonacci spiral が，実測においても高い周波数帯域まで鋭い指向性を実現可能であることを確認

◆サンプリング手法によらず，ほぼ全帯域で鋭指向性マイクロホン(DPA 4017C)より鋭い指向性を実現

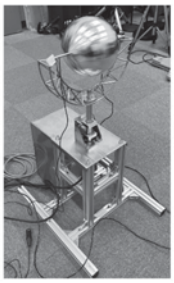


Fig.1：Spherical baffle microphone

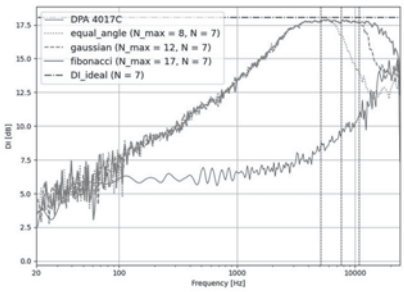


Fig.2：Directivity Index for sampling methods. Vertical dashed line : the upper frequency limit for noticeable spatial aliasing

1-7-2

1-7-2　空間シフトフィルタを用いた直線スピーカアレイ波面合成法の時間領域表現の検討

Study on time-domain representation of wave field synthesis based on spatial shifting filter using linear loudspeaker array

☆五島 優太，羽田 陽一(電通大)

◆ 直線スピーカアレイを用いた音場再生の需要が高まっている

◆ スピーカアレイを跨いで移動する仮想音源再生や，音圧の高い領域を制限するエリア再生は，受聴位置での目的音場が重要

◆ さらに，人の動作で音場を変化させるインタラクティブオーディオでは時変音場を再現する駆動信号のリアルタイム計算が必要

◆ 受聴位置で目的音場が設計されたインタラクティブオーディオへの拡張を目的に、空間シフトフィルタを用いた WFS の時間領域表現を導出

◆ 従来手法と仮想音源の精度を計算機シミュレーションで比較

- 再現精度と再現領域の広さのトレードを制御可能
- 従来手法より空間エイリアシングを低減

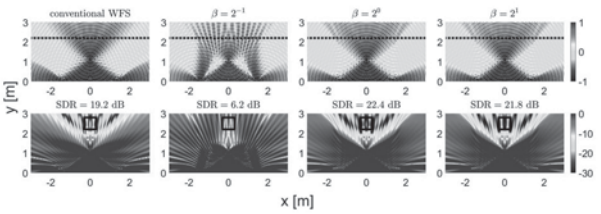


Fig.1：Wavefront (upper) and mean squared error (bottom) at 2000 Hz for different parameters (β indicates the resolution of the desired sound field)

1-7-3

1-7-3　球面および直線マイクロホンアレイでの実測 IR に基づく HOA 音場再現における遠方距離感呈示の検討

A study on reproducing distant virtual sound source perception in HOA sound fields using measured IR from spherical and linear microphone arrays

☆赤井たろ生，羽田陽一(電通大)

◆モデルベースの HOA によって実測ベースの HOA と同等の距離感呈示を行うための残響音を含む点音源レンダリング手法を検討した。

◆球面マイクロホンアレイの測定ベースの NFC-HOA(Sph_RIR)と直接音と初期反射音をそれぞれ NFC-HOA と HOA 仮想音源フィルタで生成し、後部残響音を指数減衰するホワイトノイズで模擬する提案手法(Pro_HOA)と Pro_HOA の直接音および初期反射音を単一のスピーカで簡易に再現する手法(Pro_single)を用いて半径 1.125m の 42ch 球面スピーカアレイで 1m, 2m, 4m に仮想音源を生成し、受聴して残響感および距離感を評価した。

◆残響感についてはいずれの音源種類及び距離においても有意な差は見られなかった。距離感の結果は Fig. 1 に示す。Pro_HOA で Sph_RIR と同等の距離感を呈示可能であることが示された。

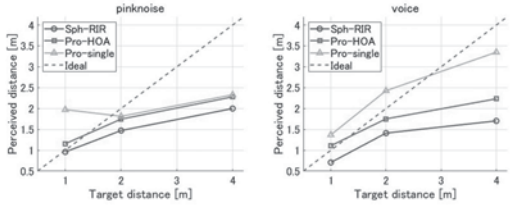


Fig.1：Median perceived auditory distance.

1-7-4

1-7-4　小型剛球マイクロホンの BoSC システムへの適用可能検討
—音像定位実験による評価—

Investigation of the applicability of a small rigid microphone to the BoSC system -Evaluation by sound image localization experiment-

☆宮原朋哉，伊勢史郎（東京電機大）

◆従来の BoSC システムでは収録系として半径 0.23 m の中空球型マイクロホンを用いてきたが、その半分の半径である小型剛球マイクロホンを用いた際の有効性がシミュレーションにより示唆されている。

◆小型剛球マイクロホンを用いて測定したインパルス応答から逆システムを設計し、その精度を従来のマイクロホンと比較した。

◆中空球型マイクロホンを用いて設計した逆システムとインパルス応答を畳み込んだ波形の周波数特性の総和を示した図を Fig.1 に小型剛球マイクロホンを用いた同様の図を Fig.2 に示す。

◆中空球型マイクロホンの課題となっていた特定の周波数でのディップが効果的に抑制されることを確認。

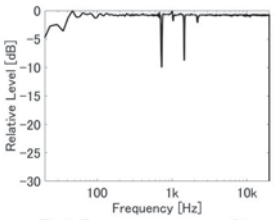


Fig.1：Frequency response of the convolved inverse system in the conventional BoSC mic.

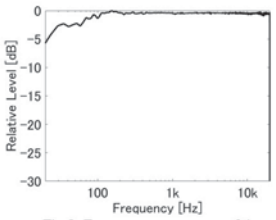


Fig.2：Frequency response of the convolved inverse system in the new BoSC mic.

1-7-5

1-7-5　コンタクトマイクロホンを活用した放射特性の取得の検討

Estimation of radiation characteristics using contact microphone

○木下光太郎，富塚ゆみ子，松井健太郎（NHK）

◆被写体を取り囲むように配置した複数台のカメラの映像から 3 次元モデルを作り出す**ボリュメトリックキャプチャ**の研究・開発が盛んとなっている。

◆ボリュメトリックキャプチャを活用した任意視点映像に対応する音声として**ボリュメトリック音声制作**を提案しており，映像と同時に音声信号および放射特性を取得できる環境を構築した。

◆ただし，放射特性は収録距離に依存して，音場再現の精度が保たれる範囲は狭まり，特に音源近傍の位置では適切にレンダリング信号が生成できない可能性がある。

◆本稿では，音源近傍に設置したコンタクトマイクロホンを活用して放射特性を取得することで，音源近傍でも適切なレンダリング信号が生成できるようになるか検証を行った。

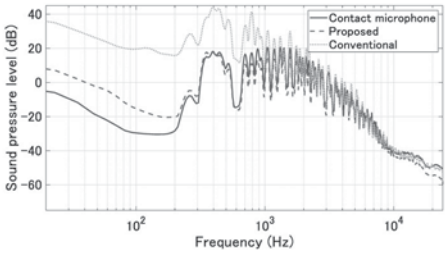


Fig.1 Spectrum of the signals in the frontal direction

1-7-6

1-7-6　ボリュメトリック音声制作に向けたオフセンター音源の放射特性取得法の検討

A study on the estimation method of radiation characteristics of an off-center sound source for volumetric audio production.

◎富塚ゆみ子，木下光太郎，松井健太郎，杉本岳大，中山靖茂（NHK）

◆任意視点映像に対応する音場を表現するため，音源を取り囲む球状マイクロホンアレイを用いて観測した音圧から**音源の放射特性**を推定・取得することを検討している。

◆音源がマイクロホンアレイの**センターから外れた際（オフセンター）にも，センターと同等**の放射特性を取得できる手法の開発を目指して，観測点が疎になる方向に観測点を補間して放射特性を推定した。

◆センターとオフセンターの放射特性間の誤差量を，補間の有り無しについて計算した（Fig. 1）。補間によって 1 dB 未満の誤差量の度数増加が期待されたが，Fig. 1(a)では減少しており改善とは言えない結果となった。一方，Fig. 1(b)では 1 dB 未満の度数の明確な増加はなかったものの，5 dB 以上の誤差が減少していることと 5 dB 未満の総数が 5 %増加していることから改善の傾向がみられた。

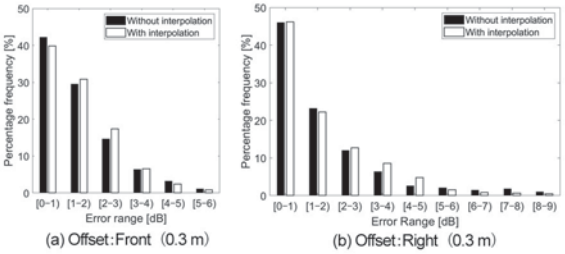


Fig. 1：Error range in estimated radiation pattern. (Values below 1% in both datasets are omitted.)

1-7-7

1-7-7　臨場感向上のためのマルチチャンネル曲げ波スピーカシステムの評価

Evaluation of Multichannel Bending Wave Loudspeaker Systems to Improve the Sence of Presence

☆橋本篤拓，河原一彦（九州大・芸工）

◆曲げ波スピーカ（BWL）は「いわゆる放射の拡散的性質」を最適となるように河原により提案されている。

◆BWL を含んだ 5 チャンネルスピーカシステムを構築し、インパルス応答による音響特徴量の算出と音楽ライブ音源を利用した再生音場の印象評価実験を実施し、システムの評価を行った。

◆音響特徴量を算出した結果、両耳間相互相関（IACC）、横方向初期エネルギー率（LF）、両耳コヒーレンス（IC）では BWL を利用することによる値の変化が見られた。

◆印象評価実験の結果、BWL を利用することによる臨場感の向上は確認できなかった、臨場感と音響特徴量との重回帰分析の結果、IC と臨場感に関連があることがわかった。

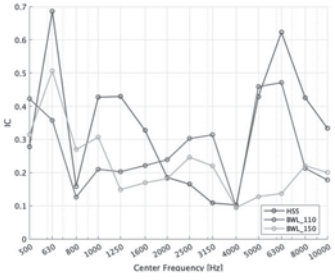


Fig.1 Inter-aural Coherence (IC) plot per 1/3 Oct. bands

1-7-8

1-7-8　複合的音響提示環境における安全な音圧レベルの考察

A Study on Safe Sound Pressure Levels in Hybrid Acoustic Presentation Environments.

○伊藤弘章，西潟愛美，野口賢一，鎌本優（NTT）

◆スピーカとオープンイヤー型イヤホンを組み合わせた複合音響提示環境における安全な音圧制御手法を検討

◆一律制御および聴覚心理モデルに基づくマスキング効果を考慮したゲイン制御手法を提案し、シミュレーションで有効性を評価。

◆聴覚マスキングを活用することで、耳の安全を確保しつつ没入感のある音響体験が可能

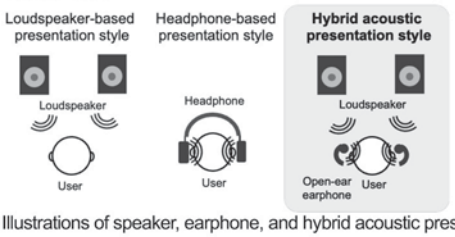


Fig.1: Illustrations of speaker, earphone, and hybrid acoustic presentation styles

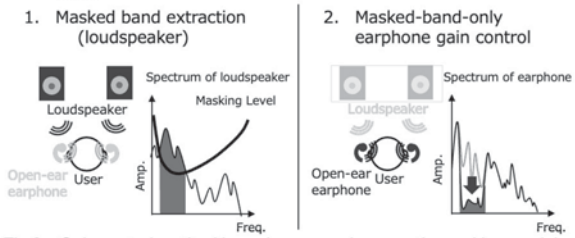


Fig.2: Gain control method based on a psychoacoustic masking model.

1-7-9

1-7-9　ライブ会場における大音量暴露による難聴リスク低減に向けた検討

A preliminary study on the risk reduction of hearing loss due to louder sound exposure at live venues

◎西潟愛美，野口賢一，加古達也，伊藤弘章，鎌本優(NTT)

◆ライブ会場ではスピーカアレイの近傍と遠方での音圧の差があり後方の音量に合わせると前方の受聴者が大音量に暴露され難聴リスクにつながる課題

◆オープンイヤー型デバイスを用いて音量を補完することで受聴時の音圧を均一化することによる難聴リスク低減を提案

◆会場の実測データを用い音圧を補償する方法を提案し評価

◆客席間の音圧差を改善し難聴リスクの低減を示唆する結果を確認

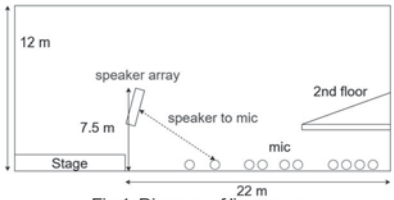


Fig.1:Diagram of live venue

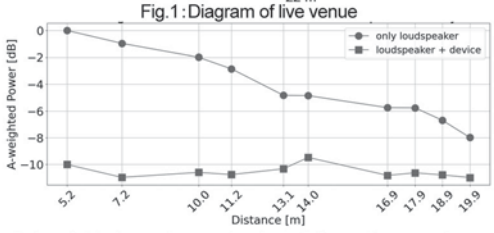


Fig.2:A-weighted sound power level vs. distance from speaker array (Participants ○:only speaker, □:speaker + open ear device)

1-7-10

単純な利得と遅延操作による 2 チャンネルの Optimal Source Distribution に基づくクロストークキャンセルシステム −マルチウェイスピーカを用いた効果検証−

Crosstalk cancellation system with adjusting simple gain and delay based on the two-channel optimal source distribution

◎原健登，星野嗣人，矢入幹記(鹿島技研)，△武内隆(OPSODIS Ltd./ISVR)

◆スピーカを用いたバイノーラル合成の課題として、バイノーラル音源の絶対数の少なさや、逆フィルタ畳み込み演算処理が持つ計算負荷や遅延，実装の煩雑さが挙げられる。

◆そのため筆者らは逆フィルタの畳み込みを必要とせず、OPSODIS 原理に基づき離散化・帯域分割された音源に、各周波数帯を代表する相対利得と遅延を付与するのみの簡易的なバイノーラル合成システムを提案してきた。

◆本報では提案システムの実証のため、筆者らが設計・製作・測定したスピーカ-ダミーヘッド間のインパルス応答を用いて、提案システムと従来システムのバイノーラル合成を数値シミュレーションし、有用性を確認した。

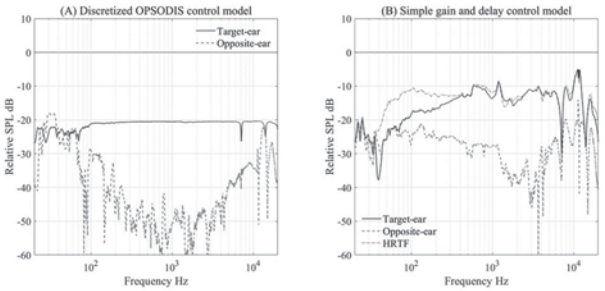


Fig. 7 Control effect of binaural synthesis and amount of crosstalk cancellation using measured data from multi-way speaker system. (A): Conventional method (inverse-filter convolution system based on discretized OPSODIS principle), (B): Proposed method (simple gain-and-delay system)

1-7-11

1-7-11　3D オーディオにおける上層のマイクロホンの高さと最適聴取範囲との関係

Relationship between upper microphone height and sweet spot in 3D audio

○亀川徹，丸井淳史(東京芸大)

◆3D オーディオにおける上層のマイクロホンの指向性と高さの違いによる最適聴取範囲との関係について調査した。

◆類似度と評定尺度の結果から，中央の聴取位置から横にずれた場合に定位や音像幅の印象が変化する傾向が見られた。

◆中層と上層のマイクロホンの間隔を広げた場合に，聴取位置が横にずれても印象の違いが少ないことから，「上下方向の相関度が低くなることで，中央の聴取位置と印象が変わらない聴取範囲が広がる」という前報の結果と同様に，実際の録音環境においても，マイクロホンの指向性や高さによって上下方向の相関値が低くなることで，中央の聴取位置と印象が変わらない範囲が広くなる可能性が示唆された。

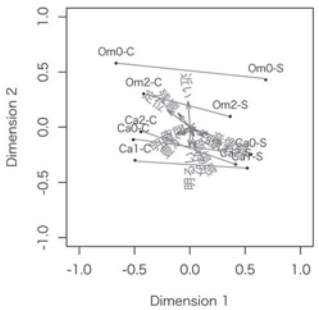


Fig.1: Result of the rating scale of Exp. 2 superimposed on the figure of distribution of the stimuli obtained from INDSCAL of Exp. 1 (cello).

1-7-12

1-7-12　ウェアラブルデバイスの音響透過性がオーディオ AR 体験に与える影響

Effects of audio AR experience
caused by acoustic transparency of wearable audio devices.

◎渡邊悠希，千葉大将，野口賢一，伊藤弘章，加古達也，鎌本優（NTT）

◆オーディオ拡張現実（オーディオ AR）体験に用いられるウェアラブル音響デバイスでは，実世界の音を劣化せず知覚できることが重要である．

◆先行研究では，実世界の音が装着デバイスを通して自然に聞き取れる程度，すなわち音響透過性について，物理的および知覚的な側面から評価されてきた．しかし，音響透過性と Audio AR 体験時の主観印象との関係は明確には解明されてこなかった．

◆本研究では，音響透過性と Audio AR 体験時の主観印象との関係を調査するため，異なる形状をもつ５つのヘッドホンやイヤホンについて評価した．

◆実験の結果，音響透過性の物理指標と実世界の音の空間印象との間に相関がみられることがわかった．また，音響透過性が高い，耳を塞がない形状のデバイス装着時は，Audio AR 体験に対する主観印象が優れていた．

Tab.1: Selected devices.

Symbols	Models	Shapes
A	AKG K712 PRO	Over-ear Headphone
B	nwm ONE	Extra-aural Headphone
C	audio-technica ATH-CM707	Intra-concha Earphone
D	Apple Airpods Pro 2	Insert Earphone
E	nwm MWE001	Open-ear Earphone

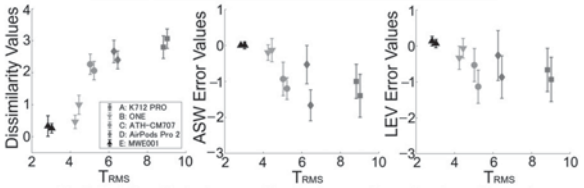


Fig.1: Relationship between subjective impressions of real-world sounds and physical indices of acoustic transparency (T_{RMS}).

1-7-13

1-7-13　ARIB 技術資料 TR-B48 に準拠したオブジェクトベース音響対応ライブ制作システムの構築

Construction of a live-production system for object-based audio complied with ARIB Technical Report TR-B48

◎久保 弘樹，岩崎 泰士，杉本 岳大，大出 訓史（NHK）

◆オブジェクトベース音響（Object-based audio: OBA）の次世代地上テレビジョン放送への導入に向けて，2025 年 3 月に音響メタデータの運用ガイドライン（ARIB 技術資料 TR-B48）が作成された．

◆TR-B48 は国際標準の音響メタデータ ADM やそのシリアル表現形式である S-ADM に準拠しつつ，ARIB で独自に拡張した音響メタデータ，gainControl を定義している．gainControl により，一つの背景音に異なる言語のダイアログを組み合わせて副音声を構築する場合などに聞き取りにくいダイアログができないよう，組み合わせによって背景音とダイアログのレベルバランスを動的に調節可能となる．

◆OBA の放送チェーンでは，各機器が音響メタデータを適切に変換しながら伝送する．既開発のライブ制作システムを改修し，gainControl を含む ADM から動的に変化するレベルバランス値を時分割した S-ADM への変換，および S-ADM の gainControl から MPEG-H 3D audio 規格で規定されている drc metadata への変換を可能とした．

◆図 1 の系で最終段の MPEG-H 3D audio 復号装置の出力を記録し，ADM に記述された通りのレベルバランス調節が行われたことをもって，構築したライブ制作システムの正常動作を確認した．

Fig.1 : Block diagram of connection verification

1-7-14

1-7-14　高速 1bit 信号と有限インパルス応答の直接畳み込み法

Direct convolution method
between high-speed single-bit and finite impulse response

☆五味佑太，櫻井意織，坂口智弘，及川靖広（早大理工）

◆高速 1bit 信号をマルチビット信号に変換せずに直接混合する手法を応用し，値の算術演算を伴わない畳み込み処理を提案する。

◆同一の高速 1bit 信号に対して，無音信号の挿入による信号の遅延と，サンプルの並べ替えによる信号同士の混合を用いて畳み込み処理を実装した。

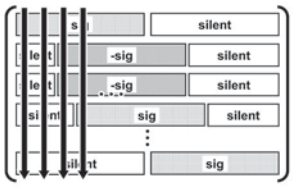


Fig.1 Proposed Method Overview

◆ハイパス特性とバンドストップ特性を含むフィルタの畳み込みにおいて，通常の畳み込み演算とほぼ同等のパワースペクトルが得られた。

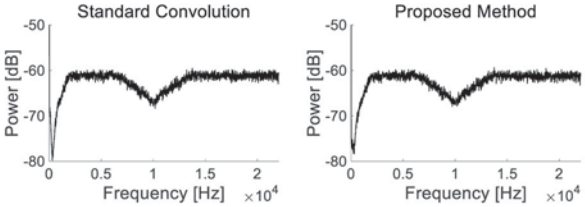


Fig.2 Power Spectrum of Standard Convolution and Proposed Method

1-7-15

劇場舞台音響システムによる高臨場感創出の歴史と最近の動向

○和田竜一，鈴木航輔，小口恵司（永田音響設計）

劇場空間において、「イマーシブオーディオ」と称される音響システムを用いた、従来の音響演出を超える没入体験が提供されるようになった。本稿では、劇場で用いられてきた舞台音響システムによる音像定位および音場支援を概観し、それらを統合したイマーシブオーディオの概要と事例を紹介する。

◆音響システムを用いた音像定位と音場支援

音像定位については、実音源の有無に関わらず仮想的な位置に音像を結ぶ「Spat」及び、実音源の定位をより確実にする補強方法の Delta Stereophony について触れる。

音場支援については、最初に実用化された Assisted Resonance に触れる。

◆イマーシブオーディオ

イマーシブオーディオの信号処理の概要を紹介する。

◆イマーシブオーディオの事例と課題

事例として「Sapporo City Jazz」や「ABBA Voyage」を紹介する。また、課題として主にオペレーションの難しさについて述べる。

1-7-16

1-7-16　Active Field Control(AFC)システムによるイマーシブ音場の実現

Immersive sound field with Active Field Control (AFC) system

〇宮崎秀生(ヤマハ)

◆音場支援システム（Active Field Control：AFC）は、演奏者や観客を取り囲む音場及び音要素を電気音響の支援により制御することでイマーシブ音環境を創り出すシステムである。元々はある教会の音響設計に際し、“講話時の話声の明瞭度”と“オルガン演奏時の豊かな響き”との相反する音響条件を一つの空間内に両立させるため、音場を制御する音響設計ツールの一つとして 1980 年代に開発された。

◆AFC は、音場の主要な構成要素である残響の長さや音の空間印象を制御する Enhance システムと、音要素の重要な情報である音の方向性や大きさを制御する Image システムから構成される。各システムを単独で用いる、又は組み合わせることで、目的に合わせた効果的な音環境創りが可能となる。

◆AFC の基本概念と共に代表的な適用事例を紹介する。

Fig.1　Enhance システム基本モデル

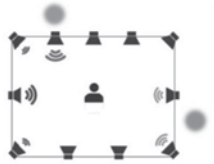


Fig. 2　Image システム概念図

1-7-17

1-7-17　大規模マルチチャンネルスピーカー環境の実装とヒビノの取り組みについて

Implementation of Large-Scale Multichannel Speaker Systems and Hibino's Initiatives

〇松井祐大(ヒビノ株式会社)

◆本発表では、非劇場型の大規模空間における高臨場マルチチャンネル音響の実装に関する取り組みを報告する。従来の音響設計手法では対応困難な空間的・演出的制約に対し、ヒビノ株式会社において開発された音響・映像・振動を統合した空間演出システム「AXYZ（仮称）」の構成と実験結果及び実際の実装事例を紹介する。内部空間における自由な移動を前提とした設計とし、立体音場再現を行った。コンテンツ制作と実装を一体的に進め、複数イベントでの応用を通じた実証も行っている。また、多受聴点環境におけるスピーカー配置の最適化に関して、東京藝術大学との共同研究を通じて定量的評価指標の確立を目指している。

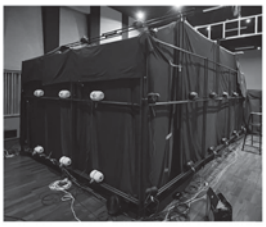


Fig. 1　実験の様子

Fig. 2　実験の設営

1-7-18

1-7-18　映画館での Dolby Atmos 設計・調整の実践

Dolby Atmos design and audio equalization practices in the movie theaters

羽金淳（Dolby Japan）

◆現在全世界および日本国内で実際に設計・調整され、実際に映画館で稼働している Dolby Atmos の具体的な設計・音響調整方法について技術的背景からの観点を中心に解説します。

◆Dolby Atmos の技術的背景、開発目標を踏まえたうえで、映画館での Dolby Atmos の設計上のガイドラインおよび、3 つの大きな要素、様々な要件に対する実践的な対応方法を解説いたします。

◆機器設置後に実施する音響調整でのガイドラインおよび、実践的な調整方法について解説いたします。

1-7-19

1-7-19　高臨場感オーディオ技術によるサウンドスケープの再生：観光プロモーションへの展開の可能性

Soundscape reproduction with highly immersive audio technology

〇末永信一(日本オーディオ協会)，△片向史一(千葉工大)，△高柳欽也(名古屋芸大)，関研一(千葉工大)

◆海と山に囲まれた自然豊かな南房総市のサウンドスケープを，様々な高低差のある音源が響く高臨場感オーディオとして収音，再生をした．印象評価，脳波計測等を用いて聴取者の反応を詳細に分析し，観光プロモーションへの展開の可能性を検討した．

◆アンビソニックマイクで収音した音を，バイノーラル，及び，立体音響システムで再生し，臨場感を高める上でのそれぞれの効果が確認できた．現地の音と地域文化を届ける「音旅トレカ」や，「スピーカーで組まれた立体音響システム」を，観光プロモーションツールとして活用する上で，具体的な知見が得られた．

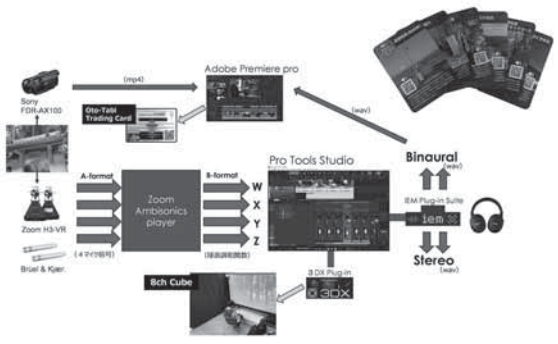


Fig.1 : Sound pickup and playback scheme, Oto-tabi trading card

1-8-1

1-8-1

デフォーカスを用いた弦振動二次元変位測定における逆 Radon フィルタ

Inverse Radon filter
for two-dimensional displacement measurement
of string vibration by defocusing

☆ 新美智也，赤石夏輝，矢田部浩平（農工大）

これまで
- デフォーカスを利用した高速度カメラによる弦振動の二次元的な測定手法を提案

問題点
- カメラとレンズの仰角や，照明条件によって弦上の計測点の検出性能が低下する

今回
- 逆 Radon 変換を用いたノイズ除去手法を提案
  - ▻ 従来のノイズ除去手法と比べて光量ムラがある場合の検出点のぶれを抑制

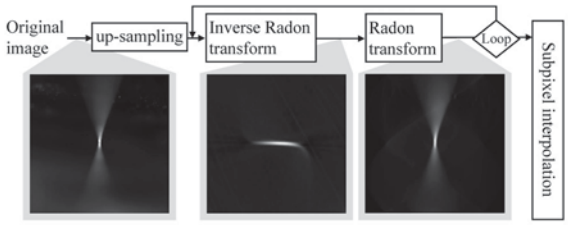


Fig. 1 Flowchart of image processing using inverse Radon transformation

1-8-2

1-8-2　ヴァイオリンにおける擦弦方向・弓圧方向の演奏動作を考慮した擦弦振動解析

Vibrational Analysis of Bowed String Considering Bowing Direction and Bow Pressure Direction in Violin Playing.

☆佐倉歌，鮫島俊哉（九大）

◆本研究では，ヴァイオリンの擦弦振動を解析するため，弓のx（擦弦）・y（弓圧）方向の 2 自由度運動と，それに対応する弦の 2 次元振動を連成させたモデルを構築した(Fig. 1)。演奏者の手の動作は時間変化するディリクレ境界条件として与え，実際の演奏を模擬した。

◆数値解析では，3 通りの手の動作（①弓圧固定，②弓圧変化，③弓を弾ませる Ricochet 奏法）を設定し，弓と弦の動的相互作用を調べた。

◆結果，①では，弓の弾性によるy方向の微小振動（弓が踊る現象），②では，弓圧変化により振動周期が跳躍する現象が確認された(Fig. 2)。③では，弓の再接触に伴う振幅増加が生じたが，接触条件に課題があり，非物理的な挙動が見られた。

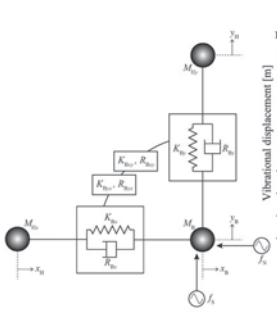


Fig.1: 2-DOF spring-mass-damper system model of a violin bow.

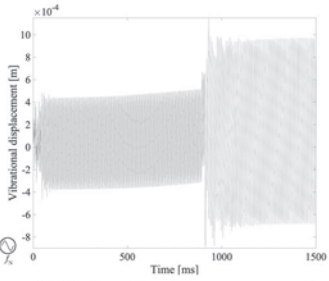


Fig.2: Transitional time response of the bowed string at the bowing point in the x-direction when bow pressure is varied with time.

1-8-3

1-8-3　脚部の質量分布の変化が振動継続時間に与える影響：振動覚音叉を用いた検討

Effects of mass distribution in tuning-fork legs on vibration duration: An experimental study using tuning forks for vibration sensation testing

☆上野嵩旺，大塚　翔，中川誠司（千葉大）

◆音叉の使用用途を鑑みると，基音はより長く持続し，倍音はより速やかに減衰することが望ましい．しかし，音叉の製作工程では持続時間の評価は行われておらず，音叉振動の持続時間には個体差が存在する．

◆前報では，サージカルテープを貼付することで音叉の脚部先端の質量を段階的に変化させて振動持続時間の挙動を観察した．左右脚部の質量分布および対称性の影響が示唆されたが，より大きな質量変化を与えることで，より明瞭な効果を観察出来るかもしれない．

◆複数の振動覚音叉で左右脚部に取り付けられた錘の位置を変化させながら打撃を行い，脚部の質量分布が振動継続時間に影響を観察した．

◆錘を脚部先端から柄部側に移動させるに従い，基音の振動継続時間が短縮する傾向が確認された．錘が先端にあるほど固有振動数が減少したことで，内部摩擦や空気抵抗などによるエネルギーの散逸が相対的に遅くなり，振動継続時間が延長した可能性が考えられる．また，脚部先端の質量を柄部側に比べて増すことで，基音がより長く持続する音叉を作成できることを示している．

◆今後，左右脚部の錘の質量や位置の対称性の影響についても検討が必要であろう．

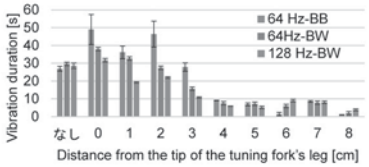


Fig.1 Vibration durations as functions of weight position on tuning fork's legs.

1-8-4

1-8-4

中国四川省の真鍮製口琴の音響分析

Acoustic analysis of brass Jew's harp
from Sichuan province, China

☆王櫟，新美智也，安藤元暉，矢田部浩平（農工大）

◆ 背景
- ➢ 口琴（Fig.1）は口を利用して演奏する楽器
  - ✧ 口琴の発音源の音高はほぼ一定だが演奏者は口腔の容積などで制御して音色を変えて表現する
- ➢ 演奏者の口腔形状制御が音響特性に与える影響など未解明
  発音の仕組みの解明が求められる

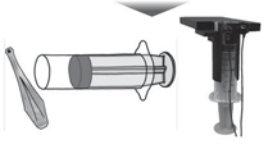
◆ 本研究
- ➢ 中国四川省に伝わる真鍮製口琴，コウシャン（Fig.1 下）を研究対象とする
- ➢ 演奏時の口腔容積の変化（Fig.2 上）を模擬するためシリンジを用いた簡易口腔モデル（Fig.2 下）を提案・製作
- ➢ 容積の違いが演奏音に与える影響を定量的に評価できた

Russian brass Jew's harp

Schematic diagram of the actual oral cavity

Brass Jew's harp from Sichuan Province, China (Kouxian, 4-pieces type)

Fig.1: Various Jew's Harps



Schematic diagram and produced actual item of the oral cavity modeling

Fig.2: Photos of actual oral cavity and its modeling for Kouxian performance

1-8-5

1-8-5　小鼓の胴と革の接触圧力と振動モードに関する研究～楽器の音響モフォロジーの研究～

Research on contact pressure and vibration modes between the body and membrane of a kotsuzumi ~Research on the acoustic morphology of musical instruments~

☆山下佳也，田野直輝，松谷晃宏，田原麻梨江（東京科学大）

◆小鼓 (Fig. 1(a))のリアルタイム張力測定に向け，革と胴の接触圧力による新たな張力測定法(TCPM)を考案した。

◆小鼓を模した治具を用いて胴と革の間の接触圧力(p)と振動スペクトルの測定を行った。その結果，ある接触圧力での張力(T)上昇の鈍化を確認 (Fig. 1(b))。考察により，表革の張力上限が裏革の物性を基準として決定される可能性が示唆された。また，TCPM と従来の基本振動数による張力測定法を用いて小鼓の張力測定を行ったところ，従来手法との誤差は8%程度となり，提案手法の妥当性が確認された。

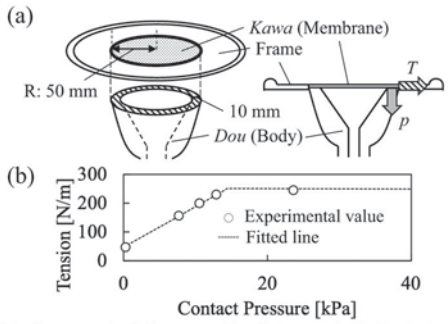


Fig. 1 (a) Basic conceptual diagram of the *kotsuzumi*. (b) Contact pressure vs. tension for jig imitating *kotsuzumi*.

1-8-6

1-8-6　計測された加振力の時刻歴を用いる数値シミュレーションによるシンバル音の合成

Numerical sound synthesis of cymbals using measured time history of excitation force.

☆堀之内太貴（筑波大・院・シス情工），若槻尚斗，海老原格（筑波大・シス情系）

◆数値シミュレーションによってシンバル音を合成する手法において，加振力の時間波形は応答に大きく影響し，従来用いられていた簡易的な加振力のモデルでは不十分であることを示した。

◆検証のために，二種類の加振力の時間波形を用いてシンバルの振動シミュレーションを行い，それぞれ結果を実測値と比較した。

◆シミュレーションに用いた加振力の時間波形は，ドラムスティックの特性を有したカセンサによる計測結果に基づいたものと，従来の研究で用いられていた簡易的なモデルから導かれたものである。

◆実測に基づいた加振力の時間波形を用いるシミュレーションにより得られた振動加速度と，実際のシンバルの振動加速度を比較すると，Fig.1 に示すように 0kHz から 5kHz の範囲で近い周波数特性を持っていることが確認できた。

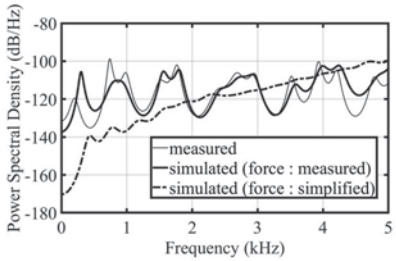


Fig.1 : Spectral envelope comparison of measured and simulated cymbal vibration acceleration.

1-11-1

1-11-1　筋膜界面状態がせん断波伝播特性に及ぼす影響の Ex Vivo 評価 ～せん断波に関する音響モフォロジー～

Ex Vivo Study on the Effects of Fascial Interface Conditions on Shear Wave Propagation – Acoustic Morphology of Shear Waves –

☆田野直輝（東京科学大），△紺野啓（自治医大），△谷口圭吾，△塩泡孝介（札幌医大），△江田康，△山越芳樹（群馬大），田原麻梨江（東京科学大）

◆ 本研究では，界面状態の変化がせん断波伝播特性に及ぼす影響を明らかにすることを目的とし，筋膜治療法であるハイドロリリース(HR)に着目した Ex Vivo 試験を実施した。鶏肉試料の大胸筋と小胸筋の間に位置する筋膜層に生理食塩水を注入し，HR を模擬した介入を行った上で，注入前後に著者らが提案した外部加振器を利用した SWE 測定を実施した(Fig.1 (a))。

◆ 得られたせん断波位相分布像(Fig.1 (b))では HR 後に筋膜境界部で急峻な位相変化が観察され，筋膜を挟んだ上下領域間の位相差は 0.29 rad から 0.75 rad へと増大した。この結果は，不均質組織の界面状態をせん断波の位相差に基づいて評価できる可能性を示唆している。

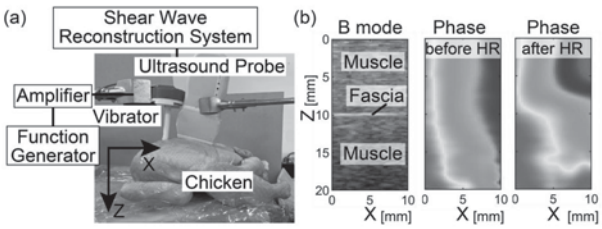


Fig. 1: (a) A picture of ex vivo measurement setup. (b) Measured B-mode image and phase maps before and after HR in ex-vivo measurements.

1-11-2

1-11-2　SWE における Push Beam 照射時間と Shear Wave の時間周波数の関係性評価のためのファントム実験

Phantom Experiment for Evaluating the Relationship Between Push Beam Duration and Shear Wave Temporal Frequency in SWE

☆髙木大世（富山大院・医薬理工学環），
長岡亮，大村眞朗，長谷川英之（富山大・工）

◆Shear wave elastography(SWE) において，ARFI を生じさせる push beam の送信時間（duration）と shear wave の時間周波数の関係についてファントム実験を用いて検討を行う.

◆結果として，push beam の照射中心近傍では duration の長さが増すほど shear wave の横波周波数は低下することが確認された.

◆今後，組織の不均一性を考慮し，縦波音速が変化する場合の関係性について検討を行う.

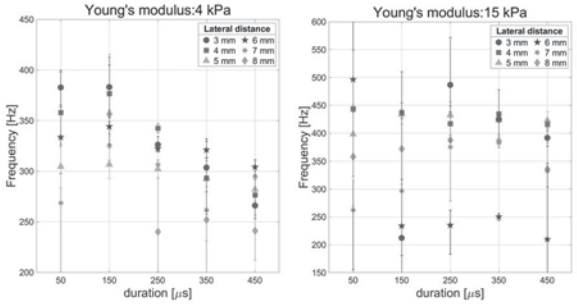


Fig. 1 Relationship between the temporal frequency of the SW and duration (left) 4 kPa, and (right) 15 kPa.

1-11-3

1-11-3　ネットワークアナライザを用いた双方向超音波加振により発生した変位の計測による粘弾性特性推定

Estimation of viscoelasticity using displacement measurement generated by dual ultrasound excitation with network analyzer

○新谷聡基，荒川元孝（東北大）

◆ 我々は，双方向からの超音波照射により、対象物表面で発生する**変位と加速度**の計測による粘弾性特性の推定法に関する研究を進めている。

◆ 先行研究では**対象物背面からの反射波**が変位計測に誤差を与えていた。背面反射波を抑える**非平行ファントム**(Fig. 1)を導入した。

◆ 非平行ファントムは平行ファントムよりも背面反射成分振幅が小さくなった（Fig. 2)。非平行ファントムは平行ファントムよりも**微小な変位を計測できる可能性があること**を示した。

◆ 非平行ファントムの変位と加速度の比（相対$H(\Delta f)$）は、**Voigt モデルに概ね従った**(Fig. 3)。

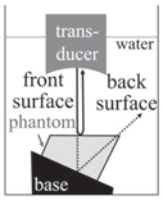


Fig. 1: Schematic of measurement for a non-parallel phantom.

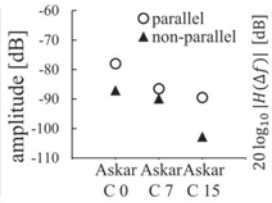


Fig. 2: Amplitude of the back surface reflection.

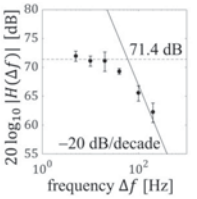


Fig. 3: Relative $H(\Delta f)$ of the C0 phantom.

1-11-4

1-11-4　超音波照射や骨接合用の圧電性高分子材料を利用した骨折治療法の開発

Development of bone fracture healing technique using low intensity ultrasound and weak piezoelectric polymer for bone fixation

☆村上優太，△北嶋昇太，△和田大河，△前田貴光，△小澤翔愛(同志社大)，
△大西政彰(兵庫医科大)，△中島美和(ドクトル鍼灸医学研究所)，
△井上基浩(宝塚医療大)，松川真美(同志社大)

◆ **ポリ-L-乳酸（PLLA）**と低出力超音波パルス（LIPUS）を併用した**新たな骨折治療法**を実験的に検討した。

◆ **ラット脛骨**に骨孔を作成し 2 週間経過後 **X 線 CT を用いて骨量を推定した。**自然治癒の Control 群、超音波照射を行う **US 群**、PLLA フィルムを挿入し超音波照射を行う **PLLA＋US 群**で比較を行った。

◆ やわらかい仮骨ではControl 群に比べ **US 群は 2.1%**, **PLLA+US 群が 10.4%増加**した。

◆ 超音波照射時の誘発電位が骨折治療に寄与する場合、骨折部の PLLA 接合材が治療の促進に役立つ可能性がある。

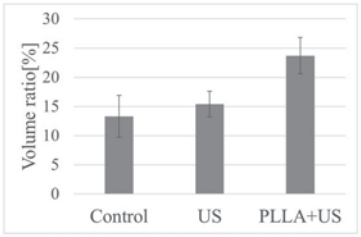


Fig.1：Changes in bone volume due to ultrasound irradiation and PLLA film.

1-11-5

1-11-5　ウェアラブル超音波トランスデューサを用いた口動作の分類 ～表面付近の生体情報に関する音響モフォロジーの研究～

Classification of mouth movements with wearable ultrasound transducers. ~Study of acoustic morphology of biological information near the surface~

☆金谷拓海，田野直輝(東京科学大学)，荒川元孝(東北大学)，
△小野雄(Carleton University)，田原麻梨江(東京科学大学)

◆ 顎二腹筋に設置した 2 つのウェアラブル超音波トランスデューサ(UT)でリアルタイムで記録したパルスエコー波形から無声発話認識によって母音を分類する手法の有効性を検証した。

◆ 得られた特徴量に対し、k-近傍法(kNN)および線形判別分析(LDA)を適用し、5 分割交差検証によって性能を評価した。

◆ 結果、kNN では約 80%, LDA では約 90%という高い総合精度を達成した。このことから、口動作を用いた母音分類の原理的な可能性が示された。

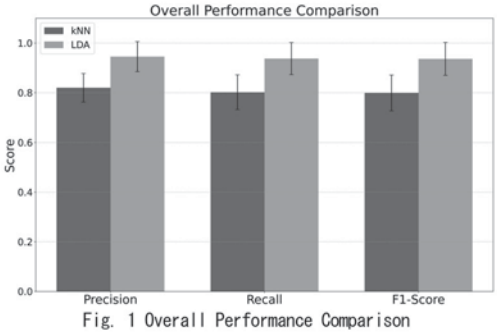


Fig. 1 Overall Performance Comparison

1-11-6

1-11-6　超音波アレイスピーカの指向性制御の試み

Attempts at controlling the directionality of ultrasonic array speaker.

○今村俊樹，△池田賢二(セコム・IS 研)

◆セコムでは独居者の見守りサービス高度化のため，超音波を使った，人の位置や状態を把握する研究を進めている。

◆先に報告した 180 度指向性スピーカは指向性が広すぎるが故，居室，特に狭い空間でセンシングすると，壁からの反射が想定以上に影響を及ぼすことが分かった。そのため今回，不要方向に超音波を放射しないよう超音波の指向性を制御する手法を検討した。

◆アレイ構造のスピーカはフェイズドアレイ手法で指向性の制御が可能である。しかし音源素子サイズが大きな従来構造では，指向性制御角度が±12deg 程度と狭く，広域の制御が困難であった。

◆そこで今回，大口径から小口径へ変換するスピーカユニットを組み合わせたアレイ構造を新たに考案し，制御角度を±90deg まで拡大できることをシミュレーションで確認した。また 3D プリンタで実機を試作した上で評価・検証し効果を確認した。

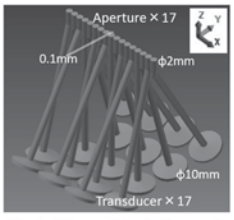


Fig.1：3D-model of ultrasonic speaker with wide-angle controllable directivity.

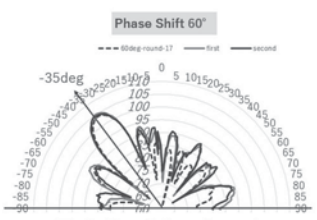


Fig.2：Directivity pattern of simulation and experiment results. (Phase shift = 60°, at 25kHz)

1-11-7

1-11-7　毛髪が VHF 音の伝搬に与える影響 -パルスエコー法を用いた髪型推定の試み-

Influence of hair on VHF sound propagation - A trial to estimate hairstyle by using the pulse-echo method

☆松下宗輝, 久保谷旭陽, 近藤奏海(神奈川工科大), 廣江正明(小林理研),
Zuhal Firdaus, 中村健太郎(東京科学大),
長谷川英之(富山大), 上田麻理(神奈川工科大)

◆VHF 音(16 kHz～32 kHz の音)の HRTF において，毛髪の影響を確認するために，パルスエコー法を用いた髪型と VHF 音の反射特性の関係について実験的に調査した.

◆ウィッグあり/なしの HATS に 16 kHz～21 kHz のトーンバースト波を照射した時の入射音と，測定対象からの反射音を計測し，反射率を比較した.

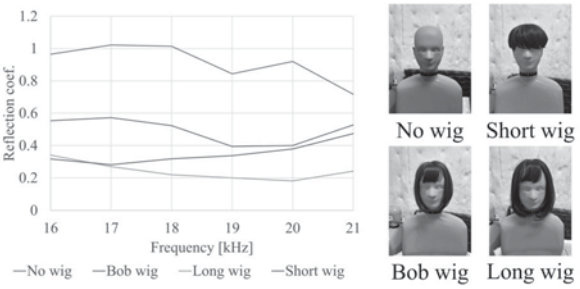


Fig.1: Frequency characteristics of reflection coefficients normalized by the reflection rate of a steel plate

1-11-8

1-11-8 VHF 領域に顕著な成分を持つ音源に対する聴こえ評価の試み

An attempt of auditory evaluation against sound sources with prominent components in the very high frequency region

☆久保谷旭陽, 松下宗輝, 近藤奏海(神奈川工科大), ZuhalFirdaus,
中村健太郎(東京科学大), 廣江正明(小林理研), 上田麻理(神奈川工科大)

◆先行研究では 16 kHz～22 kHz のバースト信号を用いた聴覚閾値計測を行ったが，18 kHz 越の周波数領域では十分な音圧レベルを呈示できず、正確な閾値計測ができなかった.

◆本研究では，高周波数成分を持つ鼠撃退器音を用いて音源の向き，高さ，受聴者との距離を変え，聴覚閾値計測と聴感印象評価を実験的に行った。

◆音源は 19 kHz 付近にピークを持ち，音圧レベルは約 105 dB(SPL)であった(Fig.1). また，音圧レベル 90 dB(SPL)付近が閾値という結果を得た(Fig.2).

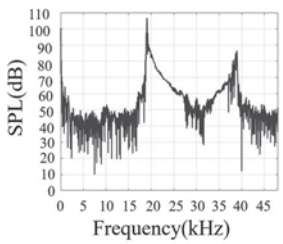


Fig.1 : Frequency characteristics of sounds from the rodent repellent at 1-m distant.

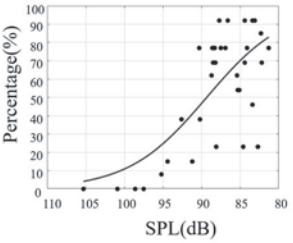


Fig.2: Relationship between PSL from the rodent repellent and response rate of inaudible participants for all conditions of auditory evaluation.

1-11-9

1-11-9　レーザ固定-探触子走査による カットオフ周波数に着目した板厚計測

Plate thickness measurement focusing on cutoff frequency by laser fixed-probe scanning

◎清水 鏡介(愛媛大院), △水谷 隼人(愛媛大),
△重松 尚太(愛媛大院), 中畑 和之(愛媛大院)

◆鋼板等の減肉の検査方法としてレーザ超音波を用いた Lamb 波のカットオフ周波数に着目した方法について研究を行っている.

◆本報告では，固定設置したレーザによって超音波を励起させ，各計測点において受信することで板厚の推定が可能であるか実験的に検証した.

◆Table 1 に実験により推定された板厚を示す.

◆結果より，5，10 mm の板厚においては 3 %以下，15～25 mm の板厚においては 7 %，30 mm の板厚においては誤差率が 13 %となった.

Table 1. Plate thickness estimated from experimental results.

Measured sample thickness [mm]	Measured resonant frequency [kHz]	Calculated sample thickness [mm]	Error rate [%]
29.77	116.01	25.78	-13.40
24.89	130.28	22.96	-7.76
19.77	165.73	18.05	-8.72
14.80	217.09	13.78	-6.91
9.84	313.10	9.55	-2.92
4.96	610.14	4.90	-1.17

1-11-10

1-11-10 段付き円形振動板とパラボラ反射鏡による非接触弾性計測～非接触弾性計測システムを用いた音響モフォロジーの研究～

Non-contact elasticity measurement using stepped-circular plate and parabolic reflectors ~Research on acoustic morphology using a non-contact elasticity measurement system~

☆山下佳也, 中村健太郎, 田原麻梨江(東京科学大)

◆著者らが提案した非接触弾性計測システムの小型化を目的とし，100 kHz で共振する小型の段付き円形振動板 (CSPT)を FEM の固有振動解析で設計し (Fig. 1(a))，機械加工により製作した。ビーム径は約 10 mm だった。

◆異なるショア硬さを持つ試料に対しパラボラ反射鏡を用いて CSPT から発生した空中超音波を集束し，試料表面を励振した。LDV で振動特性を測定し，解析した結果，ショア硬さが高くなるほど振動速度が低下した (Fig. 1(b))。この傾向は弾性モデルと同様の傾向であることから，本システムが弾性評価として利用可能であることがわかった。

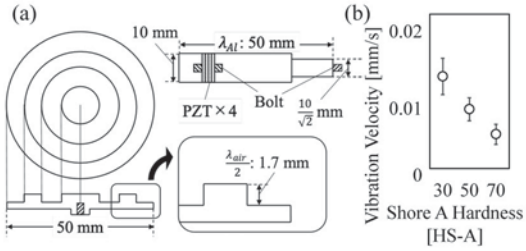


Fig. 1 (a) CSPT configuration diagram (b) Shore A hardness vs. surface vibration velocity for urethane rubbers.

1-11-11

1-11-11　相互相関を用いた花崗岩の音響インピーダンス顕微鏡映像～超音波顕微鏡を用いた音響モフォロジーの研究～

Acoustic Impedance Microscopic Imaging of Granite using Cross-Correlation Method -Study on Acoustic Morphology with an Ultrasonic Microscope -

☆上原陸太（東京科学大），△川口祐季（本多電子），小林和人（本多電子），
△江波戸宗大（農研機構），松谷晃宏（東京科学大），田原麻梨江（東京科学大）

◆鉱物表面のミクロスケール特性評価を目的として、超音波顕微鏡を用い、花崗岩の構成要素である石英、斜長石、雲母の音響インピーダンス画像を取得・評価した。Peak to Peak、RMS、相互相関の 3 手法による信号処理を比較し、最適な処理法を検討した。また、その結果をもとに花崗岩中の鉱物の識別を行った。

◆鉱物表面の音響インピーダンス画像取得で、相互相関が最も有効な信号処理方法であることがわかった。(Fig. 1)また、花崗岩に対して音響インピーダンス画像による鉱物識別可能性を示した。

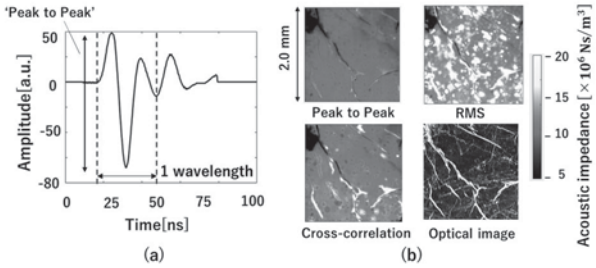


Fig.1(a) Reflected signal from mica (b) Acoustic impedance image of mica obtained by Peak to Peak, RMS, Cross-correlation and Optical image.

1-11-12

1-11-12　超音波伝搬シミュレータによる集束超音波探触子設計方法

Focused ultrasonic probe design method using ultrasonic propagation simulator

○田中雄介，　△小倉幸夫(ジャパンプローブ)

◆集束超音波探触子の焦点位置は直接波とエッジ波の重なりにより発生する。

◆焦点位置に対する直接波とエッジ波の影響度がそれぞれ最大となる点 F_D と F_E の間に焦点位置がある。

◆焦点位置に対するエッジ波の影響度 I_E を求めることで焦点位置を推定できる。

◆エッジ波の影響度 I_E は凹面音源の開口角 θ と関係し、シミュレータにより θ と I_E の関係を求めた。

◆F_D と F_E、I_E から求めた焦点計算値 F_C と実際に試作した集束超音波探触子の焦点 F が良く一致した。

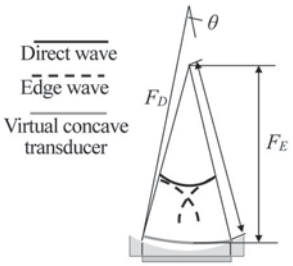


Figure 1 Virtual concave source and direct and edge wave propagation distances.

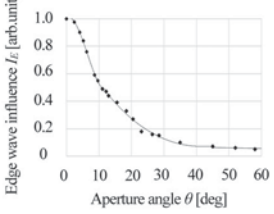


Figure 2 Relationship between aperture angle θ and influence of edge wave I_E.

1-11-13

1-11-13　MHz帯域送受信兼用型圧電高分子トランスデューサによる空気中での集積回路内部の検出について

Detecting the inside of Integrated Circuit by using one concave type P(VDF/TrFE)-transducer transmitting/receiving in air at MHz range.

○高橋貞幸(山形大学)

本研究では、前回の報告から更に発展させ、MHz 帯域での送受信兼用型方式(1 つのトランスデューサ)による、空中超音波画像[物質内部(主として集積回路/IC_11229-12/ROCKWELL)]の透過画像形成について報告する。従来のとおり使用したトランスデューサは、圧電高分子材料の P(VDF/TrFE)である。高分解能の画像を得るため、送信周波数は、約 2 MHz で行った。高周波数に伴う減衰の他、媒質(空気)や観測対象物により、超音波エネルギの減衰値は極度に増大する。送受信兼用型方式の場合は、観測対象物表面からの受信波も考慮する必要があり、加えて、超音波送受信システムで発生するノイズの影響も無視できない(受信波がノイズ中に埋没する現象も発生する)。本研究では、凹面型の P(VDF/TrFE)トランスデューサを用いた送受信兼用型透過方式(Figure 1)により IC の透過画像を試みたことを報告する。

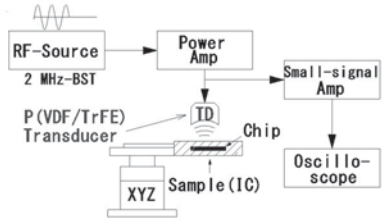


Figure 1 shows ultrasonic system to transmit/receive by the one concave type P(VDF/TrFE) transducer.

1-11-14

1-11-14　HIFU トランスデューサ加振によって生じた変位の推定手法に関する基礎検討

Basic investigation on an estimation technique for particle velocity induced by HIFU transducer excitation

○長岡亮(富山大・工)，吉澤晋(東北大院・工学研)，
梅村晋一郎(東北大院・医工学研)，長谷川英之(富山大・工)

集束超音波(HIFU: High-Intensity Focused Ultrasound)治療において，HIFU 用トランスデューサを加振源として用い，計測部位に音響放射圧を加えることで，横波やひずみを生じさせ，その横波速度やひずみ分布に基づいた治療範囲の評価手法が検討されている．また，加振音圧の強度を時間的に変化させることで，計測対象に任意の周波数成分を有する変位を生じさせ，その振幅の比に基づいた治療範囲の評価方法も検討されている．いずれの先行研究に関しても，生じた変位をより正確に推定することが，治療範囲の評価方法の精度向上につながると考えられる．そこで，本発表では，一般的に変位推定に用いられる自己相関法と私たちの研究室で検討している位相を用いた 2 次元変位推定法とを比較する．また，HIFU 用トランスデューサを加振源として用いて，生じた変位を推定する．

1-11-15

1-11-15 医用超音波画像の包絡振幅統計解析における高輝度非スペックル成分の確率画像に基づく除去法の検討

Removal of high-intensity non-speckle signals based on non-speckle probability image in envelope statistics analysis for medical ultrasound image

○森 翔平（東北大院・工学研），吉澤 晋（東北大院・工学研／東北大院・医工学研／ソニア・セラピューティクス）

◆医用超音波画像の包絡振幅統計解析において不要な信号である非スペックル（NS）信号を除去する手法を検討した。

◆NS 信号のみを抽出・除去することを目的に，NS 信号確率画像を算出し，一定確率以上の信号のみを NS 信号として除去した。本手法について，Field II による超音波送受シミュレーションにより検証した。

◆NS 信号確率が高値となる位置は，NS 構造体を配置した位置と概ね対応した（Fig. 1）。

◆従来法では NS 信号の影響を低減できなかった解析窓においても，提案法により周囲の解析窓の情報を用いることで NS 信号の影響を低減できる可能性が示唆された。

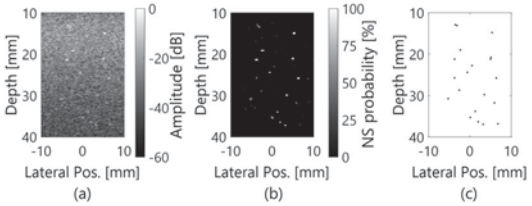


Fig.1 : (a) B-mode image. (b) Probability image of non-speckle (NS) signals. (c) Positions of NS structures.

1-11-16

1-11-16　超解像振幅包絡統計解析によるリンパ浮腫マウスの組織性状評価の基礎検討

Basic study on tissue characterization of lymphedema mise using super-resolution echo-envelope statistical analysis

◎奈良千尋, Joceline Theda Kadarman, 秋田 新介, 三川 信之（千葉大），Cameron Hoerig, Jonathan Mamou（Weill Cornell Med.），森 翔平（東北大），吉田 憲司, 山口 匡, 平田 慎之介（千葉大）

◆超音波画像の振幅包絡統計（ES）解析において，小さな ROI でも安定した結果を得るために，同一部位を撮像した複数の超音波画像を用いて，解析する ROI 内の画素数を増やす手法を提案している．

◆後肢リンパ浮腫モデルマウスが回復していく様子を超音波で経過観察し，真皮層において ES 解析を行った．±5°の範囲を 1°間隔で送信角度を変化させた平面波を用いて，提案手法による 3 次モーメントと各角度で算出した 3 次モーメントを比較した．

◆Fig. 1(a)の 0°では 3 次モーメントの四分位範囲の平均が 0.085 であるのに対し，(b)の提案手法では 0.045 となっており，提案手法により安定した ES 解析が行えたことが確認できた．

◆日数経過による 3 次モーメントの値の変化はあまりみられなかったが，今回は若い週齢のマウスを用いたため真皮の線維化がほとんど起こらず，組織性状の変化が少ないためであると考えられる．

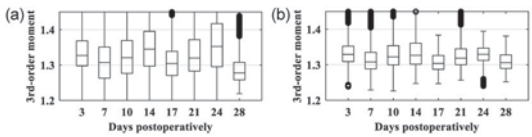


Fig. 1 Box-and-whisker plot of 3rd-order-moments of steering angles of 0° (a), and proposed method (b).

1-11-17

1-11-17　超音波顕微鏡のための生体組織音速と周波数依存減衰の推定

Estimation of sound speed and frequency dependent attenuation of biological tissue for ultrasonic microscopy.

☆粕谷美咲, 田中直彦（芝浦工大院）

◆生体組織用超音波顕微鏡のための，組織の音速と周波数依存減衰を推定する方法について検討している．

◆生体組織の音速は，Fig. 1 に示すように，試料のない場合のステージからのエコーを基準とした，試料表面と背面のエコーの伝搬時間差から求められる．また，試料背面からのエコーは，試料の周波数依存減衰の影響を受けている．

◆伝搬時間差と周波数依存減衰は，エコー波形より計算される正規化スペクトルから推定する．本研究では，正規化スペクトルのモデルとして次式を用いた．

$$R(n) = {A_1}^2\left({k_1}^2 e^{j\theta_1}\right)^n + {A_2}^2\left({k_2}^2 e^{j\theta_2}\right)^n$$

$R(n)$：正規化スペクトル，A_1, A_2：振幅，k_1, k_2：周波数依存減衰，θ_1, θ_2：伝搬時間差に対応した位相回転

◆計算機シミュレーションにより，最急降下法を用いて６個の音響パラメータ$(A_1, A_2, k_1, k_2, \theta_1, \theta_2)$が推定できることを確認した．

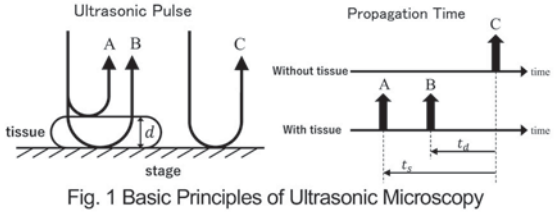


Fig. 1 Basic Principles of Ultrasonic Microscopy

1-11-18

1-11-18　高周波帯域プローブを想定した超音波顕微鏡による二次元音速分布の計測

Two-dimensional sound speed distribution measurement by ultrasonic microscopy assuming a high-frequency range probe

☆小林玲奈[1]，小林和人[2]，荒川元孝[1]（[1]東北大，[2]本多電子）

我々は，超音波顕微鏡を用いて，高周波数帯域における高解像度な音響特性分布の計測方法を検討している．しかし，プローブの帯域制限により低周波領域の情報が欠落する．そこで，正規化振幅スペクトルの周期性を利用し，低周波領域の情報を推定することで音速を算出した．本報では，中心周波数 250 MHz の超音波プローブを用いて，ヤギの血管を薄切した組織切片の二次元音速分布を計測した．その結果を Fig. 1(a)に示す．柔らかい領域(点 A)と硬い領域 (点 B)を比較すると，点 B は点 A に比べ，周期性が低く，周期の算出が困難だった．これは，より高い周波数で計測した場合の，減衰の影響が大きい場合と類似していると考える．今後，周期性が低い場合の解析方法を検討し，高周波帯域でより適切な音速の算出を目指す．

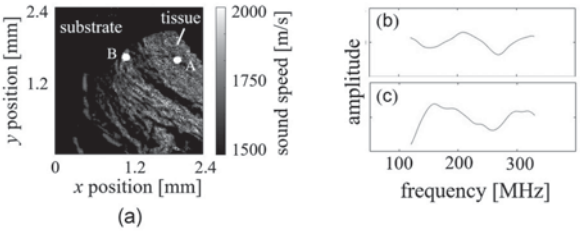


Fig.1. Sound speed of blood vessel . (a) Two-dimensional distribution, (b) normalized amplitude spectrum at point A , and (c) normalized amplitude spectrum at point B.

1-11-19

1-11-19　オリフィス板による負荷を備えた液体ピストン熱音響エンジンの熱力学モデルと実験による検証

Thermodynamic Modeling and Experimental Validation of a Liquid-Piston Thermoacoustic Engine with Orifice Load

☆猪鼻康祐，富平嗣大，△鈴木爽太，△楠戸宏城，△庄司衛太，琵琶哲志（東北大）

◆熱音響エンジンの基本設計方法の確立を目指している．

◆本研究では，熱交換器と作動流体の熱伝達を考慮することで，各部の温度と出力を関連付けるモデルを構築した．

◆往復流タービンの負荷を模擬したオリフィス板を備えた液体ピストンエンジンを用例として実験的にモデルの妥当性を検証した．

◆モデルは，少数のパラメータだけで実験装置各部の温度および出力を再現した．

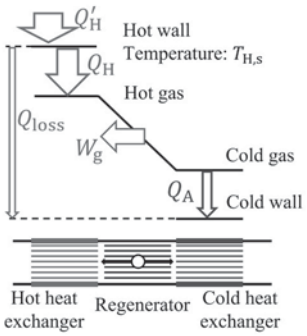


Fig.1 : Schematic diagram of the thermodynamic model.

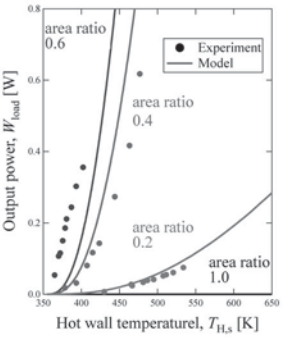


Fig.2 : Relationship between the hot solid temperature and the output power with various area ratios of orifice.

1-11-20

1-11-20　粒子群最適化法を用いた熱音響冷凍機の設計

Design of thermoacoustic refrigerator using particle swarm optimization

☆大友武，倉井陽太郎，上田祐樹（農工大）

◆熱音響冷凍機は音波による冷媒の温度変化を利用する冷凍機である．装置構造の単純さ，ノンフロン，理論効率の高さという特徴から，低コストで環境にやさしい冷凍機として注目されている．図１に示す進行波型熱音響冷凍機は，音響ドライバ，枝管，ループ管で構成されており，ループ管内には熱交換器に挟まれた蓄冷器が設置されている．

◆様々な進行波型熱音響冷凍機の研究は盛んに行われているが，その多くはループ管部分のみに注目した装置最適化についてのものであり，音響ドライバや枝管，ループ管を同時最適化した例はない．

◆本研究では粒子群最適化法（PSO）を用いて，音響ドライバ，枝管，ループ管で構成される進行波型熱音響冷凍機の設計パラメータを最適化した．また，その結果に基づいて大気圧空気を作動流体とする冷凍機を作成し，その性能を評価した．冷凍機の性能を表す冷凍出力と消費電力の比である COP_{total} の実験値は 0.5 となり計算結果の 0.9 と比べて小さい値となった．

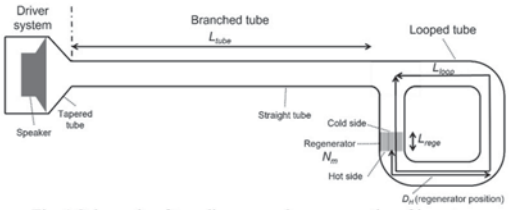


Fig. 1 Schematic of traveling-wave thermoacoustic refrigerator

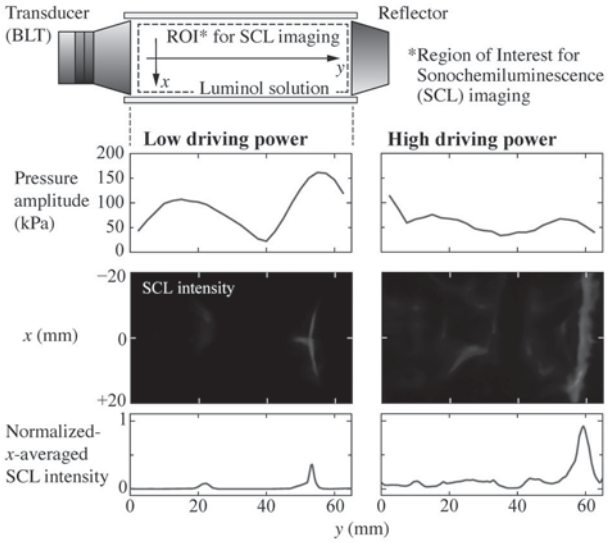
1-11-21

1-11-21　定在波型ソノリアクタにおける音圧分布の振動子駆動電力依存性計測

Measurement of dependence of sound pressure distribution on transducer driving power in standing-wave sonoreactor

○黒山 喬允，小笠原 英子，森 和義（防衛大）

- □ 定在波型ソノリアクター内の音圧分布を光偏向法により計測
- □ 計測した音圧分布をよるソノケミルミネッセンスの強度分布と比較
- ■ 低振動子駆動電力では音圧の腹節が見られ音場は定在波的音圧の腹でソノケミルミネッセンス(SCL)を検出
- ■ 高振動子駆動電力では音圧の腹節が見られず音場は進行波的気泡によるシールディングで音圧は低下したが，SCLを検出



1-Q-1

1-Q-1　Bark 尺度の音圧依存性

Level dependency on Bark scale

○岡崎聡（富山県立大）

◆人の聴覚は周波数軸上に帯域通過フィルタが並ぶフィルタ群で模擬される。

- ➢ 中心周波数によって異なる臨界帯域幅の間隔を等間隔に変換した周波数尺度を CB-rate（単位は Bark）あるいは Bark 尺度という。

◆２純音が同時に開始したと知覚される同時性の範囲は周波数距離の関数としてＶ字曲線に回帰する。

- ➢ Ｖ字の折点の位置と２音の音圧の関係は
  - ✧ 65 dB（A）のとき 0.5 Bark【これまでの知見】
  - ✧ 35 dB（A）のとき 0.3 Bark【本研究】

◆聴覚フィルタは音圧低下により帯域幅が狭くなる。

- ➢ すなわち、Bark 尺度も音圧低下により縮小される。
- ➢ 縮小は 65 dB→35 dB のとき、0.3 / 0.5 = 0.6 倍であることが示唆される。

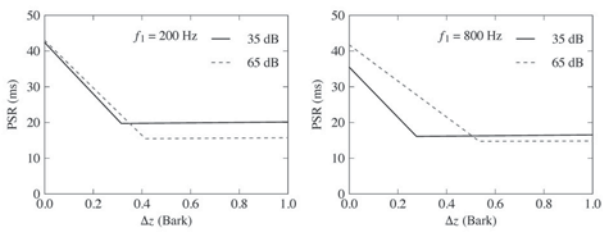


Fig. 1 The breakpoint of the perceptual simultaneity range V-curve.

1-Q-2

1-Q-2 電車内における音楽聴取レベルと環境騒音の関係の実測的検討

An Empirical Study on the Relationship Between Music Listening Levels and Environmental Noise Inside Commuter Trains

☆平國弘督，渡邉祐子，△伊藤俊介（東京電機大学・デザエ）

◆近年、イヤホンやヘッドホンを長時間・高音量で使用する若年層において，「イヤホン難聴」と呼ばれる聴力低下リスクが指摘されている。

◆本研究では若年層の音楽聴取実態を調査することを目的とし、首都圏の大学生を対象に通学時の電車内で音楽再生音圧レベルと環境騒音の音圧レベルの測定を行った。被験者が測定中に音量変更を自由に行うことで得られた変更回数やタイミング、レベル変化を用いて、聴取行動の違いに着目し、傾向や特徴について分析をした。

◆その結果、聴取行動として音楽再生音圧レベルを一定に保つA群と環境騒音の変動に応じて音楽再生音圧レベルが変動するB群と環境騒音の変動とは関係なく音楽再生音圧レベルが変動するC群の3群に分類できた。

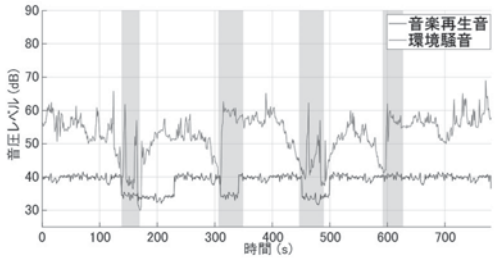


図1：B群の被験者の例

1-Q-3

1-Q-3 ASMR 音源の音響的加工に伴う聴取印象の変化とその音響的要因の分析

Analysis of changes of perceptual impression and acoustic factors induced by acoustic modification of ASMR sounds

☆森嵜隼一郎，坂野秀樹，旭健作（名城大院・理工学研）

◆**ASMR（自律感覚絶頂反応）**とは，人が聴覚や視覚への刺激によって感じる，**心地よい**，**脳がゾワゾワする**といった反応・感覚のことである。

◆本研究では，ASMR に寄与する音響的要因の解明を目的として，**重畳処理**によって音信号が定常的なノイズに近づき，時間的変動や両耳間の音響差異が減衰する条件下においても ASMR 効果が知覚されるかを検証する実験を行った。

◆主観評価と音響分析の結果，重畳による予測困難性の増加はゾワゾワ感を高め，特に咀嚼音や新聞紙のように**不規則な音の動き**が含まれる音源でその傾向が顕著に見られた。一方で，鈴のように SPL の変動が小さい音源では重畳によって刺激が単調化し，ゾワゾワ感が減少する傾向が示された。

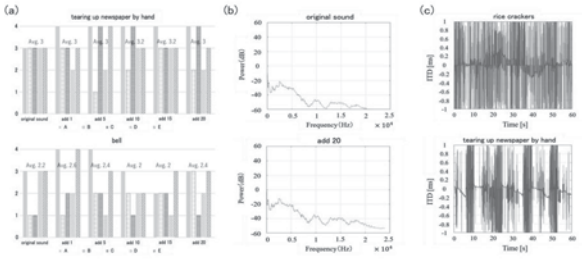


Fig. 1: (a) Results of subjective evaluation experiment, (b) Long-term average spectrum of rice crackers, (c) ITD of the original sounds

1-Q-4

1-Q-4 音質評価指標とスペクトロテンポラルモジュレーションの関係についての検討

Relationship between sound quality metrics and spectrum-temporal modulation

◎磯山拓都（都立高専），鵜木祐史（北陸先端大）

【背景】音質評価指標（SQMs）は音の快・不快を見積もることができるが，その値は音信号の時間・周波数両方向の音響特徴に依存する．スペクトル時間変調（STM）分析は時間・周波数両方向の音響特徴を同時に把握でき，SQMs の特性を探るのに有効であると考えられる

【目的】STM スペクトル平面のどの領域が SQMs にとって重要であるかを明らかにすること

【方法】STM と SQMs の関係を調べるための３つのステップ

１．STM 分析による STM 特徴の抽出　２．計算モデルによる SQMs の計算　３．統計的解析による STM と SQMs の関係性の検討

【結果】Fig.1 より，ラウドネスとシャープネス：TM が 0 Hz，SM が 0 cycle/ERB 付近の成分，ラフネス：TM が 70 Hz 付近，SM が 0 cycle/ERB 付近の成分，変動強度：TM が 0 から 5 Hz 付近，SM が 0 cycle/ERB 付近の成分が重要であることがわかった．

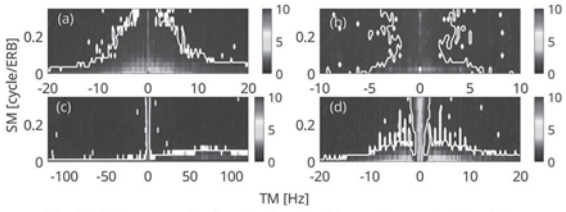


Fig.1：VIP scores obtained using SQMs as dependent variables: (a) loudness, (b) sharpness, (c) roughness, (d) fluctuation strength.

1-Q-5

1-Q-5 DNN による非侵入型 Cross-Domain 評価モデルの開発 ～難聴音声了解度予測及び音質評価～

A Novel Deep Neural Model on Prediction for Speech Intelligibility and Quality for Persons with Hearing Loss

○時政和征，羅文煜，孫浩正，入野俊夫，陳金輝（和歌山大）

◆高齢難聴者の音声了解度を予測できる客観評価指標の開発を目指している。補聴器等の信号処理技術開発においては、音声了解度だけでなく音声の品質も同時に評価できることが望ましい。

◆GESI や PESQ は、高精度な音声了解度・音質評価が可能な指標であるが、基準音声を必要とする侵入型であり、さまざまな場面での応用がしにくい。

◆このような背景から、本研究では、深層学習(DNN)を用いることにより、基準音を必要としない音声了解度予測および音質評価を可能とする非侵入型 DeepPIQ の開発を行った。

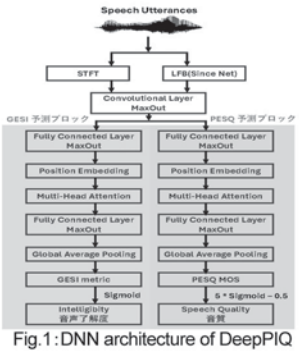


Fig.1：DNN architecture of DeepPIQ

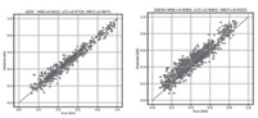


Fig.2：Speech intelligibility prediction results of DeepPIQ on CPC2

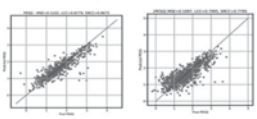


Fig.3：Speech quality prediction results of DeepPIQ on CPC2

1-Q-6

1-Q-6　健聴者の雑音下文聴取閾値の推定 ～年齢，純音聴力レベル，時間分解能指標を用いた検討～

Estimation of speech reception thresholds in noise for normal-hearing individuals: A study using age, pure-tone hearing levels, and temporal resolution indices.

◎岡龍也，森本隆司(リオン)，山本弥生(国際医療福祉大)，△小渕千絵(筑波大)，△岡本康秀(済生会中央病院/慶大)，△神崎晶(東京医療センター)，森周司(九州大)

◆これまで標準検査である標準純音聴力検査より得られる純音聴力レベルおよび被験者の年齢より，雑音下語音聴取能を推定することを検討してきた（NF model）.

◆本報告では，若年および高齢健聴者のデータを用いて EHF 純音聴力レベルを加えたモデル（EHF model），2 種類の時間分解能指標（GDT，TMTF）をそれぞれ加えたモデル(GDT model, TMTF model)を評価し，NF model に対し調整済み決定係数は向上することが分かった.

◆一方，推定精度を若年健聴者のデータを用いて評価したところ NF model が今回評価した model のいずれよりも高かった.

Table 1 : Results of multiple regression analysis

Regression coefficients		第1主成分	第2主成分	第3主成分	第4主成分	第5主成分			Intercept	Adj. R^2	p
EHF model	Coeff	0.752	-0.359	-	-1.604	-			-4.824	0.701	***
	p	***		-	*	-			***		
Regression coefficients		**$HL_{0.125}$**	**$HL_{2.0}$**	**$HL_{4.0}$**	**GDT**	**TMTF PS (dB)**	**TMTF CF (Hz)**	**Age**	**Intercept**	**Adj. R^2**	**p**
NF model (2024)	VIF	3.894	1.892	1.560	-	-	-	1.720	-		
	Coeff	0.025	0.029	0.033	-	-	-	0.046	-7.717	0.405	***
	p			***				***	***		
GDT model	VIF	2.937	2.668	3.175	1.301	-	-	7.273	-		
	Coeff	0.028	-0.056	0.015	1.914	-	-	0.071	-8.610	0.778	***
	p		.		.			***	***		
TMTF model	VIF	-	2.893	3.084	-	1.714	-	6.552	-		
	Coeff	-	-0.048	0.0191	-	0.104	-	0.071	-5.283	0.774	***
	p	-		.				***	**		

1-Q-7

1-Q-7　騒音環境における難聴者のための音声強調処理モデル GESI-GAN

Speech Enhancement Model "GESI-GAN" for Hearing-Impaired Listeners in Noisy Environments

☆孫 浩正，羅 文煜，時政 和征，山本 絢子(和歌山大院)
陳 訓泉(県立広島大)，入野 俊夫，陳 金輝(和歌山大)

◆本研究では、難聴者のため、深層学習に基づく新たな音声強調モデル「GESI-GAN」を提案する。模擬難聴音声と雑音を入力とし、了解度（GESI）および音質（PESQ）の両指標を同時に考慮した GAN ベースにより学習を行う。

◆提案モデルは 1 つの生成器と 2 つの識別器から構成され、参照音声も導入して、学習の安定化を実現した。日本語の模擬難聴音声データセットを用いた実証実験では、その有効性が確認された。

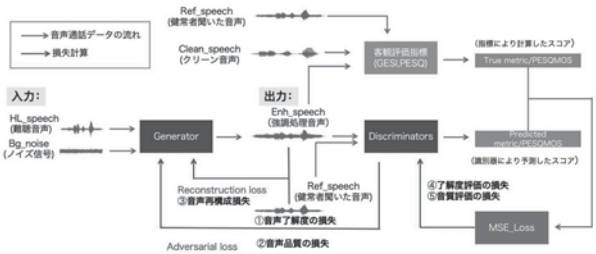


Fig.1 : Overall architecture of the proposed model GESI-GAN

1-Q-8

1-Q-8　DeepGESI:難聴音声了解度を予測するための非侵入型評価モデル

DeepGESI: A Deep Model on Intelligibility Prediction for the Hearing-Aid Speech

☆羅 文煜，孫 浩正，時政 和征，山本 絢子（和歌山大院）
陳 訓泉(県立広島大)，入野 俊夫，陳 金輝(和歌山大)

◆本研究では深層学習を用いて難聴者の音声了解度を高精度で効率的予測できる非侵入型モデル DeepGESI を提案する。

◆DeepGESI は難聴音声に対する予測精度が、GESI に遜色ない。

◆DeepGESI は非侵入型（non-intrusive）の計算方式で、基準音声を必要としない。

◆DeepGESI は軽量なモデル構成であり、リアルタイムで難聴音声の了解度を推定可能とする。

◆DeepGESI は文音声や複雑音響環境における会話音声にも適用できる。

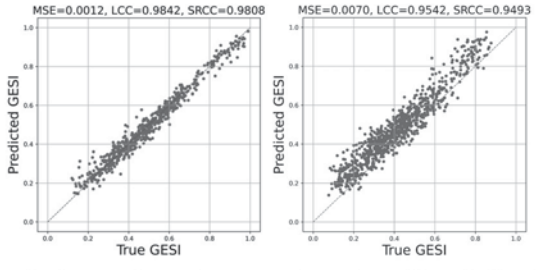


Fig.1 : Scatter plots of speech assessment predictions of DeepGESI on the Seen (left) and Unseen (right) datasets.

1-Q-9

1-Q-9　模擬難聴環境下での少ない評価音数での音声了解度推定関数導出

Derivation of speech intelligibility estimation functions under simulated hearing-loss conditions with limited speech samples.

〇小林洋介(室工大)

◆音声了解度は、恒常法による心理物理実験として、聴覚障害や補聴器の評価、音声通信システムの評価、拡声システム、建築音響での評価に広く利用されている。しかし、恒常法では評価刺激数が多くなる傾向がある。評価時間短縮を意図した主観評価法は種々提案されているが、50%刺激閾値の SRT 導出に特化する傾向にある。

◆著者らは、SRT 導出に用いる物理量の変化から主観評価値を推定する音声了解度推定関数（心理測定関数）を、少ない評価音数（刺激数）から導出することを目指している。

◆本報告では、模擬難聴システム WHIS を用いて作成した加齢性難聴模擬音を、20 代の実験参加者が評価する系において、了解度推定関数の導出を対象とした。

◆Fig. 1 に音声レベル別の評価結果とその推定を示す。恒常法による推定関数（青）と提案法の関数（赤）との SRT 差は最大でも 2.68 dB と、近い値であった。

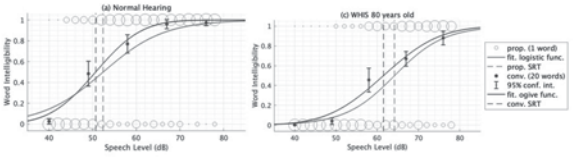


Fig. 1 Comparison of estimation functions by speech level.

1-Q-10

1-Q-10　Mamba を用いた難聴者のための音声了解度予測法の検討

Study of speech intelligibility prediction model for hearing-impaired listeners using Mamba

◎山本克彦，宮崎晃一（サイバーエージェント）

◆先行研究：Transformer を用いた音声了解度の予測モデル
- 利点：リファレンス信号不要で高い予測性能
- 問題点：自己注意機構で比較的大きな計算量とメモリ

◆提案手法：先行研究の時間情報変換ブロックに Mamba を導入
- 実時間で動作しながら音声信号の長距離依存関係を維持
- Transformer よりも低計算能力・低消費電力で動作

◆片耳・両耳のモデルによる実験結果：
- 少ないパラメータ数で先行研究と同等以上の予測性能

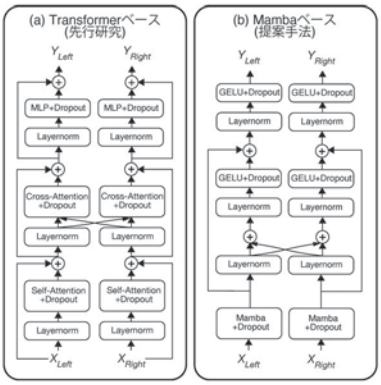


Figure 1: Binaural temporal transform block by Transformer-based model (a), and the Mamba-based model we proposed (b).

1-Q-11

1-Q-11　Integrating Frequency and Temporal Resolution into Speech Intelligibility Prediction for Hearing-Impaired Listeners

☆ Xiajie Zhou, Candy Olivia Mawalim, Masashi Unoki (JAIST)

◆ **Aim：** To investigate how simulating frequency and temporal resolution deficits caused by different types of hearing loss can improve the prediction of speech intelligibility.

◆ **Problem:** Hearing-impaired listeners often experience reduced frequency and temporal resolution, impairing their ability to understand speech, particularly in noisy environments.

◆ **Solution:** Frequency and temporal resolution are simulated according to different types of hearing loss. Frequency resolution is modeled by broadening the bandwidths of the Gammatone filterbanks, and temporal resolution by low-pass filtering with cutoff frequencies reflecting reduced temporal sensitivity.

◆ **Evaluation:** The prediction methods were assessed using the Clarity challenge data with two metrics: Pearson correlation coefficient (ρ) and root-mean-squared error (RMSE).

◆ **Summary:** Compared with HASPI, the proposed method improved ρ by 7.3% and reduced RMSE by 15.1%.

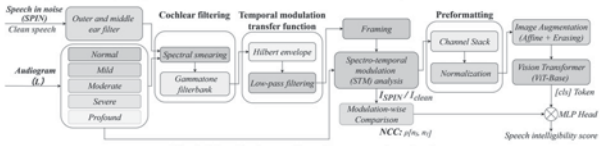


Fig.1: Simplified overview of proposed method

1-Q-12

1-Q-12　騒音下でのロンバード音声が模擬難聴状態での音声知覚能力に与える影響の調査

Investigating the effect of Lombard speech under noisy conditions on speech perception ability in simulated hearing loss conditions

☆寺村実優，坂野秀樹（名城大院）

◆本研究では，騒音レベルの大きいような場所で自然と大きな声を出してしまうロンバード音声に注目し，ロンバード効果を含んだ発声を模擬難聴状態で知覚した際の音声知覚能力を調査する．

◆主観評価実験の 4 モーラ単語書き取り実験における正答率の結果を Fig. 1 に示す．

◆Lom85_WHIS の単語正答率が一番低くなっており，この正答率を下げた原因は 1 モーラ目の正答率が低いことであると考えられる．

◆1 モーラ目の正答率が低くなった理由として音声の立ち上がり部分が聞き取りづらい可能性が示唆されたが，これに関してはより詳細な分析が必要であると考える．

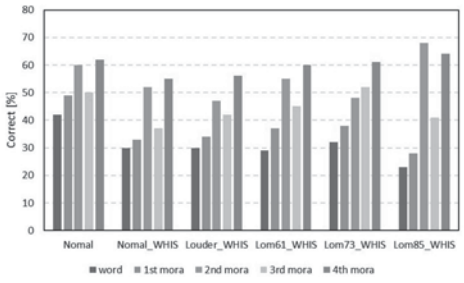


Fig.1：Correct answer rate for each word and mora

1-Q-13

1-Q-13　骨導無発声音声の明瞭度向上に向けた音声強調システムの検討

Investigation of Speech Enhancement Systems for Improving Intelligibility of Bone-Conducted Whispered Speech

☆草田晃彦，小澤賢司，鳥谷輝樹（山梨大）

◆**背景**：無発声音声(ささやき声)の明瞭度を向上させる技術
→声帯摘出者のコミュニケーション手段として有効

◆**目的**：骨導マイクロホンで収音した無発声音声の明瞭度向上

◆**方法**：目標となる気導通常音声と同期した人工骨導無発声音声を合成することで，教師あり音声強調システム(Fig. 1)を適用する．

◆**結果**：骨導無発声音声と人工骨導無発声音声の異聴表における異聴傾向の差分は，全体的に 0 に近い値
人工骨導無発声音声の強調音声では，一部の被験者において，母音の明瞭度が向上(Fig. 2)→全ての母音で一定の正解率

◆**考察**：子音は信号強度が弱いことから，骨導経路での信号の減衰が考えられる．

◆**今後**：実際の骨導無発声音声の明瞭度向上を目指す．

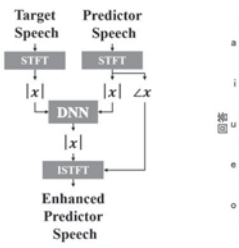


Fig. 1: Overview of the Speech Enhancement System

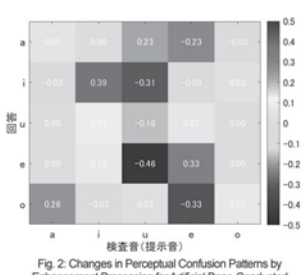


Fig. 2: Changes in Perceptual Confusion Patterns by Enhancement Processing for Artificial Bone-Conducted Whispered Speech: Vowels (Subject 3)

1-Q-14

講演取消

1-Q-15

1-Q-15　空間知覚機能の計測を目的としたスピーカマトリクスシステムの開発

Development of Loudspeaker Matrix System for Measuring Spatial Perception Function,

○藪 謙一郎(東大), 三浦 貴大(産総研), 伊福部 達(東大)

◆世界的な高齢化のなか、高齢者の認知機能や身体機能を簡易的に把握するための計測法へのニーズや関心が高まっている。

◆視覚に障害のある高齢者は支援の必要性が高いにもかかわらず、ワーキングメモリの視空間スケッチパッドなどの視覚を用いる認知機能の検査が適用できないことが、問題の早期発見の障壁となる。

◆発表では、視覚障害者のための空間情報の代替提示方法としてスピーカマトリクスを利用して、音源定位能力や空間知覚機能を定量的に測定できるようにするためのシステムを提案し、システムの試作を行ったので報告する。

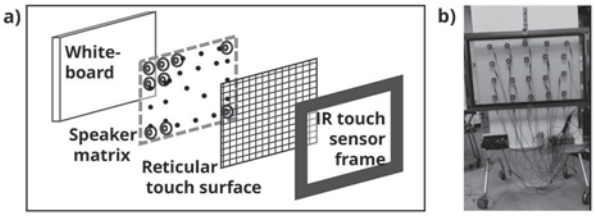


Fig.1: Configuration of System and Prototype

1-Q-17

1-Q-17　正面 0.5m にある音源の放射方向知覚の傾向調査

Investigation of trends in perception of radiation direction of sound sources located 0.5m in front.

☆大島若菜（阪産大･デザインエ), △米澤杏瞳（阪産大･デザインエ),
中山雅人（立命大･情理), 高橋徹（阪産大･情報デザイン）

【背景】聴取者付近の音源の放射方向は，高臨場感再生に必要な特徴である．放射方向の知覚傾向を明らかにすれば，高臨場感再生の音源制御に必要な放射方向の角度解像度を知ることができる．

【目的】正面 0.5m に位置する発話者の頭部方向を聴取者がどのように知覚するかに注目し，知覚上の角度分解能を調査する．

【手段】頭部方向を 30 度ごと変化させた刺激音をバイノーラルマイクロホンで収録した(Fig.1)．これら刺激音から任意ペアを作成し，被験者に提示した．刺激音ペアから受ける印象を 5 名に評価させた．聴取者に近い向きに発話していると感じる方を強制選択する課題を与えた．

【結果】-30 から 30 度付近では区別がつきにくい傾向が示唆された (Fig.2).

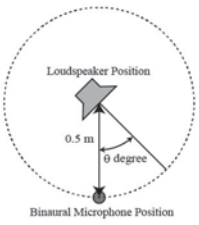


Fig.1:Layout of loudspeaker and binaural microphone.

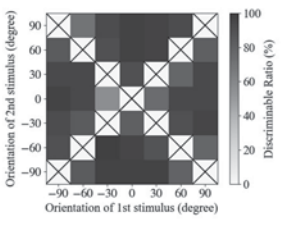


Fig.2:Average discriminable ratio.

1-Q-18

1-Q-18　ヘッドセットおよびダミーヘッド録音音声における距離知覚の比較

Comparison of Perceived Distance between Headset and Dummy Head Speech Recordings.

☆佐藤樹, 小林洋介, 寺岡諒(室蘭工大)

◆本研究では，発話者が聴取者との距離に応じた適応的な発話の変化を分析するため，発話者の口元にあるヘッドセット (HS) によるモノラル音声と聴き手を模擬したダミーヘッド (DH) によるバイノーラル録音音声とで発話距離に関する主観評価を行い，録音方式を比較する。

◆Fig. 1 に録音時の機器構成を示す。最初に校正音を録音し，次に主観評価に用いる対話を模擬した音声を HS と DH で同時に録音した。

◆主観評価は，録音した発話音声（５文章）を HS および DH で録音した音声刺激とした。参加者は男女 10 名（平均年齢 21.2 歳）で，2.0 m を基準として提示し，知覚した距離を回答してもらった。

◆Fig. 2 より，1.0 m 以内の近距離では，DH 刺激が HS 刺激より平均で 0.72 m 大きな RMSE を示した。これは，DH 刺激が持つ空間的手がかりが距離知覚に影響を与えた可能性を示唆する。

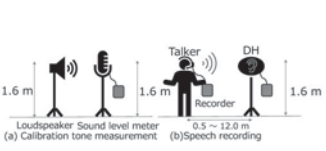


Fig.1:Recording setting.

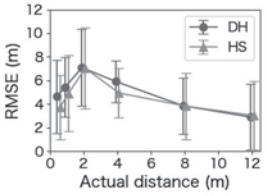


Fig.2:RMSE of Perceived Distance.

1-Q-19

1-Q-19 聴覚―発話運動制御における母語の影響：フォルマント補償応答と母音分布との関係

Influence of native language on auditory-speech motor control: Relationship between formant compensation response and vowel distribution

○上江洲安史，村上 正悟，赤木 正人，鵜木 祐史（JAIST）

◆**目的**：フォルマント変形聴覚フィードバック（AAF）下で与えた摂動と母音分布の関係を定量化し、F1 補償応答との関係を調査した。

◆**手法**：日本語・中国語・ベトナム語話者各 10 名を対象に、母音/e/に対する F1 摂動 AAF 実験を実施した。各言語の短母音分布を F1-F2 平面で分析し、摂動と分布の関係を定量化する「影響度」を算出した。

◆**結果**：F1-F2 平面上の母音分布は各母語で特徴的な様子を示し、AAF 補償量もまた母語ごとに異なる様子を示した。算出された影響度は、この補償量の傾向とおおむね一致した。

◆**考察**：各言語特有の母音分布構造がフォルマント AAF 摂動に対する補償量を規定し、母音マグネット理論に基づく母語による弁別感度の差異が聴覚―発話運動制御に反映される可能性を示唆する。

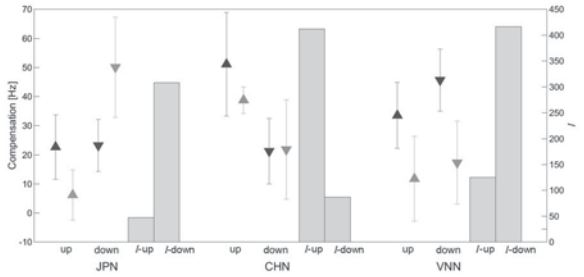


Fig.1: F1 compensation amounts (△: upward shift, ▽: downward shift, dark color: native language, light color: English) and influence index / (upward and downward) for each language group.

1-Q-20

1-Q-20　視聴覚統合が残響誘発音声の明瞭度に与える影響

The effect of congruence of virtual image and reverberation-induced speech on speech intelligibility

○程島奈緒，濱本和彦(東海大)，水町光徳(九工大)，榊原健一(北海道医療大)

◆視覚情報が CG である視聴覚情報の一致度が、残響下の残響誘発音声の音声明瞭度に与える影響を調査した

◆Table 1 の 5 条件で聴取実験を行った結果を Fig. 1 に示す

◆視聴覚情報が一致した残響誘発音声の明瞭度＞静音下で発話された音声・視聴覚が不一致の残響誘発音声の明瞭度

◆視覚情報が CG である視聴覚情報の一致度は、音声明瞭度に影響を与えなかった

Table 1 Experimental conditions

Condition	Visual Condition	Audio Condition	RT (s) Recording	RT (s) Listening test
A		Without reverberation		1.5
B				12
C (congruent)	Larger room	Reverberation -induced	12	12
D (incongruent)			1.5	1.5
E (incongruent)	Smaller room		12	12

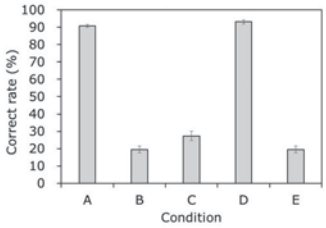


Fig.1: Mean correct rate and standard error of the target words.

1-Q-21

1-Q-21　The Origin of Sound Symbolism: an Articulatory-based Approach

○ Tess Spiteri, △Hidekazu Nagamura, Kohta I. Kobayasi (Doshisha Uni., Fac.Medical and Life Sciences)

◆Sound symbolism involves non-arbitrary links between speech sounds and perceptual meanings. While well studied through perception-based tasks (e.g., forced-choice, ratings), its articulatory basis remains underexplored.

◆This study introduces a production-based method using articulatory synthesis, grounded in the motor theory of speech perception—which suggests that speech is processed through inferred articulatory gestures, not just acoustics.

◆23 native Japanese speakers used a slider interface to adjust, in real-time, 7 articulatory parameters randomly (e.g., tongue height, lip rounding) and generate synthetic stimuli corresponding to the resulting vocal tract configuration, via an articulatory synthesis tool. Each trial targeted one of four concepts: small, large, round, or spiky.

◆Results showed clear trends for size contrasts, but weaker effects for shape. Vowel-only stimuli may not reliably trigger shape-symbolism. Some articulatory effects (e.g., velum position) may also reflect volume rather than symbolic association.

◆This work proposes a new framework for studying sound symbolism that bridges perceptual judgments and motor representations.

1-Q-25

1-Q-25　聴覚障害者の異なる速度での文章音読における発話音声の韻律的特徴（２）補聴器装用者を対象に

Prosodic features of speech sounds in the reading aloud of sentences at different speech rates by hearing-impaired people(2)：For hearing aid users

○湯浅哲也(大教大･教)

◆補聴器装用聴覚障害者４名を対象に，異なる速度での文章音読の音声を録音し，発話速度・音圧・基本周波数の変化を検討した。

◆その結果、以下の点が明らかになった。

◆①速度が遅くなるにつれて，１モーラ当りの持続時間が長くなること，発話時間よりポーズ時間の割合が大きくなることが示された。

◆②速い発話は音圧及び基本周波数が高くなる傾向が示された。

◆③人工内耳装用者の音声的特徴とは、異なることが明らかになった。

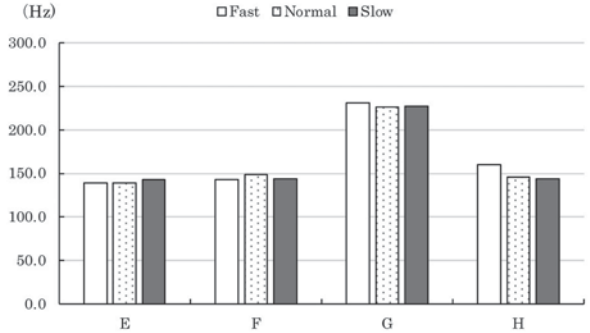


Fig.1: Fundamental frequency at different speech rates

1-Q-26

1-Q-26　加齢性音声障害の回復過程を可視化するインクルーシブデザイン型リハビリ支援システムの開発

Development of an Inclusive Design Rehabilitation Support System for Visualizing the Recovery Process of Age-Related Voice Disorders

☆三藤早葵，石光俊介（広島市立大学院）

◆加齢性発声障害(Presbyphonia)は高齢者のコミュニケーション能力を低下させるだけでなく，誤嚥性肺炎の一因ともなり，健康寿命に直結する.

◆本研究では，加齢性音声障害に対し，AI とインクルーシブデザインに基づくリハビリ支援システムの構築を目指す.

◆本報では，VFE(Vocal Function Exercise)訓練時の発声機能の変化や発声持続の経過時間の提示の有無での違いを調査する.

◆VFE が加齢性発声障害罹患者と男性女性健常者の発声持続時間と音階上昇下降の音高変化が多くの項目において改善された.

◆視覚情報の介入が発声機能改善に繋がることが示唆され，この知見もシステムの UI 構築の一助となる.

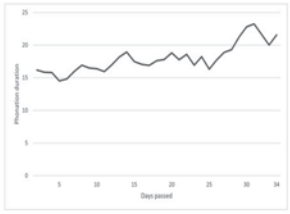


Fig.1：Patients with presbyphonia [Humming]

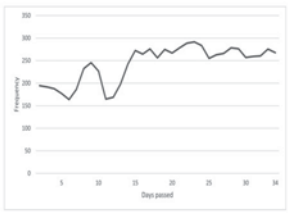


Fig.2：Patients with presbyphonia [Difference in pitch ascent]

1-Q-27

1-Q-27　パラメータ加重平均モデルによる明瞭な脊髄性筋萎縮症者音声の合成

Synthesis of Intelligible Speech for Individuals with Spinal Muscular Atrophy Using a Parameter-Weighted Averaging Model.

☆八木雄大[1]，高島遼一[2]，佐々木千穂[3]，滝口哲也[1]
([1]神戸大学，[2]立命館大学，[3]熊本保健科学大学)

◆本研究では，脊髄性筋萎縮症による構音障害を有する話者のコミュニケーションを支援するためのテキスト音声合成 (TTS) モデルの作成を目的とする。

◆脊髄性筋萎縮症者の収録音声を学習した TTS モデルは，構音障害の発話特徴を示す，コミュニケーション支援に向かない音声を合成するという課題がある。

◆本研究では，構音障害を有しない話者と脊髄性筋萎縮症者の TTS モデルの各パラメータを，モデルの構成要素ごとに加重平均し，新しい TTS モデル (パラメータ加重平均モデル) のパラメータとする手法を提案する。

◆実験では TTS モデルに VITS を使用してパラメータ加重平均モデルを作成し (Fig. 1)，その合成音声が声質を維持しつつ明瞭性の向上を達成しているかどうかを，客観評価およびアンケートによる主観評価を通じて検証した。

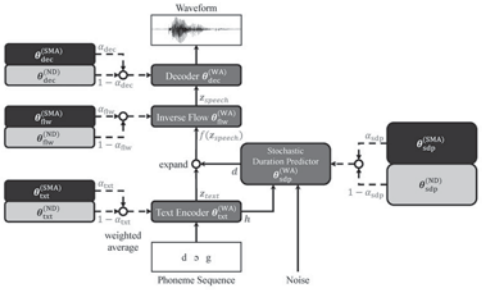


Fig. 1 Synthesis procedure using a VITS-based parameter-weighted averaging model.

1-Q-28

1-Q-28　選択的状態空間モデルを用いた long-form 音声認識の有効性に関する実験的評価

An experimental study on the effectiveness of long-form speech recognition using selective state space models

◎松井清彰，三村正人，松浦孝平，浅見太一（NTT）

◆選択的状態空間モデルの一種である Mamba を音声認識のエンコーダとして活用した Mambaformer について、カリキュラム学習を用いて段階的に学習データの発話長を長くしていくことで、long-form 音声認識に対する精度向上を目指した。

◆学習・評価セットとして、欧州議会の実音声を収録した EuroParl を使用。コーパスのセグメント情報に基づいてセグメント単位にデータ分割を行い、それをベースに発話長の異なるデータセットを作成した。

◆Mambaformer 17 層による実験では、発話長の異なるデータセットに対してカリキュラム学習を行うことで long-form 音声に対する認識精度を改善させることができたが、これらのデータを足し合わせたコーパスで学習を行った場合が最も高い精度を達成できた。(Fig. 1)

◆また、Conformer との比較では、約 60[%]以上計算量を削減しつつ、同等程度の認識精度を達成した。

curriculum	num_epoch			lr_dec_rate	WER[%] of MEP-test	
	concat1	concat2	concat3		segment	session
1	50	-	-	-	12.2	15.3
3	-	-	50	-	25.7	26.8
1→2	35	15	-	0.1	**12.0**	11.3
				0.5	12.1	12.4
				1.0	12.3	10.4
1→2→3	35	5	10	0.1	12.2	10.2
				0.5	12.1	10.0
				1.0	12.2	9.7
1+2+3	17			-	12.2	**9.1**

Fig.1：WER of Mambaformer using curriculum learning

1-Q-29

1-Q-29　アライメントを用いない Transducer ベース複数話者音声認識モデル構築の検討

A Study on Alignment-Free Training for Transducer-based Multi-Talker ASR

森谷 崇史，堀口 翔太，デルクロア マーク，増村 亮，
芦原 孝典，佐藤 宏，松浦 孝平，三村 正人（NTT）

◆本研究では複数話者の音声が含まれる混合音の各話者の音声を同時に認識する複数話者音声認識（MT-ASR）タスクを実施

◆本研究では Transducer（RNNT）ベースの MT-ASR モデルを構築

◆先行研究では各話者の出力ラベルの各トークンをタイムスタンプの順序にしたがって１つの系列に並び替えた正解データを用いて学習する手法（MT-RNNT-tSOT）を提案

× 時刻情報の取得には学習済みの音声認識システムを用いて学習データのアライメントの事前準備が必要

× 推論時は全ての話者の出力トークンが混在するため認識結果が複雑な形式となり，言語モデル（LM）を用いることが困難

◆本研究ではアライメントを用いずに学習可能であり，推論時は話者ごとの認識仮説が出力可能な手法（MT-RNNT-AFT）を提案

➢ 実験により MT-RNNT-AFT は MT-RNNT-tSOT と同程度の認識性能を達成し，提案手法である MT-RNNT-AFT の有効性を確認

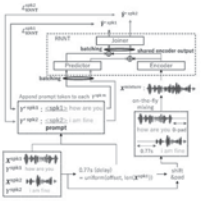


Figure 1 : Overview of the MT-RNNT-AFT pipeline

Table 1 : cpWERs [%] (1spk/2spk) of each MT-ASR system on LibriSpeechMix.

System	Offline	Streaming
Single-Talker ASR	2.7/64.5	4.1/66.6
MT-RNNT-tSOT	2.6/4.0	4.2/6.5
MT-RNNT-AFT+KD	2.6/3.7	4.1/6.7
+LM	2.4/3.4	4.0/6.3

1-Q-30

1-Q-30　大規模多言語 E2E 音声認識モデルに基づく CTC ブランチを活用した時刻アライメントの検討

Exploring Time Alignment Using the CTC Branch in a Large-Scale Multilingual End-to-End Speech Recognition Model.

◎柏木陽佑，二見颯，角尾衣未留，朝川智（ソニーグループ）

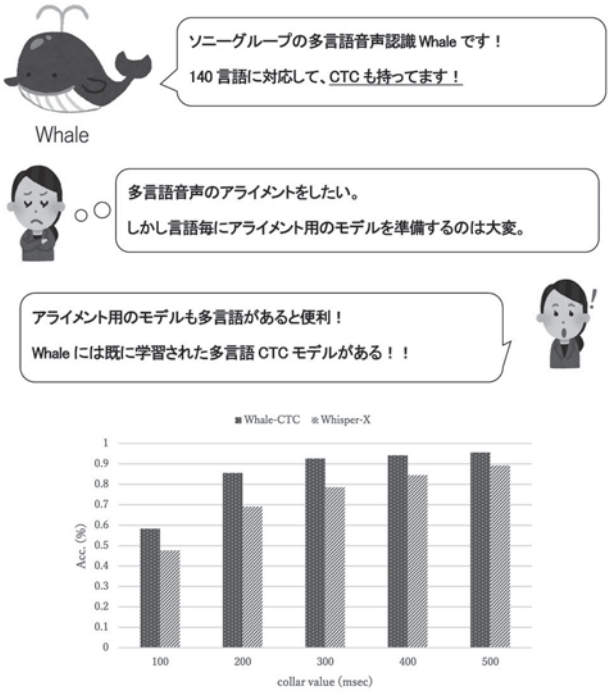


Fig.1：Timestamp accuracy on CSJ data.

1-Q-31

1-Q-31　音声の不変構造を活用した話者非依存な End-to-End 音声認識の検討

A Study of Speaker-Independent End-to-End Speech Recognition Utilizing the Invariant Structure

☆岡田翔太，齋藤大輔，峯松信明（東大）

◆幼児が限られた話者から言葉を学び、未知話者の言葉も認識できるようになることに着想

◆End-to-End 音声認識器において、未学習話者に対する認識精度向上手法の検討

◆話者非依存な言語的特徴を捉える「不変構造」と呼ばれる音声特徴量を活用

◆従来の Transformer の Attention 機構における内積計算と不変構造との類似性に着目し、新たな Attention 機構の構造を検討（Fig.1）

◆学習話者１名の条件で、一部の未学習話者に対する認識精度が向上

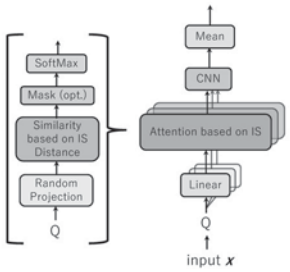


Fig.1: Proposed model architecture: direct feature extraction based on invariant structure (IS) similarity

1-Q-32

1-Q-32　ベイズ深層学習モデルを音声認識システムに適用

Application of Bayes' Deep Learning Model to Speech Recognition System.

☆坂井秀彌(県広大院)，折本寿子(県広大)

◆音声認識技術は近年飛躍的に向上しているが，雑音混入下で音声認識性能が著しく低下するという課題が存在する．

◆一般的には，雑音除去や音声前処理を音声認識前に行うことが多い．

◆本研究では音声認識の精度向上を目的として，深層学習モデルにベイズ推定を適用した，ベイズ深層学習モデルを提案する．

◆深層学習内で雑音を考慮する信号と重みパラメータを最適化することによる認識率の向上を目指す．

◆音声認識結果から S/N 比が上がるにつれ認識率の低下を低減できていることから，本手法の有効性が確認できる．

Table 1: Recognition

S/N	Word	Proposed method	Comparision method 1	Comparision method 2
1/1	あるとき	86.7%	100.0%	100.0%
	北風	100.0%	93.3%	100.0%
	太陽	100.0%	100.0%	100.0%
	力比べ	100.0%	100.0%	100.0%
	Average	96.7%	98.3%	100.0%
1/2	あるとき	33.3%	20.0%	26.7%
	北風	100.0%	100.0%	93.3%
	太陽	100.0%	86.7%	100.0%
	力比べ	100.0%	100.0%	100.0%
	Average	83.3%	76.7%	80.0%
1/3	あるとき	33.3%	13.3%	0.0%
	北風	93.3%	93.3%	86.7%
	太陽	100.0%	46.7%	93.3%
	力比べ	100.0%	100.0%	100.0%
	Average	81.7%	63.3%	70.0%

1-Q-33

1-Q-33　TS-SUPERB：目的話者音声処理タスクによる自己教師あり学習モデルのベンチマーク

TS-SUPERB: A Target Speech Processing Benchmark for Speech Self-Supervised Learning Models

◎芦原孝典[1]，Peng Junyi[2]，Delcroix Marc[1]，落合翼[1]，Plchot Oldrich[2]，荒木章子[1]，Černocký Jan[2]（[1]NTT，[2]ブルノ工科大学）

◆音声 SSL モデルの成功に伴い，その音声表現そのものの有効性を検証・比較するためのベンチマークが様々提案されている．

◆しかし，これまでのベンチマークは主に単一話者が静音環境下で発話するシナリオに焦点を当てている．

◆本稿では，雑音環境下で複数話者が混在するより実践的な状況を想定し，その混合音声から目的話者の音声情報を抽出するタスク(TSE・PSE・PVAD・TS-ASR)で構成されるベンチマーク TS-SUPERB を提案する(Figure 1)．

◆実験から，TS タスクの精度傾向は，関連する単一話者タスクの精度傾向と明確な相関を示さないことが明らかとなり(Figure 2)，既存のベンチマーク指標のみでは，より実用的なタスクに適した SSL モデルを選定するための判断基準とはなり得えない事が示唆された．

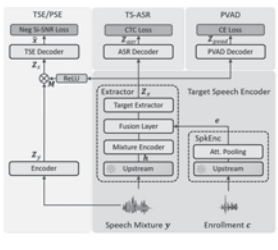


Figure 1: TS-SUPERB system

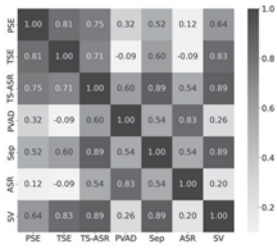


Figure 2: Spearman's rank correlation

1-Q-34

1-Q-34　Early Exit における退出層への注意機構導入による音声認識の高速化

Speeding up of autoregressive speech recognition by introducing attention mechanisms to exiting layers in early exit

◎河田尚孝，折橋翔太，鈴木聡志，田中智大，庵愛，牧島直輝，山根大河，増村亮（NTT）

◆ Early Exit
- ➢ 単一退出層を持つ学習済みモデルの各デコーダーブロックに対して，追加退出層を追加
- ➢ メリット：閾値を用いて正しく推論できる追加退出層で処理を中断することで，精度を維持したまま推論の高速化を達成

◆ 従来手法
- ➢ 浅い層における追加退出層の精度が悪い
  →浅い層で推論を中断することができないため高速化率が低い

◆ モチベーション
- ➢ 追加退出層のパラメータ増加に伴う精度向上およびEarly Exitによる高速化の適切なバランスの発見

◆ 提案手法
- ➢ 追加出退出層と学習済みモデルの間に注意機構を含んだ Transformer Decoder Block を追加
- ➢ メリット：エンコーダーの出力を考慮しながら追加退出層で推論できるようになるため，線形層を増やしてパラメータ数を大きくするよりも効率的に精度の改善が見込める

◆ 結果
- ➢ 浅い層の精度向上および高速化を達成

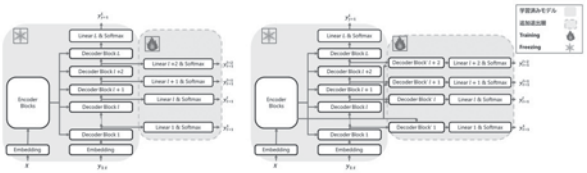


Fig.1：Left figure shows training with conventional exiting layers. Right figure shows training with proposed exiting layers.

1-Q-35

1-Q-35　音声離散トークンの韻律把握能力に関する分析的研究

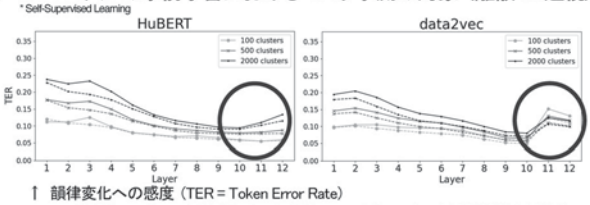
An Analytical Study on the Prosodic Encoding of Discrete Speech Tokens

◎恩田健太郎（東大/産総研），深山覚（産総研），齋藤大輔，峯松信明（東大）

① 絶対変化 ＜ 相対変化

This WAS easy for us　特定単語のみ強調　（相対変化）← より敏感に反応
THIS WAS EASY FOR US　発話全体を同様に強調（絶対変化）

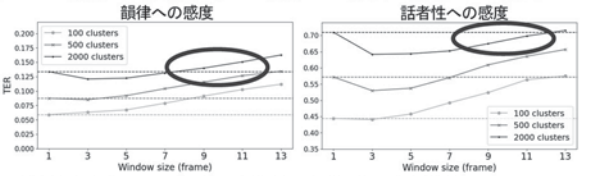
② SSL モデルの事前学習におけるマスク予測の対象（離散 vs.連続）



↑ 韻律変化への感度（TER = Token Error Rate）
HuBERT（離散）→ クラスタ数が大きいほど高い ＝ 小クラスタでも言語情報を捉える
data2vec(連続）→ クラスタ数が小さいほど高い ＝ 大クラスタ(≒連続)で言語情報を捉える
＊ ただし，最終層付近のみ．中間層ではともに似た傾向

③ K-means を感情音声で学習すると韻律への感度 UP

④ SSL 特徴量に移動平均 → 韻律への感度＆話者への頑健性 UP



クラスタ数 2000 → 9～11 フレーム(≒200ms)で移動平均をかけると，
韻律への感度↑ ＆ 話者性への感度↓（Window size=1 との比較）
＝ 音声言語モデルへの応用？

1-Q-36

1-Q-36　CTC と WFST を用いたハイブリッドシステムによる脳性麻痺者の音声認識

Speech Recognition for Individuals with Cerebral Palsy Using a Hybrid System with CTC and WFST

☆大谷赳[1]，相原龍[2]，高島遼一[3]，滝口哲也[1]，△田口進也[2]
（[1]神戸大学，[2]三菱電機，[3]立命館大学）

◆本研究では脳性麻痺を対象とした音声認識を目的とする。

◆構音障害者の音声認識モデルの学習を行う際，構音障害者の音声を大量に収集することが困難であるという課題がある。

◆本研究では，限られた音声データを効率的に用いることを目的に，重み付き有限状態トランスデューサ (Weighted Finite-State Transducers: WFST) をベースとしたハイブリッド音声認識モデルに着目する。また，CTC-WFST のアプローチを導入し，音響モデルの性能向上のために自己教師あり学習モデルを活用する。

◆データ不足の課題に対処するため，音声合成によるデータ拡張を行い，音響モデルの段階的なファインチューニングを行う (Fig. 1)。

◆論文では，特定話者音声認識モデルにより有効性を示す。

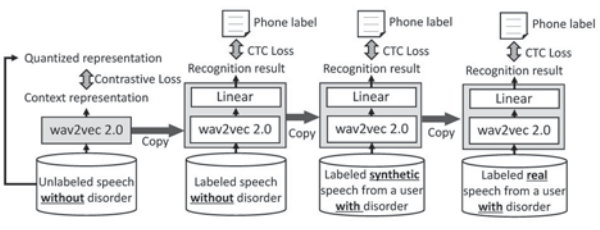


Fig. 1：Our acoustic model multi-step training procedures.

1-Q-37

1-Q-37　Parakeet-TDT を用いた構音障害音声認識の高精度化

Improving accuracy of dysarthric speech recognition using Prakeet-TDT.

☆北條圭悟，高橋快斗，坂井利光，若林佑幸，北岡教英（豊橋技科大）

◆本稿では，Interspeech2025 Speech Accessibility Project Challenge にて一位を獲得した，構音障害音声のための音声認識モデルを報告する。

◆構音障害者の音声は，調音器官の運動障害により音韻の生成が不明瞭となるため，一般的な音声に比べ認識性能が低下しやすい。

◆構音障害音声認識では，自己教師あり学習モデルを用いたアプローチが広く検討されているが，Whisper などの最先端モデルとの認識精度の比較があまり行われていない。

◆そこで，公開されている認識精度の高い事前学習済みモデルを調査し，その中で構音障害者音声に対して，最も精度が高いモデルをベースモデルとし，fine-tuning した（Table 1）。

◆実験では，ベースモデルとして Parakeet-TDT 1.1B を採用し，汎化性能向上のためのデータ拡張や，限られた推論時間で高精度化するためのデコード戦略の適用による認識精度の比較を行った。

Table 1：WER and SemScore for the existing models trained on standard speech.

model	*Dev*		*Test1*	
	WER	SemScore	WER	SemScore
wav2vec 2.0 large (fine-tuned on LibriSpeech 960h)	32.55	64.12	N/A	N/A
HuBERT large (fine-tuned on LibriSpeech 960h)	25.46	71.16	N/A	N/A
Whisper large v3	21.66	81.49	23.64	83.38
Canary 1B	15.35	82.43	N/A	N/A
Parakeet-RNNT 1.1B	12.88	82.64	11.02	85.73
Parakeet-TDT 1.1B	**12.79**	**83.15**	**10.40**	**86.75**

1-Q-38

1-Q-38　Whisper と LLM による生成的誤り修正に基づく構音障害音声認識システムの構築

Development of Dysarthric Speech Recognition System Based on Generative Error Correction Using Whisper and LLM

○高城異成, 若林佑幸, 北岡教英（豊橋技科大）

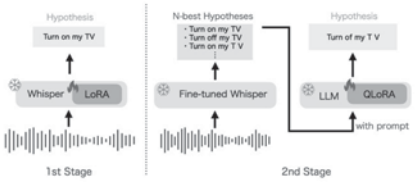


Fig. 1　Overview of the proposed system's training.

- Whisper と LLM を用いた生成的誤り修正 (Generative Error Correction: GER) を構音障害者音声認識に適用.
- 2ステージの学習において, データセット選択と LoRA や QLoRA を活用し, 大規模な計算機環境なしでも効率的に学習可能(Fig. 1)
- 結果として, 構音障害者音声認識の精度を競う INTERSPEECH 2025 Speech Accessibility Project Challenge にて３位に入賞.

Table 1　WERs and SemScores on the SAPC.

Method	*test1*		*test2*	
	WER	SemScore	WER	SemScore
WLV2 (baseline)	14.97	82.26	17.8	75.85
WLV3	10.38	86.26	-	-
+batched inference	9.65	86.99	-	-
+LoRA (1st stage)	8.84	88.19	-	-
+GER (2nd stage)	8.43	88.89	10.90	84.42

1-Q-39

1-Q-39　脳性麻痺による構音障害患者のための音声認識システムにおける発話特性の比較検討

A comparative study of speech characteristics in an automatic speech recognition system for dysarthria patients with dysarthria due to cerebral palsy

△戸叶遥斗(阪大・基), ◎細田侑也(立命館大・情理)

◆本発表では, 構音障害患者の発話に含まれる言い淀みや不安定な抑揚などの発話特性が音声認識システムに与える影響について調査.

◆性別 や 年齢, 構音障害の型 が異なる話者７名の音声を収録.
- 母音 ５種類（可能な限り継続して発声）
- オーラルディアドコキネシス ３種類（５秒間連続して発声）
- 長文（北風と太陽）と 短文「藪の中からウサギがぴょこんと出てきました」
- コマンド単語 10 種類 と 数字読み上げ 10 種類

◆同じ型の患者でも異なる発話特性をもち, 音声の周期性 や 構音器官の滑らかさ が音声認識システムの性能に影響することを確認.

Table 1: Acoustic measures and keyword recognition accuracy

ID	長母音			ODK [syll/sec]			認識率 [%]	
	Jitter	Shimmer	NHR [dB]	/ pa /	/ ta /	/ ka /	Cmd.	Num.
Spk1	1.56	7.71	18.40	3.07	2.87	2.40	16	20
Spk2	0.40	2.93	25.22	4.64	5.32	5.12	60	72
Spk3	1.98	8.66	16.14	2.53	2.90	2.50	16	26
Spk4	0.47	3.05	24.50	4.93	4.53	4.00	92	94
Spk5	0.37	2.23	29.44	6.27	8.23	6.83	94	80
Spk6	0.68	2.43	26.03	0.93	0.77	0.71	30	24
Spk7	0.77	6.06	19.34	4.50	3.73	2.70	84	86

1-Q-40

1-Q-40　Wild-card CTC によるキーワード検出と中間層バイアスを用いた再学習不要な稀少語音声認識

WCTC-Biasing: Retraining-free Contextual Biasing ASR with Wild-card CTC-based Keyword Spotting and Inter-layer Biasing

◎中込優, Michael Hentschel (LINE WORKS)

◆ End-to-end 音声認識では、認識精度は学習データ中の語彙に依存し、固有名詞など稀少語の認識精度が低下する課題がある。

◆ 本研究では、追加のモデル学習や音声合成を用いることなく、CTC モデルの稀少語認識精度を向上させる手法を提案する。

◆ 推論時に音響エンコーダの中間層において、キーワード検出を行い、検出されたキーワードをエンコーダ後続層に条件づけることで、対象キーワードの認識率を改善する。

◆ 稀少語のキーワード検出には、曖昧な一致にも柔軟に対応でき、高速な計算が可能な Wild-card CTC を採用した。

◆ 日本語の音声認識実験において、稀少語に対する F1 値が 29%向上することを確認した。

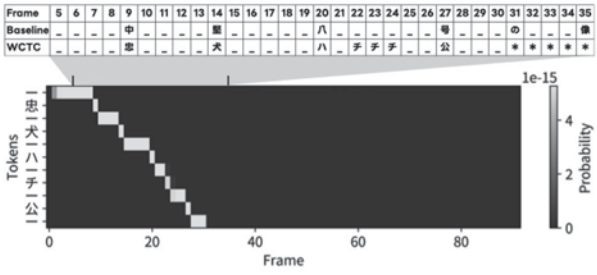


Fig.1 Wild-card CTC decoding path for the target keyword "忠犬ハチ公".

1-Q-41

1-Q-41　中間層における Contextual Biasing を用いた稀少語音声認識

Intermediate Contextual Biasing for Rare-word Speech Recognition

◎Michael Hentschel, 中込優 (LINE WORKS)

◆End-to-end 音声認識システムにおけるユーザー固有語彙の音声認識への適応は、依然として未解決の課題である

◆稀少語の認識を改善する手法として、neural contextual biasing が提案されている

◆本研究では、self-conditioned CTC と組み合わせた、neural contextual biasing の新たな応用を提案する

◆具体的には、全ての self-conditioned 層に neural contextual biasing を適用し、稀少語の事後確率の向上を図る

◆日本語音声データを用いた実験により、本手法は稀少語の認識率を25%～62%向上させることを確認した

◆Table 1 示す CSJ、Common Voice（CV）、JSUT、TEDxJP のデータセットを用いた実験により、提案手法の有効性を確認した

Model	Eval1	Eval2	Eval3	CV	JSUT	TED
SC-CTC ベースライン	5.1	4.2	4.0	22.2	13.6	17.5
提案法	4.9	3.8	3.9	20.7	12.7	16.7

Table 1: CSJで学習した音響モデルにおけるgreedy CTCデコードのCER[%]

1-Q-42

1-Q-42　音声認識における固有名詞対策のための Tool-use

Tool-use for Proper Noun Handling in speech recognition.

◎菅野竜雅，佐藤裕明，佐久間旭，△熊野正，
△河合吉彦（NHK），渡部晋治（CMU）

◆事前に想定できない固有名詞の対策手法を提案
◆E2E 音声認識モデルは、低頻度語である固有名詞で精度が著しく劣化
◆出現する固有名詞が想定できるとき → Biasing や Fine-tuning で対策可能
◆出現する固有名詞が想定できないとき → 本提案
- 音声に含まれる固有名詞の位置を特定 (Named entity recognition)
- 固有名詞のフリガナ(音素列)を出力 (Phoneme recognition)
- 外部の知識(検索 API)から正しい表記を取得 (Tool-use)

◆バイアスリストや辞書なしで固有名詞の認識精度の改善を確認

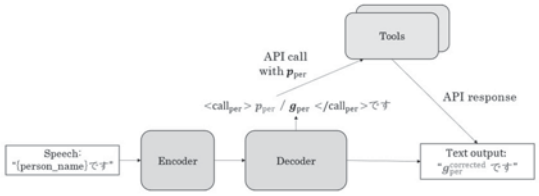


Fig.1：Proposed method

1-Q-43

1-Q-43　日本語諸方言と標準語の知識を共有した音声認識のための自己教師あり学習モデルの適応手法

Adaptive method for self-supervised learning models on automatic speech recognition based on shared knowledge of Japanese dialects and standard Japanese

☆淺川夏生瑠，△高橋尚希，甲斐充彦，中川聖一（静岡大）

◆本研究では，日本語の諸方言(COJADS)と標準語の両者に適応した音声認識モデル構築のため，SSL モデル XLS-R の適応手法について検討した。
◆先行研究の 3 段階ファインチューニング手法をベースに，方言と標準語を同時に学習に用いる手法，マルチタスク学習として行う方言識別にソフトラベルを用いる手法，標準語データセットとして ReazonSpeech を新たに用いる手法の 3 つについて検討した。
◆方言と標準語を同時に用いる手法では，標準語のデータ量を方言データの 1 方言地域あたりの時間に合わせることにより，方言音声では 28.5%，標準語では 5.5%の CER を単一の ASR モデルで達成した。

Method (Training Data)	CER[%] *CSJ*	CER[%] *COJADS*
adapter-based ft (full ft)		
(CSJ:0h, COJADS:63h)	11.2 (11.1)	29.2 (28.8)
(CSJ:3.7h, COJADS:63h)	5.5 (6.6)	**28.5** (28.8)
(CSJ:60h, COJADS:63h)	4.3 (4.7)	29.2 (28.7)
(CSJ:230h, COJADS:63h)	**3.7** (**3.7**)	29.4 (29.3)

Table.1：CER of a model trained using both standard language and dialect speech simultaneously in the final stage of a three-stage fine-tuning method

1-Q-44

Model-free Speculative Decoding for Transformer-based ASR with Token Map Drafting

○ Tuan Vu Ho, Hiroaki Kokubo, Masaaki Yamamoto, Yohei Kawaguchi
(Hitachi, Ltd.)

Transformer-based end-to-end ASR systems like Whisper achieve high accuracy but suffer from slow autoregressive decoding, limiting use on CPU-only devices. Speculative decoding (SD) speeds up inference using a draft model, but this is impractical without GPU support. We propose **Token Map Drafting**, a model-free SD method that uses a precomputed n-gram token map from domain-specific data to suggest candidate tokens. This approach accelerates decoding in structured, low-perplexity domains with minimal overhead. On the CI-AVSR and internal datasets, our method achieves up to 1.37× speed-up without loss in accuracy, outperforming CPU-based Distill-spec by 10% in decoding speed.

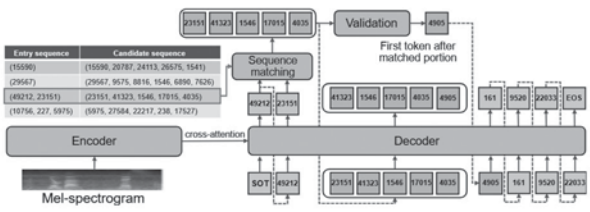


Figure 1: Overview of our proposed speculative decoding with Token Map Drafting.

1-Q-45

A Unified Speech Tokenizer from SSL Features: Disentangling Content and Style with FSQ-VAE

1-Q-45　SSL 特徴量に基づく統一音声トークナイザ: FSQ-VAE による内容とスタイルの分離

☆ Zhijie HUANG, 齋藤 大輔, 峯松 信明 (東大院・電気系工学)

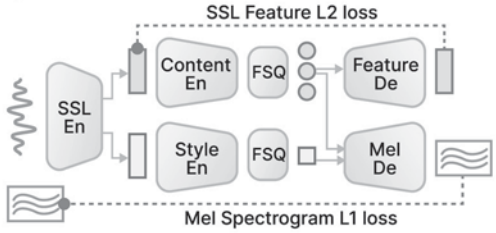
Motivation

complex, inefficient

SpeechLMs often rely on **2-stage tokenization.**

We propose a **unified speech tokenizer** for SLMs.

Design　**Unsupervised disentanglement. Multi-task training.**



Results

Single-stream 25Hz content tokens rich in linguistics and prosody.
Utterance-level style embedding for speaker identity.
High-quality reconstruction: **4.06 UTMOS, 2.66% WER, 300bps.**
Voice conversion experiment shows **effective disentanglement.**

1-Q-46

1-Q-46　講複数話者 TTS における深層話者埋め込み特徴を用いた話者再現性推定法の検討

Estimating speaker identity preservation in multi-speaker TTS using features with deep speaker embedding.

☆張 珠焕, 西村 竜一(和歌山大)

◆複数話者 TTS が出力した合成音声の話者再現性（参照音声の話者を再現できているか）を推定（Fig. 1）

◆参照音声から抽出した深層話者埋め込み特徴（x-vector）と話者再現性を対としたロジスティック回帰（教師あり学習）を採用

- 話者再現性（Table 1，4,507 個）：参照音声と合成音声を聴取し、聞き比べて主観で定義

◆入力特徴量の構成を 21 条件で比較（x-vector の次元数の削減）

◆正解率 0.65、ROC 曲線の AUC 面積 0.70（10 分割交差検証）

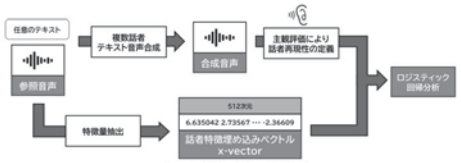


Fig. 1:提案手法の全体図

Table 1:主観による話者再現性の定義結果(コーパス別内訳)

コーパス名	再現OK	再現NG	合計
CSJ[1]	177	423	600
JVS[2]	143	857	1,000
JNAS[3]	483	1,317	1,800
方言[4]	338	749	1,087
Xeno-canto[5]	0	20	20
合計	1,141	3,366	4,507

[1] https://doi.org/10.15084/00001357
[2] https://arxiv.org/abs/1908.06248
[3] https://doi.org/10.32130/src.JNAS
[4] https://doi.org/10.32130/src.GSR-JD
[5] https://www.xeno-canto.org/

1-Q-47

1-Q-47　追加学習不要な TTS モデル操作に向けたモデルマージとタスク算術の検討

A study on model merging and task arithmetic toward training-free TTS model manipulation

◎榎本悠久，今井柊平，金子剛士，中村泰貴（Parakeet）

- テキスト音声合成で、複数の学習済みモデルの重みを足し引きするだけで、追加学習なしに新しいモデルを作る**モデルマージ技術**の可能性を検討
- **線形補間**：モデル A と B から、中間のモデルを作る
  - $M := (1 - \alpha) * A + \alpha * B$
- **タスクベクトル**：モデル A に足すことで特徴を付与する
  - あるモデル B と、「B＋**何らかの特徴**」の B_task を考える
  - 「B_task – B」はその特徴を表すベクトル（タスクベクトル）
    - $M := A + \beta * (B_task - B)$
  - 例えば M := A + **(B_囁き – B)** で A に囁き付与できるかも？
- **実験**：単独話者 VITS で、**コンポーネントごとに係数を変えて**マージ（Text Encoder, Flow, Decoder, Duration Predictor）
- **結果**
  - 線形補間では、話者性は Decoder の影響が圧倒的に高い
    - 話者性だけ A、他は B、のような TTS モデルが作れる
  - 囁きタスクベクトルでの囁き付与もある程度成功

1-Q-48

1-Q-48　系列整合性評価に特化した高難度日本語テキスト音声合成コーパスの検討

A Study on a High-Difficulty Japanese Text-to-Speech Corpus Specialized for Sequence Consistency Evaluation

◎今井 柊平, 榎本 悠久, 金子 剛士, 中村 泰貴 (Parakeet)

- 自己回帰型の TTS 手法は韻律の自然性に優れる一方で，単語の繰り返しやスキップなどの系列の非整合が構造上起こりやすい．
- 日本語においてこれらを評価するためのコーパスが存在しないため，以下の 4 つのサブセットから構成される評価コーパスを構築した．

Table 1: 作成したコーパスの各サブセットに含まれる文章例

Subset	Samples
short	「え？」「雨」「朝」「トマト」「食べる」
repetition	「えっと、えっと、えっと、えっと、えっと、えっと、えっと、なんだっけ？」「わ、わわわわ、私は、そんなつもりじゃ、ありません！」
rhyme	「会社の会議で、開発係の改革案を解決するため、会合が開催された。」「母校の素行の方向について、高校で出稿した思考の変更を促す投稿が成功。」
continuation	「ねぇ、だから言ったで」「本日の献立は、魚料理、煮物、あとはみそ汁の」

- 作成したコーパスにより，日本語に対応しているオープンソースの自己回帰型 Zero-Shot TTS 手法である CosyVoice2, OpenAudio S1-mini, xTTS-v2 を評価した．
- 実験結果から，系列整合性を適切に評価できる評価用コーパスを構築できたことを確認した．

Table 2: 作成したコーパスを用いた Zero-shot TTS 手法の評価

Model	short			repetition			rhyme			continuation			Cos.
	cer_b	cer_a	cer_w	cer_b	cer_a	cer_w	cer_b	cer_a	cer_w	cer_b	cer_a	cer_w	sim
CosyVoice2	24.41	69.45	117.3	7.841	14.99	27.79	**0.1774**	1.440	4.647	**0.4614**	5.386	15.69	**0.7072**
OpenAudioS1-mini	**1.575**	**14.49**	42.52	12.16	35.63	80.60	0.5321	1.348	**3.051**	**0.4614**	**1.292**	**2.249**	0.6439
xTTS-v2	6.299	14.65	**29.92**	**7.494**	**12.03**	**18.11**	**0.1774**	**1.029**	**3.051**	0.6344	1.465	3.114	0.6214
FastSpeech2 (JSUT)	-	8.661	-	-	3.672	-	-	0.1064	-	-	0.05767	-	-

1-Q-49

1-Q-49　動的特徴量を制御可能な TTS モデルによるポップアウトボイスの生成手法に関する検討

A study on the generation method of pop-out voices by a TTS model with controllable dynamic features.

☆上杉亮介, 坂野秀樹, 旭健作(名城大院)

◆背景雑音から際立って目立つ音声をポップアウトボイスと呼ぶ。

◆スペクトルの変動量を示す指標である動的特徴量の値が大きい音声は，よりポップアウトしやすいことが分かっている．

◆TTS モデルである JETS の variance adaptor に動的特徴量を予測する DF predictor を追加し，これを制御可能な TTS モデルの構築を行った．

◆生成された音声を聴取した印象は既存モデルと提案モデルとで大きく変わらなかった．

◆生成音声の動的特徴量の平均値を比較すると，既存モデルと提案モデルで大きな差は見られず，聴取した印象と一致している．

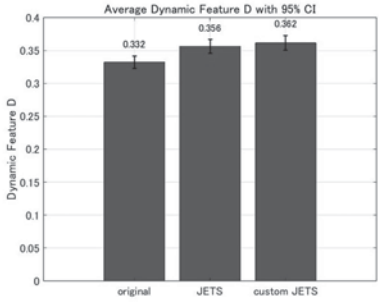


Fig.1 : Average dynamic features of speech signals of the original, the original JETS, and the proposed custom JETS.

1-Q-50

1-Q-50　WTNet：ウェーブテーブル合成に基づいた軽量なニューラルボコーダ

WTNet: Lightweight Neural Vocoder based on Wavetable Synthesis

☆廣畑和音（岡山大学院・HS 統合科学研），原直，阿部匡伸（岡山大）

◆f_0再現性の高い軽量なニューラルボコーダ・WTNet を提案

◆フレーム毎に一周期分の波形をもつウェーブテーブルを推定し，f_0に基づいてウェーブテーブルから値を取り出すことで波形を生成

◆信号処理技術と軽量な畳み込みアーキテクチャを活用することで，パラメータの大幅な削減と高速な推論を両立

◆HiFi-GAN や Vocos などの既存モデルと比較して軽量でありながら，客観評価指標においてこれらのモデルに匹敵する評価を獲得

◆主観評価実験において，f_0を変化させない条件で MOS 値 4.48 を記録し，従来モデルと遜色ない品質の音声を合成可能であることを確認

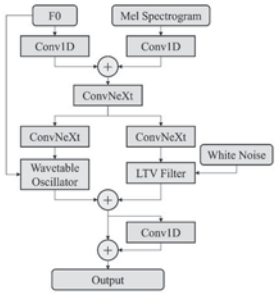


Fig. 1: Architecture of WTNet

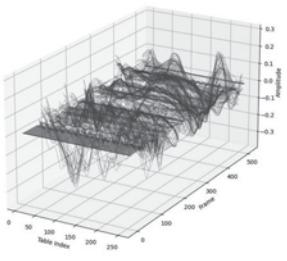


Fig. 2: Example of Generated Wavetable

1-R-1

1-R-1　ダクト内の定在波の節と腹の各誤差を観測する適応騒音制御の収束特性

Convergence characteristics of active noise control by observing errors at a node and an antinode of standing wave in a duct

☆柴崎高徳，武藤憲司（芝工大）

◆本研究では，ダクト内の定在波の節と腹で各誤差を観測し，これらの信号を加算して１つの誤差信号とする適応騒音制御を行った．

◆定在波を有するダクト騒音制御における収束性能向上のために，本研究で提案する定在波の節と腹に配置する誤差観測点 2 ヶ所による手法と，従来の誤差観測点１ヶ所による手法で，収束特性を比較した．

◆400 Hz の正弦波を対象に適応騒音制御を行った．定在波の腹のみで誤差を観測したときの収束時間は6.57 s であり，定在波の節のみで誤差を観測したときの収束時間は 10.33 s であった．また，定在波の節と腹で各誤差を観測したときの収束時間は5.72 s であり，定在波の節のみの場合と比較して 4.61 s(44.6 %)の短縮された．

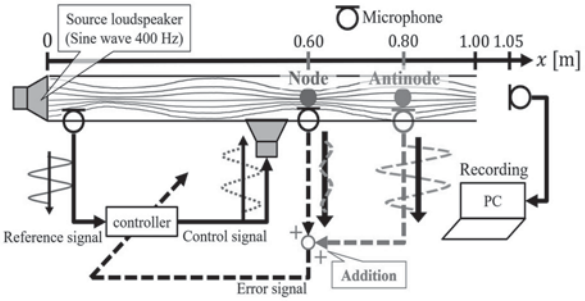


Fig.1：Proposal method of FF-ANC
by observing errors at a node and an antinode

1-R-2

1-R-2　制約付き最適化によるマルチチャネル ANC システムの消音領域拡大に関する初期検討

Basic Study on Expansion of Zone of Quiet in Multi-channel Active Noise Control System Using Constraint Optimization

◎岩居　健太（阪産大）

◆本稿では，複数の誤差マイクロホンを用いたマルチチャネルアクティブノイズコントロール (ANC) システム (Fig. 1) に関して，消音領域の拡大方法を検討する．

◆Filtered-x 法に基づく ANC システムは，誤差マイクロホン地点における騒音が最小となるように，騒音制御フィルタを更新する．そのため，10 dB の騒音低減効果が得られる領域は，誤差マイクロホンを中心に波長の 10 分の１程度の領域に限定される．

◆本稿では，２つの誤差マイクロホン地点での残差パワーの最小化を制約条件として制約付き最適化により，消音領域の拡大のための騒音制御フィルタ更新アルゴリズムを導く．

◆計算機実験の結果 (Fig. 2)，２つの誤差マイクロホン地点で一定の消音効果が得られることが示された．

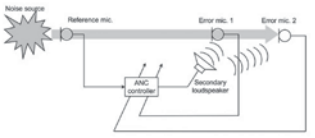
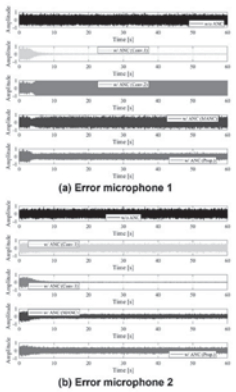


Fig. 1 Structure of proposed ANC system



(a) Error microphone 1

(b) Error microphone 2

Fig. 2 Time waveform of error signals

1-R-3

1-R-3　光レーザマイクロホンを用いたマルチチャネルフィードフォワードアクティブノイズコントロールシステムの複数実環境下における実機性能評価

Performance evaluation of multichannel feedforward active noise control system with optical laser microphone in various real environments

☆水谷真紘（立命館大院），岩居健太（阪産大），
中山雅人，西浦敬信（立命館大），添田喜治（産総研）

◆ フィードフォワードアクティブノイズコントロール(FFANC)システムにおける因果律の制約を緩和するため，光レーザマイクロホンを用いた FFANC が提案されている．

◆ しかし，本システムで取得される参照信号は空間の反射特性を含まないため，残響成分を取得できず，残響環境下では本 FFANC システムの騒音低減性能が劣化する．

◆ 本稿では，残響成分を考慮するために，参照マイクロホンとして気伝導マイクロホンを追加したマルチチャネル FFANC システム (Fig.1) を提案する．提案システムは，因果律の制約を緩和すると同時に，残響成分を含む参照信号の取得が期待できる．

◆ 実機実験の結果，複数実環境下で提案システムが最も高い騒音低減性能を持つことが示された (Fig.2)．

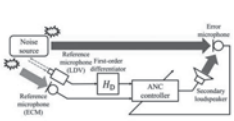
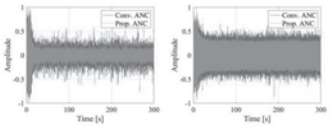


Fig. 1：Structure of proposed ANC system



(a) Soundproof environment　(b) Office environment

Fig. 2：Waveforms of error signals

1-R-4

1-R-4　デュアル ANC システムのマルチチャネル化に関する検討

Study on Multichannelization of Dual Active Noise Control Systems

☆仙藤拓馬, 豊岡祥太(関西大), 岩居健太(阪産大), 梶川嘉延(関西大)

◆ANC システムは騒音に対して同振幅，逆位相の擬似騒音を生成し，騒音と重ね合わせることで騒音低減を行うシステムである．

◆本稿では，デュアル ANC システムをマルチチャネル化した，マルチデュアル ANC システムを提案する．

◆デュアル ANC システムとは，パーティションで仕切られた隣接空間における，お互いの回折音をそれぞれ同時に制御することが可能である．

◆提案システムでは，デュアル ANC システムの各空間に複数個のマイクロホンと二次音源を設置することで，各空間内において複数の地点で騒音低減を可能とする．

◆Fig.1 に提案システムのシステム構成図を，Fig.2 に空間 2 の所望地点における騒音低減量を示す．提案システムでは，どちらの地点においても約 15 dB の騒音低減効果が得られた．

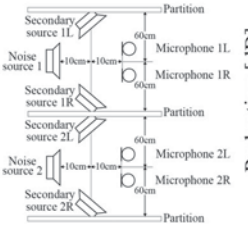


Fig.1：System configuration of the proposed system.

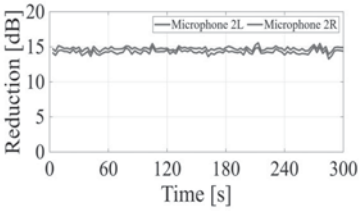


Fig.2：Noise reduction at the desired position in space 2.

1-R-5

1-R-5　ANC システムにおけるフィードバック経路および二次経路の最適なオンラインモデリングに関する検討

Study on Optimal Online Modeling of the Feedback and Secondary Paths in ANC Systems

☆辻脇楓馬, 豊岡祥太(関西大), 岩居健太(阪産大), 梶川嘉延(関西大)

◆ANC システムは騒音に対して同振幅，逆位相の擬似騒音を生成し，騒音と重ね合わせることで騒音低減を行うシステムである．

◆本稿では，広帯域な騒音環境において，二次経路およびフィードバック経路をオンラインで同時に推定可能な ANC システムを提案する．

◆提案システムでは，新たに導入した適応フィルタ H(z)により，線形予測フィルタを用いることなく，フィードバック経路の推定に混入する外乱信号を分離可能とする．

◆従来のオンライン二次経路モデリングと，改良したオンラインフィードバック経路モデリングを組み合わせることで，両経路の変動に追従可能である．

◆Fig.1 に提案システムのブロック図を，Fig.2 に所望地点における騒音低減量を示す．提案システムでは，２回の経路変動後も約 10 dB の騒音低減効果を維持した．

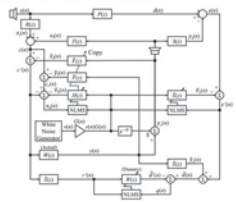


Fig.1：Block diagram of the proposed system.

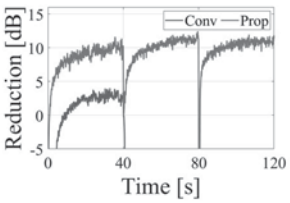


Fig.2：Noise reduction at the desired position.

1-R-6

1-R-6　能動騒音制御を目的とした物理法則に基づく深層学習を用いた音場推定の高速化

High-speed sound field estimation using physics-informed neural networks for active noise control

☆駒場颯斗, 佐藤元, 倉田験, 池田雄介(東京電機大)

◆背景

- 多点能動騒音制御(多点 ANC)は広範囲の騒音を低減できる一方，制御領域(ROI)内のマイクロホンにより行動が制限される。
- 物理法則に基づくニューラルネットワーク(PINNs)を用いて ROI 外の少数マイクロホンから音場補間を行い，ANC に利用する手法が提案されている。
- 従来の PINNs には学習に多くの時間を要する課題がある。

◆提案手法

- PINNs と高速に学習が可能な単層フィードフォワードネットワークモデル(ELM)を組み合わせた PIELM を用いて音場の補間を行う。

◆シミュレーション実験

- PIELM を用いて 200Hz 正弦波を補間する。
- 4.6 秒で学習を完了し，補間位置において NMSE で約 -10.5 dB と良好な結果を示した。

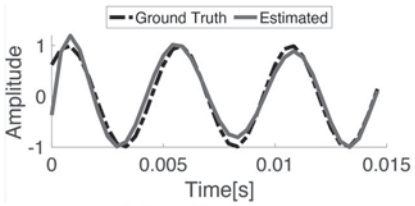


Fig.1：PIELMで推定した補間位置における波形

1-R-7

1-R-7　次の最先端を拓く汎用音響信号エンコーダの動向

Towards the Next Frontier: General-Purpose Audio Encoder Trends

○仁泉大輔, 竹内大起, 安田昌弘, グエンピンティエン, 大石康智, 原田登 (NTT)

◆ICME 2025 音響信号エンコーダチャレンジは既存手法や特定分野に忖度しない音声・環境音・音楽の各分野 27 の評価タスクにおいて汎用音響信号エンコーダの性能を評価するコンペティションである。提出された多様なアプローチによる解法を分析し、音を活用する生成 AI との融合における最新の動向を読み解いた。

◆結果は大規模音響言語モデル(LALM)である Qwen2-Audio を音響信号のエンコーダとして利用する解法がバランス良く良い性能を示した。

◆チャレンジには音声タスクが多く含まれ、これらにおける性能が良いことが必要であり、Whisper をベースに音楽や環境音を LALM の枠組みで学習することが効果的であることが示された。

◆無音の影響、十分な時間分解能の必要性、可聴音一般のデータセット AudioSet でのファインチューニングなど、多様なタスクにおいて効果的な特徴量を得るための知見が得られた。

◆チャレンジのタスク構成からは、現在の音を扱うマルチモーダル LLM に期待される能力も伺える。

1-R-8

1-R-8　ノルム制約･スパース正則化付き KL 基準 NMF の補助関数法に基づく最適化

Auxiliary-function-based optimization of KL NMF with norm constraint and sparsity regularization

☆和気佑弥，北村大地，綾野翔馬（香川高専）

◆NMF は低ランク近似手法として数多くの応用が提案されてきた．
◆補助関数法に基づく最適化を用いることで，目的関数の単調非増加性を保証しつつステップサイズが不要な乗算更新式が得られる．
◆L1 ノルムに基づくスパース NMF も数多く提案されており，スケール任意性問題や単調非増加性に対し様々な最適化手法が用いられる．
◆本研究では，ノルム制約を課すことで従来の手法が抱える課題を一挙に解決する乗算更新式を提案し，実験により収束挙動を評価する．
◆この結果，提案法は基底数や重み係数の変化に対し単調非増加性を維持しつつ，従来法と同等の性能と収束挙動の安定性を示した．

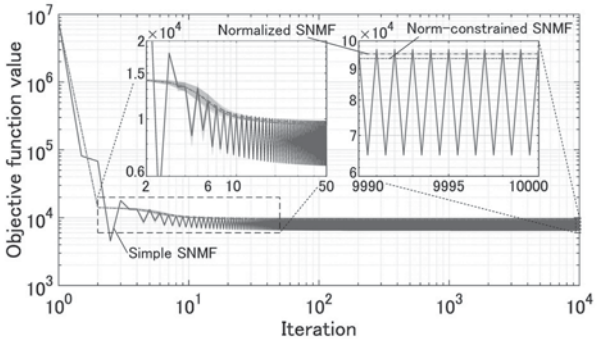


Fig. 1 : Convergence behavior. Proposed norm-constrained SNMF achieved stable convergence with comparable performance.

1-R-9

1-R-9

データ忠実性項を逐次更新する再重み付きスパース audio declipping

Reweighted sparse audio declipping with sequential update of data fidelity terms

☆ 角田清香，松本和樹，矢田部浩平（農工大）

スパース最適化に基づく audio declipping

► 時間周波数領域でのスパース性を誘導することで振幅超過により歪んだ振幅を復元

従来手法：制約集合を逐次改良する再重み付きスパース audio declipping

► 振幅を積極的に大きくするために制約の逐次更新を行う

△ 振幅値が過度に増長すると後続の反復で小さくできない

提案手法：データ忠実性項を逐次更新する再重み付きスパース audio declipping

► 制約の更新をペナルティの更新に緩和し，反復過程で過度に増長された振幅値を小さくできる手法を提案

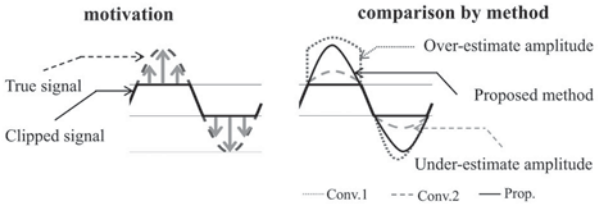


Fig. 1　Motivation for audio declipping (left) and an overview chart comparing conventional and proposed methods (right)

1-R-10

1-R-10

局所的な周波数帯域の部分更新による反復型位相復元

Iterative phase retrieval with partial updates in local frequency bands

◎ 新井慶大，矢田部浩平（農工大）

背景　位相復元：与えられた振幅スペクトログラムから位相の情報を復元する技術

従来　Griffin–Lim アルゴリズム：交互射影

- 所望の振幅を持つスペクトログラムの集合への射影
- 無矛盾なスペクトログラムの集合への射影

無矛盾化は**時間領域**を介して行われる

提案　**周波数領域**を介した無矛盾化を導入
＋ 局所的な**周波数帯域ごとに部分更新**

► 適切な帯域幅で PESQ・ESTOI が向上

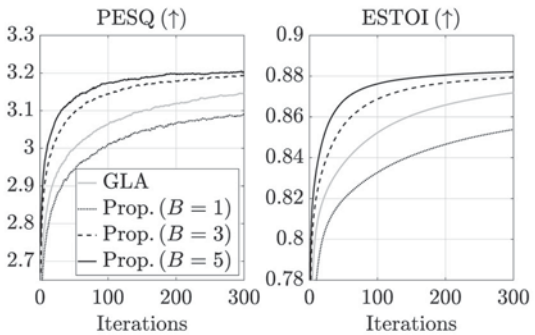


Fig. 1　The transition of averaged PESQ and ESTOI over 201 samples from TIMIT dataset. B denotes the width of frequency bands.

1-R-11

1-R-11

観測信号の周波数特性に基づく時変計量を用いた適応フィルタの実験的検討

Experimental study on adaptive filtering with time-varying metrics based on the spectral characteristics of observation

☆ 村田佳斗，松本和樹，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）

問題　空間伝達系の推定において観測信号の周波数特性の偏りが推定性能の悪化につながる

提案手法　反復ごとの更新方向を決定する計量を導入できる適応射影劣勾配法を用いて信号の特性を考慮

結果　真のインパルス応答を重みとした従来のNLMS法よりも高い推定性能を示した

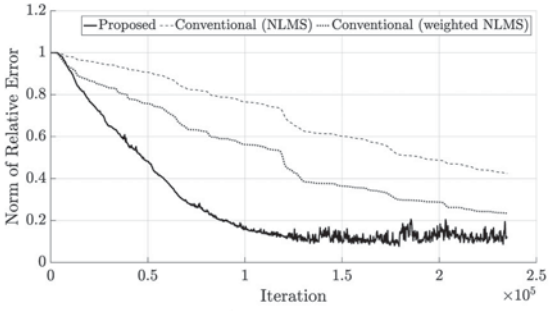


Fig. 1　Norm of relative error per iteration

1-R-12

1-R-12　音声・雑音の無相関性を考慮した Double Deep Priors 音声強調

Speech Enhancement Based on Double Deep Priors Considering the Uncorrelatedness Between Speech and Noise.

☆松村和樹, 中田優翔, 宮崎亮一（徳山高専）

◆Double Deep Priors に基づく，2 つの Deep Neural Network を用いた音声強調手法 (RK) は事前学習なしで高品質な音声強調を実現する一方，推定結果には試行間でばらつきが見られる.

◆本稿では，RK における推定精度の安定性を高めることを目的に，clean と noise の無相関性を考慮する損失項を RK の損失関数に追加する手法 (RKC) を提案する.

◆実験より，RKC は RK に比べて低周波数帯域のスペクトル削れを抑制することができ，推定精度の安定性が向上した．加えて，音声の品質についても改善が見られた (Fig. 1, Fig. 2) .

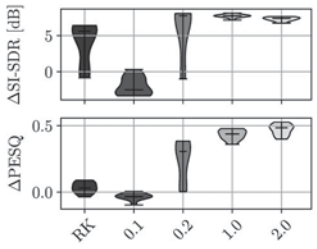


Fig. 1: Evaluation of SI-SDR and PESQ for RK and RKC with varying correlation weights.

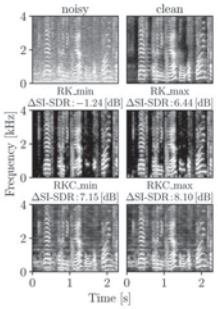


Fig. 2: Spectrograms of estimated clean.

1-R-13

1-R-13 位相を含む時空間スペクトル画像に対する画像修復に基づく雑音抑圧の検討

Noise Suppression Based on Image Restoration of Spatiotemporal Spectral Images Including Phase Information

☆両角有真（山梨大院・総教），鳥谷輝樹，小澤賢司（山梨大院・総研）

◆**背景**：雑音抑圧において，拡散モデルは強力な画像復元性能を有するが，時空間スペクトルの振幅情報のみに基づく従来手法では雑音の完全な除去が難しく，高周波帯域に残留雑音が生じやすい.

◆**目的**：振幅に加えて位相関連情報を統合した時空間スペクトル画像を用いることで，雑音抑圧性能の向上を図る.

◆**方法**：マイクロホンアレイから得られた音圧分布画像に対し 2DFFT を行い，振幅，相対位相，cos(相対位相)を 3 チャネル画像とし，画像修復型の拡散モデルに入力する.

◆**結果**：従来手法に比べ，雑音抑圧量（NSL）で最大約 4.8 dB，SDR で最大約 5.2 dB の改善を確認（Fig. 1, Fig. 2）．低周波成分の明瞭性向上や雑音抑圧の広域化が見られた.

◆**今後**：NSL で−20 dB の抑圧性能を目指すとともに，複数話者・複数雑音条件への拡張に取り組む.

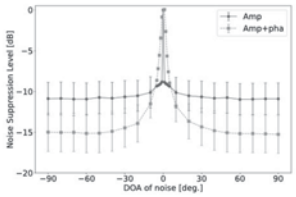


Fig. 1 Performance Evaluation Based on Noise Suppression Level (NSL)

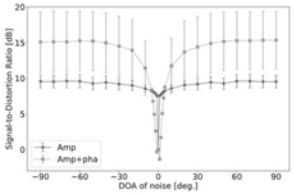


Fig. 2 Performance Evaluation Based on Signal-to-Distortion Ratio (SDR)

1-R-14

1-R-14　小型マイクロホンアレイを用いた雑音抑圧システムの実機リアルタイム化

Real-time ISSS noise suppression system using a small microphone array

鳥谷輝樹（山梨大・総研），△原口侑梧，△花形駿太，☆粟盛智也（山梨大・エ），小澤賢司（山梨大・総研）

◆ **背景**：時空間スペクトル推定に基づくスペクトル減算（ISSS）法による雑音抑圧（Ozawa et al., 2018）
→ 小型アレイを用いた実装が望まれる.

◆ **目的**：ISSS 法をシングルボードコンピュータに実装し，雑音抑圧性能を評価すること.

◆ **方法**：Raspberry Pi 4 model B に対して実装を行った（Fig.1）．マイクロホンの入力信号を 512 点ごとに切り出し，ハニング窓をかけ，半窓長ごとにオーバーラップさせながら ISSS 法へ入力しリアルタイムの出力を行った.

◆ **結果**：目的音に対して約 10~20 dB の雑音抑圧効果が見られた（Fig.2）．信号対歪み比（SDR）も約 10 dB の改善が見られた.

◆ **今後**：複数方向から到来する複数の雑音に対する抑圧性能を検証するとともに，スマートフォンへの実装を目指す.

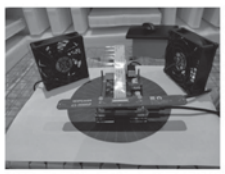


Fig.1：Overview of the apparatus for implementing the ISSS method.

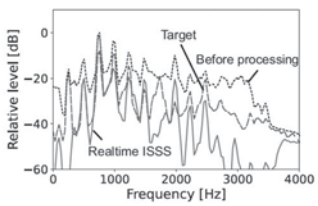


Fig.2：Amplitude spectra of the observed sound.

1-R-15

1-R-15　時間周波数マスキングを用いた微小異音検出手法の検討

A Detection Method for Subtle Anomalous Sounds Using Time-Frequency Masking

○阿部寛太郎，△宗戸俊樹，△坂本湧暉，笠原太郎（アイシン）

◆製造現場では熟練の官能検査員による異音の合否判定が行われており，省人化・自動化による生産性の向上が求められている．本稿では，製品作動時に発生する微小異音(擦れ音)を検出する方法を提案する.

◆従来では調波打楽器分離（Harmonic/percussive source separation: HPSS）を用い，得られた打音成分を元に異常品の検出を行っていた．しかし，HPSS 処理だけでは残存する製品作動音成分に微小異音成分が埋もれてしまい，異音の検出が困難であった.

◆そこで，HPSS による打音成分抽出に加え，残存する製品作動音成分を抑制する時間周波数マスキングを適用することで，異音部分のみを抽出する手法を検討した。提案手法による処理結果を Fig.1 に示す.

◆従来手法と提案手法で，正常データの統計量 3σ を閾値として異音検出性能を比較した結果，過検出率は同程度でありながら，提案手法では全ての異音を検出でき，検出精度の向上が確認された.

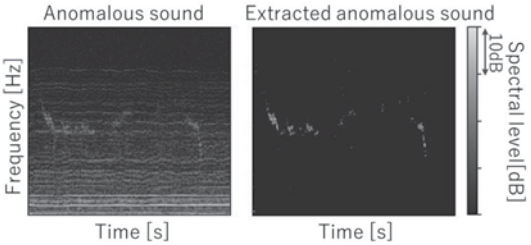


Fig.1：Extraction results using the proposed method

1-R-16

1-R-16

完全再構成性を保証した End-to-End 音声強調の実験的評価

Experimental evaluation of end-to-end speech enhancement with guaranteed perfect reconstruction

◎ 内田蓮，矢田部浩平（農工大），中村友彦（産総研）

背景：TasNet アーキテクチャ

- 時間領域信号を畳み込み層で特徴量に変換後，マスキング
- 逆変換に転置畳み込みを用いるため，完全再構成性を保証しない
- マスキングをしない場合でも入力信号を変化させてしまう
  - ▶ 前回，**直交畳み込み層**を用いて構造的に**完全再構成性を保証**

今回：提案手法による完全再構成性の影響を評価

複数の MaskNet とデータセットを用いて従来手法と**提案手法**を比較

Encoder, Decoder の冗長度や重み共有の影響を調査

提案手法に用いる直交畳み込み層の影響を調査

▶ 提案手法による完全再構成性が性能に及ぼす影響は**少ない**

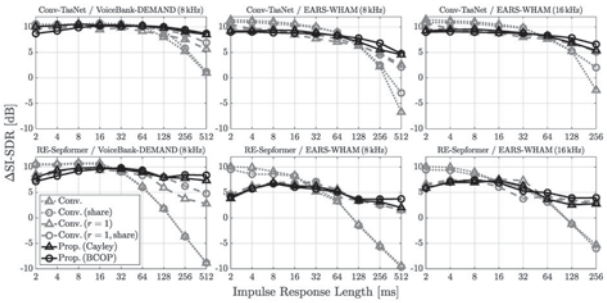


Fig. 1: ΔSI-SDR with various impulse response lengths. The rows and columns represent the results for different MaskNets and datasets, respectively.

1-R-17

1-R-17

正負ラベルなし学習を用いた半教師付きモノラル音声強調のハイパーパラメータ検討

An investigation on hyperparameters of semi-supervised monaural speech enhancement using positive-negative-unlabeled learning

◎小川諒，伊藤信貴，高宗典玄，山岡洸瑛，齋藤佑樹，猿渡洋(東大)

◆**正負ラベルなし(PNU)学習**を用いた半教師付き音声強調

- ➢ 少量の教師付きデータ(雑音あり・なし音声の組)と大量の教師なしデータ(雑音あり音声のみ)を活用

◆本研究：PNU 学習のハイパーパラメータ η，π_P を詳細に検討

◆実験結果

- ➢ 教師付き訓練データと教師なし訓練・検証データで**異なる雑音環境**を使用(前者は工場雑音、後者は地下鉄雑音)
- ➢ SI-SDRi は η=0.2，π_P=0.25 で最大値をとった
  - ✧ π_P=0.25 は教師なし訓練データに適合した値
- ➢ η>0 (Fig. 1 右側)が η<0 (Fig. 1 左側)より高い SI-SDRi を示した

※SI-SDRi: scale-invariant signal-to-distortion ratio improvement

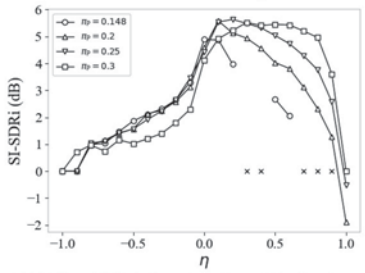


Fig.1: Validation SI-SDRi under mismatched noise environment. Symbol × indicates cases where validation SI-SDRi was negative infinity.

1-R-18

1-R-18

実環境雑音への適用を目的としたスペクトル平坦化による Deep Prior 音声強調

Deep Prior-Based Speech Enhancement Using Spectral Flattening for Robustness to Real-World Noise

☆中田優翔，宮崎亮一（徳山高専）

◆大量のデータ収集が困難な状況下では，観測信号のみで音声強調を行う Deep Prior に基づく音声強調手法 (DP-SE) が有用である．

◆我々は DP-SE の発展手法として，反復型の音声強調手法である Iterative DP-SE (IDP-SE) を提案した．

◆本稿では DP-SE，IDP-SE の実環境雑音への適用を目的として，スペクトル平坦化 (Spectral Flattening: SF) を導入する．

◆実験より，SF を導入した手法である DP-SE (SF)，IDP-SE (SF) は DP-SE，IDP-SE と比較して実環境雑音への性能が向上し (Fig. 1)，強調後の音声の品質が改善した (Fig. 2)．

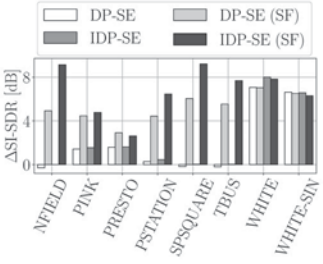


Fig. 1 Evaluation results for each speech enhancement method at 5 dB SNR.

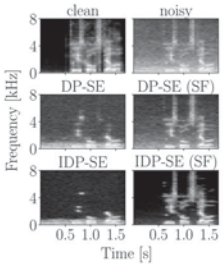


Fig. 2 Spectrograms for PSTATION noise.

1-R-19

1-R-19

パラ言語・非言語情報の記述文をクエリとした目的音声抽出

Target Speech Extraction Based on Para-linguistic and Non-linguistic Descriptions

◎関 健太郎，伊藤信貴，山内一輝，岡本悠希，山岡洸瑛，齋藤佑樹(東大)
高道慎之介(東大／慶大)，猿渡洋(東大)

❖ 目的音声抽出の課題：適切なクエリ形式の選択
- ➢ ボイスプリントを用いる方法は登録音声と目的音声の間の発話スタイルに差が生じることで性能低下を招く

❖ PNTP-TSE：目的音声抽出における新たなタスクを提案
- ➢ パラ言語・非言語情報の記述文をクエリとして利用
- ➢ 記述文をクエリとすることで柔軟な応用を実現

❖ PromptTSE：目的音声抽出のためのデータセットを提供
- ➢ 言語クエリを用いた目的音声抽出での利用を想定
- ➢ 言語・パラ言語情報・非言語情報を網羅

❖ 実験的評価によりPNTP-TSEの有効性を検証
- ➢ PNTP-TSE はボイスプリントを上回る性能を達成

Fig.1：PromptTSE dataset is available from here.

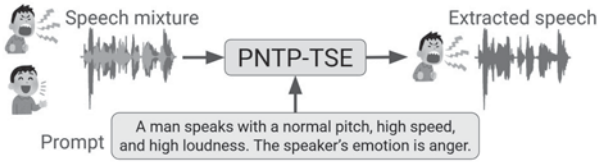


Fig.2：Overview of our proposed method.

1-R-20

1-R-20　音源推定における３群構成マイクロホンアレイの配置の検討

Investigation of the three microphone arrays configuration for sound source estimation

☆長塚哲平，武藤憲司（芝浦工大）

◆本研究では，マイクロホンの総数を変えずにマイクロホンアレイの構築を検討し，音源推定の精度向上を目指した．

◆従来使用していた３個，５個の２群構成と，新たに構築した２個，３個，３個の３群構成の２種類のマイクロホンアレイを用いて，音源推定の精度と座標トレース結果を比較した．

◆音源推定する空間を偏りなく調査できるように，アブラゼミ(Large brown cicada)とミンミンゼミ(Robust cicada)の鳴き声を再生する小型スピーカを 20 箇所に配置し，音源推定の誤差を求めた．

◆新たに構築した3群構成マイクロホンアレイは従来の2群構成と比較し，特定の位置で生じる大きな誤差が抑制することができた．

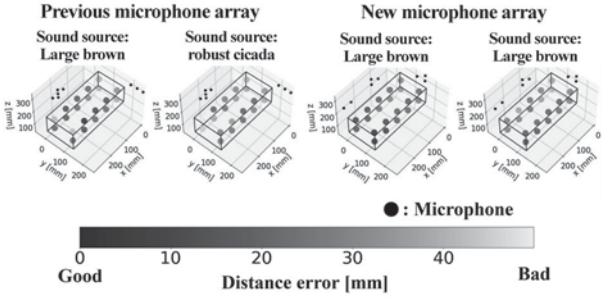


Fig.1. Accuracy of sound source estimation using each microphone array

1-R-21

1-R-21　7マイクアレイによる時空間的音響エネルギー可視化における最適解析フレーム長の検討

An analysis of optimal frame length for visualizing spatiotemporal acoustic energy by using seven microphone arrays

☆上林大輝(阪産大院･工学研)，中山雅人(立命大･情理)，高橋徹(阪産大院･工学研)

【背景】
対話活動可視化システムにおいて音源定位課題の解決[1]は必要不可欠であった．広範囲に適用可能な複数の非同期分散マイクロホンアレイで解決が試みられた．提案手法は空間の音圧エネルギー分布を推定するアイデアに基づいている．音源定位のために推定された音圧エネルギー分布は，そのまま対話活動を可視化したものとみなすことができる[2]．本研究でリアルタイム可視化の最適フレーム長を検討した．

【目的】
主観的にも客観的にも最適なフレーム長を求める．

【評価手段】
空間音響エネルギー分布の時間変化動画の印象．
音源位置のエネルギー変化とソース信号のエネルギー変化を最大化．

【フレーム長の主観評価】
200ms では，句読点のポーズ区間に対応して変化している印象．
400ms では，一つの文章に対応して変化している印象．

【フレーム長の客観評価】
Fig.1 は各フレーム長で分析したエネルギー時間変化の相関値の分布を表す．200,400ms の時，相関値は 0.95 を超えている．

【結論】
対話活動のモニタリングには，200，400ms が適当であることが示唆される．

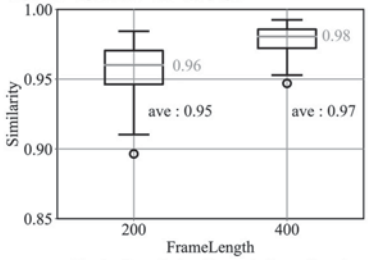


Fig 1 : Correlation for each frame length

[1]　上林+，信学技報,124(271), pp.1-6, 2024.
[2]　T. Kanbayashi+, GCCE, 2025, (in printing).

1-R-22

1-R-22　スマートフォンを用いた音源位置可視化システムにおける手ブレ計測

Measurement of hand tremor for visualizing a cicada sound source using a smartphone

☆山田和熙，武藤憲司（芝浦工大）

◆我々は音源位置可視化システムを作成している．音源位置可視化システムとは，スマホと 2ch マイクロホンを使用し，音源方向を画面上に表示するものである．

◆本研究は，音源位置可視化システムの結果表示に悪影響を及ぼす手ブレ量を数値化した．

◆正面方向の印をスマホ画面中央に写しながらスマホを回転する動作を行い，手ブレ量を測定した．手ブレ量は，2m 地点の正解位置と目印座標の距離で算出した．手ブレの平均値は 0.9 cm，最大値は 31.1 cm であった．

◆平均値は蝉１匹分であるため問題ないが，最大値は蝉５匹分となるため，手ブレ補正が必要である．

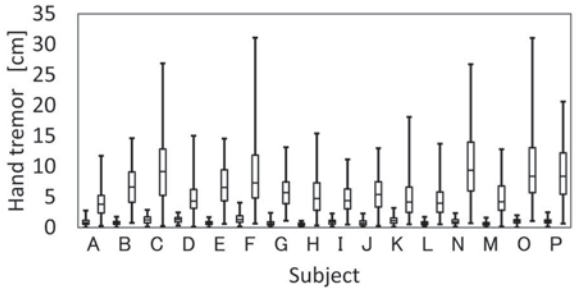


Fig.1 : Deviation from the target due to hand tremor (per subject)

1-R-23

1-R-23　Development of Sound Localization Evaluation System Using AR Technology: Sound Localization Test with Real Sound Sources

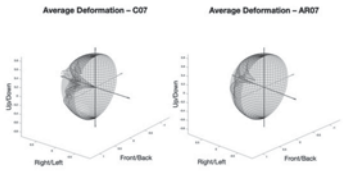
☆Nur Fakeehah Binti Rusli (Tokyo Denki University) /Yumi Fujii/Yasushi Honda/Yoshitaka Murayama (Cear Inc) / Yuko Watanabe (Tokyo Denki University)

◆In sound localization test, distance is harder to judge accurately compared to direction. To counter this, we developed a new test method using AR technology.

◆To see its applicability, we compared it with a conventional method and visualize the data.

◆The results shows that the proposed method produced less deformation of the perceptional space compared to the conventional method.

◆As the proposed method offers higher accuracy in localizing an object, it may be applicable in sound localization test.



Conventional Method　　Proposed Method

Fig 1. Deformation based on collected data for 0.7 m

1-R-24

1-R-24　深層学習を用いたマイクロホン一体型スピーカの同時計測による位置推定

Positional estimation of microphone-integrated loudspeakers by simultaneous measurement using deep learning

☆西原ちひろ，佐藤元，堀越光樹，津國和泉，池田雄介（東京電機大）

◆背景

- 立体音響技術において，実環境では理想的にスピーカを配置できない場合があり，音場制御への応用のため事前にスピーカ位置を把握することが重要である。
- マイクロホン一体型スピーカで音声の指向性を用いた位置の推定手法も提案されているが角度の推定精度が不足している。

◆提案手法

- マイクロホン一体型スピーカを円環上にランダムに配置し同時に再生・測定した信号を用いて深層学習で座標を推定。
- ネットワークアーキテクチャは畳み込み再帰型ニューラルネットワーク（CRNN）を適用。

◆シミュレーション実験

- ３つのネットワークアーキテクチャで座標の推定精度を比較。
- 実験の結果，CRNN を用いることで距離誤差は0.03~0.1 m 他の手法より高い精度を達成。

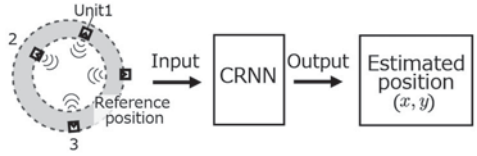


Fig.1：Overview of the proposed method.

1-R-25

1-R-25　深層学習を用いた歌声自動採譜における歌唱者の影響

The influence of singers on automatic singing transcription using deep learning.

☆田﨑晃基，小坂哲夫（山形大院・理工学研）

◆歌声は，歌唱者によって声質や歌唱表現が異なるため，すべての歌唱者に対応した自動歌声採譜モデルの作成は難しいと考えられる.

◆しかし，多くの自動歌声採譜の研究では，**学習に用いた歌唱者と用いていない歌唱者にどういった関係性があるのか調査されていない.**

◆本研究では，**自動採譜モデルの学習に用いた歌唱者(クローズ)と，用いられていない歌唱者(オープン)にデータを分けることで，歌唱者の違いが採譜精度にどういった影響を与えるのか調査する.**

◆自動採譜の入力には,バンド音源から音源分離を経て抽出した歌声音源を使用する.

◆結果として，**自動採譜モデル学習に評価と同じ歌唱者を用いることで，採譜精度が向上することを確認した.**

Table.1：Transcription results with the proposed method.
（Ⅰ and Ⅱ are group names for the three singers respectively.）

Data		State	F-measure [%] (↑)		
Test	Train		Frame	Note	Note w/offset
Ⅰ	Ⅰ	close	**81.60**	**78.29**	**55.43**
	Ⅱ	open	72.94	61.16	34.96
Ⅱ	Ⅱ	close	**82.47**	**71.95**	**58.45**
	Ⅰ	open	80.50	66.36	49.23

1-R-26

1-R-26　周波数方向のスペクトル伸縮と瞬時周波数補間によるピッチシフト

Pitch shifting by frequency-directional squeezing and instantaneous frequency interpolation

☆田中智己，赤石夏輝，矢田部浩平（農工大）

従来　　*信号の長さを変えず音高を変える処理

周波数方向に振幅スペクトログラムを伸縮するピッチシフト*

スペクトログラムは窓関数のメインローブ幅分，周波数方向に広がる

音高を上げるピッチシフトにおいては
広がりも同時に伸長され，音質を損なう恐れ

提案

スペクトログラム伸長後に振幅のピーク周辺を周波数方向に収縮
その後，元のスペクトログラムの周波数サンプリング位置で補間

結果

- 合成信号：スペクトルの広がりを抑える効果を確認
- 実信号：一部他手法を上回る音質を示した

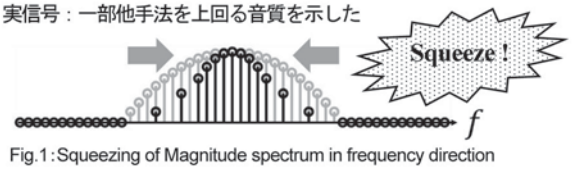


Fig.1：Squeezing of Magnitude spectrum in frequency direction

1-R-27

1-R-27　被り音混合系における独立低ランク行列分析のセミブラインド推定

Semi-blind estimation of independent low-rank matrix analysis for bleeding sound mixtures

☆谷野宮蒼士，北村大地（香川高専），高宗典玄，山岡洸瑛，猿渡洋（東京大），高橋祐，近藤多伸，△山川颯人（ヤマハ）

◆音楽のライブ演奏では，ボーカル（Vo.）用マイクに他の楽器やスピーカの音（被り音）が混入（Fig. 1）.

◆Vo.用マイクの被り音抑圧手法としてセミブラインド独立低ランク行列分析（ILRMA）を提案し，実際のライブハウスで録音し実験（Fig. 2）.

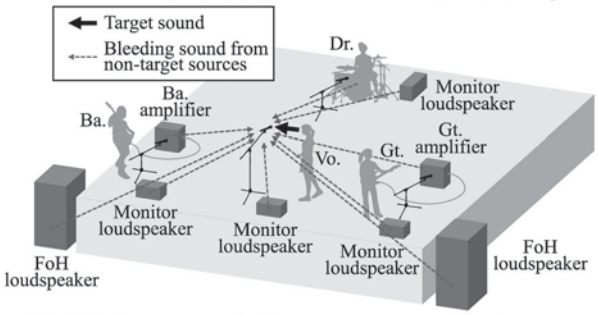


Fig. 1：Spatial arrangement of close microphones and sound sources.

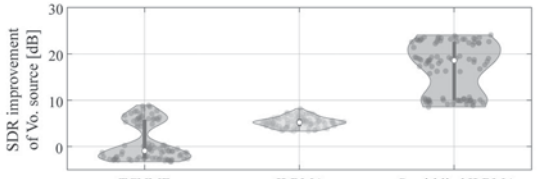


Fig. 2：Violin plot of SDR improvements of the Vo. source

1-R-28

1-R-28　目的関数の変化に基づくEarly Stoppingを用いたケプストラム基底分解型独立低ランク行列分析の基礎検討

A Fundamental Study of
Cepstral-Basis-Decomposed Independent Low-Rank Matrix Analysis with Early Stopping Based on Changes in the Cost Function

☆林航希，大島風雅，中山仁史(広島市大院)

◆これまでに，音声信号の低ケフレンシ成分に着目し，スペクトル構造の過学習を防ぐことを目的としたケプストラム基底分解（CBD）を導入したCBD-ILRMAを提案した．

◆CBD-ILRMAは安定かつ高性能な音声分離を実現したが，複雑性の低いモデリングは推定信号の詳細なスペクトル構造の表現が難しく，分離性能が十分に発揮されていない現状がある．

◆これを解決するため，動的リフタリングを用いた手法や設定したイタレーションに到達した際にCBDを打ち切る手法を提案した．

◆本研究では目的関数の変化に基づきCBDを早期停止させる新たな手法を提案し，より高精度な音源分離を実現することを目的とする．

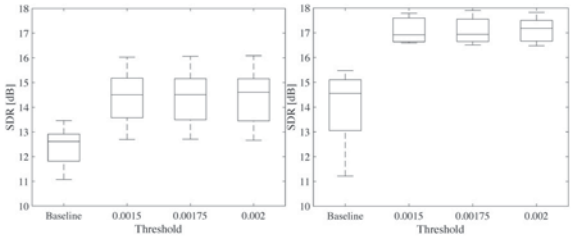


Fig. SDR comparison with and without Early Stopping

1-R-29

1-R-29　補助関数型ステアリングベクトル推定法を用いた空間正則化付き独立低ランク行列分析における正則化項の一般化

Generalized regularization for spatially regularized independent low-rank matrix analysis with auxiliary-function-based steering vector estimation

◎平田爽太，高宗典玄，山岡洸瑛(東大)，北村大地(香川高専)，
猿渡洋(東大)，高橋祐，近藤多伸(ヤマハ)

◆我々はこれまでに，到来時間差によってパラメタライズされたステアリングベクトル（SV）を独立低ランク行列分析（ILRMA）に対するL_2正則化項として活用した，フルブラインド空間正則化付きILRMA（FBSR-ILRMA）というブラインド音源分離手法を提案した．

◆本稿では，「理想的な分離行列とSVの誤差はスパースになる」という実験的観察に基づき，スパース正則化項への一般化を提案する．

◆L_p正則化項($0 < p \leq 2$)および多変量複素Student's t分布に着想を得た正則化項の２通りを提案し，補助関数法に基づく更新式を導出する．

◆数値実験により，スパース正則化による分離性能向上を確認する．

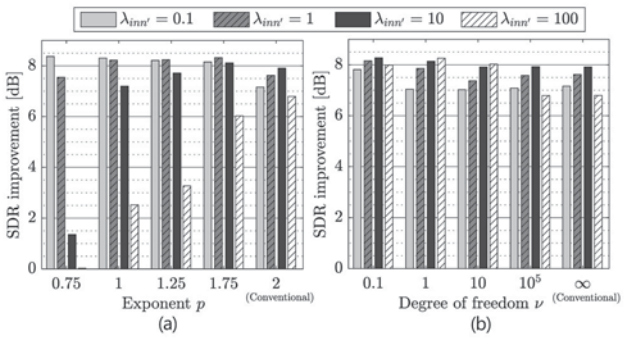


Fig.1: Average source-to-distortion ratio (SDR) improvement of FBSR-ILRMA with (a) L_p and (b) Student's t-inspired regularizations, compared with conventional one

1-R-30

1-R-30

クラスタリングに基づいて周波数帯域をグループ分けする木構造IVA

TreeIVA that groups frequency bands based on clustering

◎ 岸本麗央，松本和樹，矢田部浩平（農工大）

木構造 IVA（第 152 回 研究発表会 で発表）
周波数帯域をグループ分けし，階層的なスパース性を考慮する
優決定ブラインド音源分離手法

★ 調波音や歯擦音が主要な成分である帯域は音声によって異なる

前回　観測信号に適応したグラフの生成法を提案
- ペナルティ関数値が小さくなる周波数帯域のグループ分けをする
- グラフが２分木 に限られるという制限があった (Fig. 1：左)

今回　周波数成分をクラスタとしてグループ分けするグラフを生成
- 周波数帯域を任意の数にグループ分け可能 (Fig. 1：中央)
- 帯域ごとの特徴をより反映した周波数帯域のグループ分けを実現

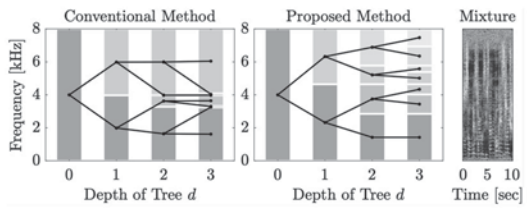


Fig. 1 Visualization of generated graphs (black lines) based on mixture and grouped frequency bands (gray rectangles).

1-R-31

1-R-31

優決定ブラインド音源分離における Projection Backのパラメータ表現

Parametric representation of demixing matrix after projection back for determined blind source separation

☆ 下川凜，松本和樹，矢田部浩平（農工大）

従来　音源数とチャネル数がともにKのとき，
各周波数ごとの分離行列の自由度は
実部と虚部併せて$2K^2$

結果　Projection Backの操作を前提とした
分離行列の別表現から2音源2チャネルのときの
分離行列のパラメータ表現を導出

提案　2音源2チャネルのとき，パラメータ表現によって
分離行列の自由度を8から5に削減

従来の分離行列

$$\begin{bmatrix} a+bi & c+di \\ e+fi & g+hi \end{bmatrix}$$

行列の自由度
実部と虚部併せて 8

提案の分離行列

$$\frac{1}{a+e+(b+f)i}\begin{bmatrix} a+bi & c \\ e+fi & -c \end{bmatrix}$$

行列の自由度　位相不定性を解消すると
実部と虚部併せて 5

1-R-32

1-R-32

優決定BSSのための 高速ADMMアルゴリズムの実験的評価

Experimental evaluation of fast ADMM algorithm for determined BSS

◎渡會博子，松本和樹，矢田部浩平 (農工大)

- ◆ 我々は以前，ADMMに基づく優決定BSSアルゴリズムを提案
  - ► 音源モデル設計の自由度の高さを維持しながら，PDSに基づく手法の高速化を実現
  - ► 空間白色化の性質を用いることで，さらに高速化
- ◆ 本稿では，ADMM-IVAとADMM-ILRMAをさらに効率化し，2，3，4話者の音源に対する実験的評価を行った
  - ► 従来のADMM-BSSの高速化・分離性能向上を確認
  - ► 特に4話者の場合において，補助変数法を上回る性能を達成

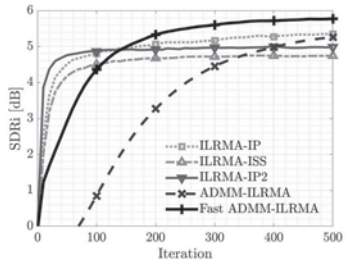


Fig. 1 : Average separation performance per iteration for 4 sources

1-R-33

1-R-33　マルチチャンネルモデルを用いた知識蒸留による単一チャンネル音声分離手法

Single-Channel Speech Separation Using Knowledge Distillation from Multi-channel Models

☆二通大地，ローランドハルタント，篠田浩一（東京科学大）

- ◆音声分離技術は，音声認識や音声処理のフロントエンドとして利用される重要な技術であり，近年，深層学習の発展により，大きく進歩している。
- ◆マルチチャンネル音声分離モデルは複数のマイクロフォンを用いることで，**音源到来方向（DOA; Directionof Arrival）や位相差などの空間情報を活用でき，高い分離性能**を実現する。
- ◆一方，単一チャンネルモデルは**空間情報を利用できず**，時間・周波数領域の特徴に依存するため，**性能は劣る**。
- ◆本研究では，単一チャンネルモデルの性能向上を目的として，**マルチタスク学習で訓練されたマルチチャンネルモデルから知識を蒸留する手法**を提案する。

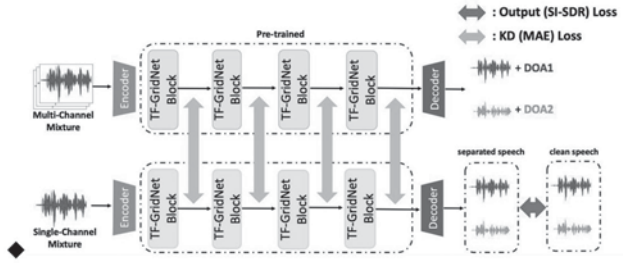


Fig. 1 : 提案手法の概略図

1-R-34

1-R-34　音量比特徴量の重み付きクラスタリングによるドラムセット録音時の被り音抑圧

Bleeding-sound reduction for drums recording using weighted clustering of volume-ratio features

☆鈴木慶，北村大地（香川高専）

- ◆ドラムのマイキングではドラムの各音源にマイクロホンを近接
  - ➢ 被り音の混入を防ぐためであるがマイキングだけでは難しい
- ◆ドラムのマイキングでは音量比が大きいという特徴を持つ（Fig.1）
- ◆特徴を用いて音量比の特徴量空間を生成し（Fig.2）クラスタリングを行い被り音抑圧を実現する

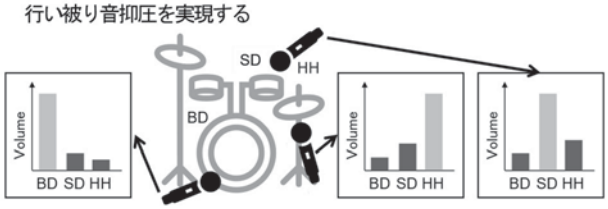


Fig. 1 Relative volume levels in multi-channel signal.

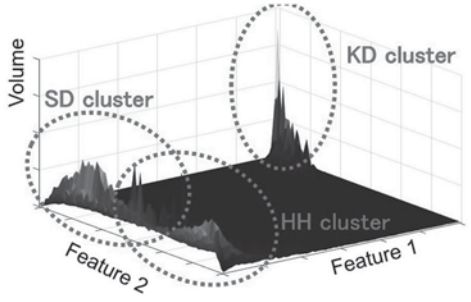


Fig. 2 Feature space of the observed multi-channel signal.

1-R-35

1-R-35　ディリクレ分布に基づく非負値テンソル因子分解のステレオ音楽分離

Application of Dirichlet-distribution-based nonnegative tensor factorization to stereo music source separation

◎小川遼，北村大地（香川高専）

- ◆目的
  - ➢ パニングで作成されたステレオ音楽信号の方位分離について検討
- ◆提案手法:ディリクレ分布正則化付きNTF
  - ➢ NTFの変数Zの各列に**ディリクレ分布**の事前生成モデルを導入
    - ✧ 変数Zの各列の**総和が1となる制約（ノルム制約）**を担保
    - ✧ 変数Zの各列を**one-hotベクトル**及び**denseベクトル**に誘導
  - ➢ 変数Zの各列は分解成分の左右チャネルへの分配重みに対応
    - ✧ one-hot及びdenseベクトルへの誘導が方位分離を促進
- ◆ステレオ音楽信号の方位分離実験
  - ➢ 信号対歪み比の平均は，方位クラスタリングが6.76 dB，提案手法が11.88 dBとなった（最適条件下）.

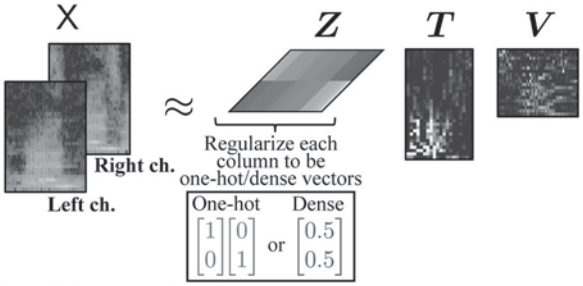


Fig.1 : NTF decomposition of a stereo spectrogram into channel-partitioning weights Z, spectral bases T, and temporal activations V.

1-R-36

1-R-36　小規模マイクロホンアレイから得た時空間スペクトルに基づくスパースベイズ音源分離

Sparse Bayesian Sound Source Separation Based on Spatiotemporal Spectrum Obtained from a Small Microphone Array

△有泉 千太(山梨大院･総教)，鳥谷 輝樹，○小澤 賢司(山梨大院･総研)

◆マイクロホンアレイから得られる出力を画像として捉え，その時空間スペクトル(2 次元スペクトル)を利用した音源分離手法を提案する。

◆時空間スペクトルの各時間周波数ビンにおいて，複数音源に関して重畳する空間スペクトルに対して，スパースベイズ推定によって音源方向ごとのスペクトルを算出する。それを全ての時間周波数ビンについて行い，音源方向ごとに時間周波数軸に沿って並べることで，音源ごとのスペクトルを得る。逆離散フーリエ変換により分離波形を得る。

◆マイクロホン 8 個を 2 cm 間隔で並べた直線状アレイについて，5 音源に関して計算機シミュレーション実験により SDR を評価した(Fig. 1)。背景雑音の SNR が 40 dB 程度あれば，十分な分離が可能である。10 音源の場合でも，ある程度の分離が可能である。

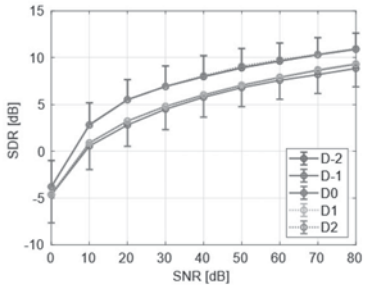


Fig. 1: SDR of five separated sound sources as a function of SNR, with an angular spacing of 20 degrees. (D-2 to D2 indicate the directions of the sources, where D0 is orthogonal to the array. Error bars represent the standard deviations.)

1-R-37

1-R-37

凸最適化に基づく劣決定音源分離における支配的な音源方向の選択法

Dominant source direction selection method for underdetermined source separation based on convex optimization

◎ 皆川朋樹, 矢田部浩平 (農工大)

劣決定音源分離

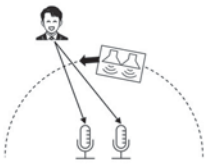
- 事前に空間モデルを定義することで凸最適化問題として扱える
- 方向が既知な仮想的な音源（仮想音源）からの混合を仮定

従来法

- **多数**かつ**固定**された仮想音源への分離
- エネルギーの高い仮想音源を選択
- ▶ 方向の多さが原因で**計算コストが重い**

提案法

- **少数**かつ**移動可能**な仮想音源への分離
- 移動が収まったときの仮想音源を選択
- ▶ **計算コストの軽減と，**
  従来法と**同程度の分離性能**



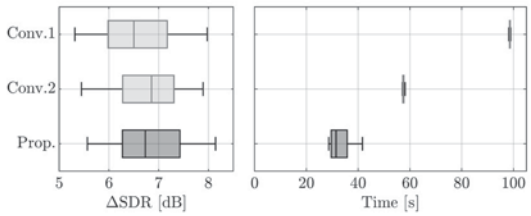


Fig. 1 Comparison of separation performance and processing time

1-R-38

1-R-38　機械学習を用いた音楽音源分離のためのデータ拡張手法の検討

Exploring Data Augmentation Techniques for Machine Learning-based Music Source Separation

☆平原裕雅（龍谷大院･先端理工学研），片岡章俊（龍谷大･先端理工）

◆ 機械学習を用いた音楽音源分離では，著作権等の関係から利用可能な音楽データが限られ，モデルの汎化性能が課題である．

◆ 本研究では，この課題に対し分離性能の向上を目的とし，音楽的整合性を考慮したデータ拡張手法を検討した．

◆ 分離モデルには Wave-U-Net と HT Demucs を使用し，既存楽曲の楽器音源をランダムに抽出し組み合わせる基本手法 (random mix) に加え，音楽的整合性を考慮した以下の手法を検討した．

- Method (A): 長さを考慮
- Method (B): ジャンルを考慮
- Method (C): ジャンルと長さを考慮

◆ Table 1 が示すように，データ拡張は分離性能を向上させ，特に Method (C)が両モデル共に最も性能が向上した．

◆ この結果から，音楽データを拡張する際は，音楽的整合性を考慮することで分離性能の向上に有効であると示された．

Table 1 Comparison of source separation performance for each data augmentation method [dB]

Model	Method	SDR	ISR	SIR	SAR
Wave-U-Net	default	1.73	6.15	2.95	**3.82**
	random mix	2.07	6.58	4.14	3.60
	Method (A)	2.06	6.66	4.27	3.54
	Method (B)	2.11	6.46	4.02	3.65
	Method (C)	**2.17**	**6.90**	**4.38**	3.74
HT Demucs	default	5.06	9.39	7.30	4.40
	random mix	5.47	9.97	8.14	4.79
	Method (A)	4.86	9.78	8.22	4.41
	Method (B)	5.47	10.13	8.34	4.83
	Method (C)	**5.57**	**10.22**	**8.78**	**4.83**

1-R-39

1-R-39

楽譜情報を用いた混合 Differentiable Digital Signal Processing モデルによる合成パラメータ推定手法の検討

Investigation of Synthesis Parameter Estimation Method Using Musical Score Information for Differentiable Digital Signal Processing Mixture Model

☆竹本健悟(東大/産総研)，中村 友彦(産総研)，猿渡 洋（東大）

◆混合 DDSP モデル [Kawamura+, 2022] は，DDSP 自己符号化器 [Engel+, 2020] を拡張し，混合楽器音から各楽器の基本周波数，音色特徴量，ラウドネス（合成パラメータ）を直接抽出

◆一方で，混合 DDSP モデルは合成パラメータの初期値に鋭敏であり，初期化において各音源と時刻同期した楽譜情報が必要

◆本稿では，楽譜が混合音と時刻同期されていない場合でも利用できる，動的時間伸縮法（DTW）を用いた合成パラメータ初期化法を提案

◆実験により，音源分離手法を用いる場合や線形に時間伸縮させる場合に比べ，提案法は高精度に推定可能であることを確認

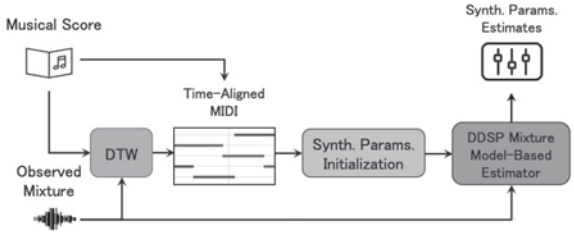


Fig.1 : Schematic of synthesis parameter extraction based on DDSP mixture model using proposed initialization method.

1-R-40

1-R-40　SoundBeam with KD：非因果モデルからの知識蒸留に基づくリアルタイム目的音抽出

SoundBeam with KD: Real-Time Target Sound Extraction Using Knowledge Distillation from a Non-Causal Model

◎若山圭吾，川瀬智子，森谷崇史，デルクロアマーク，佐藤宏，落合翼，安田昌弘，荒木章子（NTT）

◆混合音信号から指定した音クラスの信号をリアルタイムで分離する目的音抽出（TSE; target sound extraction）システムを提案する。

◆本システムは、非因果的 TSE モデルからの知識蒸留 (KD; knowledge distillation）によって学習された因果的 SoundBeam に基づいている。

- Conv-TasNet TSE では、因果的な実装で生じる性能低下を抑えるため、より高性能な非因果的教師モデルからの KD を導入。
- 特に、因果的拡張 Conv 層と非因果的 LN（layer normalization）で構成される部分的非因果的教師モデルによる KD が有効。
- 同様のアプローチでリアルタイム TSE モデルを構築（Fig.1 参照）。

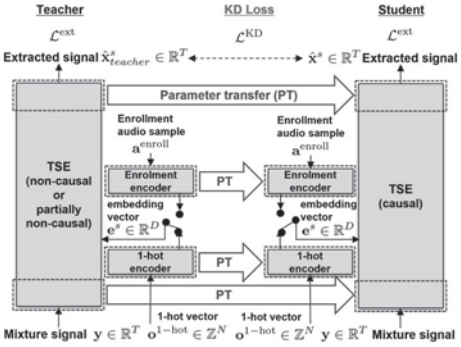


Fig.1：Semantic diagram of “SoundBeam with KD”.

1-R-41

1-R-41　映画音源分離のための非言語音声を含むデータセット

Dataset containing non-verbal sounds for cinematic audio source separation

◎蓮実拓也，藤田雄介（LINE ヤフー）

◆映画音源分離（cinematic audio source separation: CASS）は，映画のオーディオを音声，音楽，効果音に分離する技術である．

◆従来の CASS データセットは，音声のステムとして読み上げ音声のみを含んでいるため，学習した CASS モデルが笑い声や叫び声といった非言語音声を音声として分離できない．

◆本発表では、CASS モデルが非言語音声も音声として分離できるように，音声ステムに読み上げ音声と非言語音声の両方を含むデータセットである DnR-nonverbal を提案する．

◆提案するデータセットを用いて CASS モデルを学習させることで，合成映画音源と実映画音源のいずれにおいても，非言語音声を含む音声の分離性能が向上することを示す．

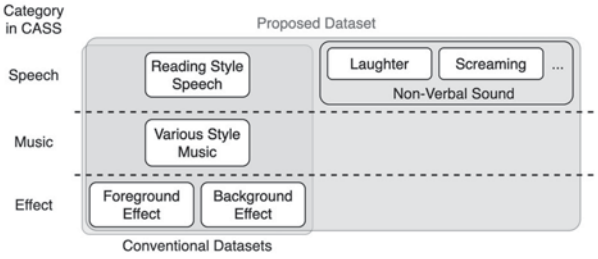


Fig.1：Comparision of conventional CASS datasets and proposed DnR-nonverbal. Unlike conventional datasets such as DnR-v2, our dataset covers non-verbal sounds often observed in movie audio.

1-R-42

1-R-42　IRMOS による同時 IR 推定を用いた音響測距精度向上の検討

Investigation of Distance Estimation Accuracy Improvement Using Simultaneous IR Estimation by IRMOS

☆徳富 響(NBU)，副島 翔太(NBU)，沖田 和久(NBU 卒(現 NJMC))，近藤 善隆(J-Tec)，福島 学(NBU)，△松本 光雄()，柳川 博文(arsl)

◆音響測距法

パワースペクトルに距離に応じた包絡線変調に着目

フーリエ変換し変調周期を求める（距離スペクトル）

◇0 収束しない成分（ノイズ）が包絡線変調を歪ませる

距離スペクトルに存在しない距離の推定値を生じる

◆インパルス応答（IR）推定手法

DLR：クロススペクトル法（CS 法）で生じる歪を解消

IRMOS：1 系統の同時 2 計測値取得を可能とした手法

確定値（IR）が 2 計測値に等しく存在することに着目

差分から DLR より正確な IR 推定が可能

◆これまで：推定ノイズレベルで打ち切り処理 (cut only)

提案手法：推定ノイズレベルを減算処理した上で打ち切り処理

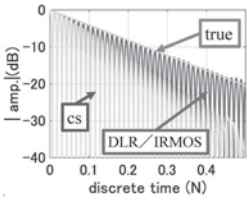


Fig.1 The difference between DLR/IRMOS and CS(Left)

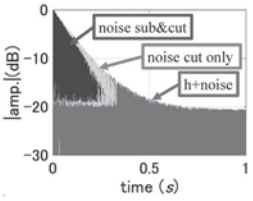


Fig.2 The result of Estimated IR (h+noise), cut only and sub&cut

1-R-43

1-R-43　カラー構造化照明と楕円当てはめを用いた振動計測の検証

Investigation of vibration measurement using color structured illumination and ellipse fitting

☆ 安藤元暉，矢田部浩平（農工大）

手法　カラー構造化照明 (RGB縞) を用いた形状推定
非接触での振動計測が可能
► 縞の歪みによって推定精度が低下

調査　撮影前後の補正（フィルタ，色補正，γ補正）と位相推定手法 (3-step algorithm，楕円当てはめ) の組み合わせによる推定精度を比較

結果　単方向の縞を用いて空間的に解析することが可視化に優れた手法であることを確認した

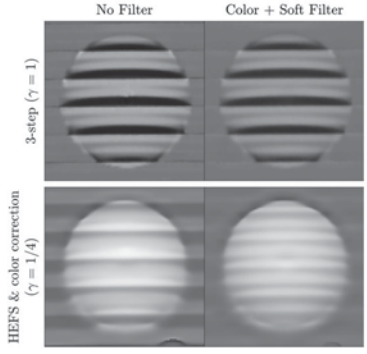


Fig. 1 A hemispherical shape estimated by 3-step algorithm (top) and ellipse fitting (bottom) without filters (left) and with lens filters (right).