

1-1-1

1-1-1 音像強調におけるハイパーソニック・プリバーチャルリード信号による音像定位キューの評価

Evaluation of sound-image localization cues using pre-virtual-leading hypersonic signals in sound-image augmentation ☆今中峻太(立命館大院), 歌毓庭, 中山雅人, 西浦敬信(立命館大)

- ◆従来、両耳効果に基づく音像定位手法が提案されている。先行音効果は両耳効果の1つであり、音の知覚される方向を強調することが可能である。また、人間の超音波知覚について、可聴音の知覚に影響を与える超音波の効果であるハイパーソニック・エフェクトや、超音波に低周波の包絡を与えることで触覚によって知覚される振動覚が報告されている。これまで、超音波帯域信号であるハイパーソニック・プリバーチャルリード信号(Fig. 1)を用いることで、音像定位強調が可能であること、当該超音波信号が主として耳周辺の感覚器により知覚されることを確認している。◆本稿では、受聴者正面以外に構築した音像に対してハイパーソニック・プリバーチャルリード信号を付加して放射することで、可聴音・超音波音それぞれの音像定位キューの評価を行った。実験結果から、一部の条件において、提案手法によって与えた超音波の音像定位キューが有効であることを確認した(Fig. 2).

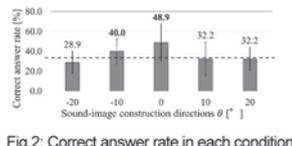
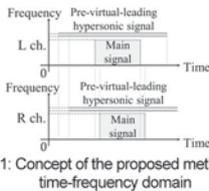


Fig. 1: Concept of the proposed method in time-frequency domain

Fig. 2: Correct answer rate in each condition

1-1-3

1-1-3 空間エイリアシングのスペクトル非一貫性に着目した音源定位法

A Sound Source Localization Method Focusing on Spectral Inconsistency in Spatial Aliasing ☆篠山俊介, 志賀芳則, △陶山健仁(東京電機大院)

- ◆一般的な音源方向推定技術は音源数が既知であり、マイクロホン数未満の音源数に限られる。◆本研究では、2マイクロホンで複数の音源方向を推定する。更に、音源数が未知である状況を想定する。◆提案手法では、MUSIC法を使用する。空間エイリアシングが発生する帯域も使用するが、空間エイリアシングによる偽のピークは周波数に依存し変化するため、ピークの分布に着目し推定を行う。◆ピークの数には雑音や反響の影響が小さいため、周囲の雑音環境が変化しても同一のパラメータで推定が可能であると考えた。異なる雑音環境の検討として、無響室での推定を行った。Fig.1に提案法の推定例、Fig.2に無響室での音源数推定結果を示す。

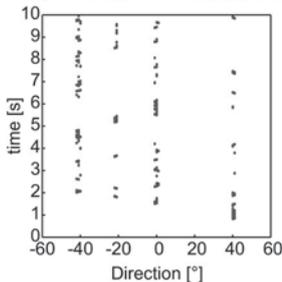


Fig. 1: Estimated Example

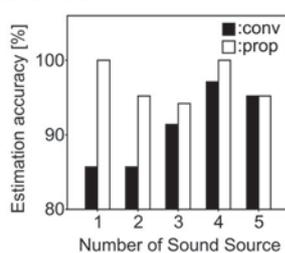


Fig. 2: Number of Sound Source Estimation Results in an Anechoic Chamber

1-1-2

1-1-2 実映画の動画像を用いた音響メタデータ自動生成における音像定位の評価

Evaluation of sound image localization in automated acoustic metadata generation based on actual film scene ☆加藤昂(立命館大院), 中山雅人, 西浦敬信(立命館大), 添田喜治(産総研)

- ◆オブジェクトベース音響は、マルチチャンネル音響の再生方式であり、再生空間内の座標情報を用いて、指定した位置に音像を構築することができる。◆しかし、音像位置は音響コンテンツ制作者が手作業で指定する必要があり、特に、長編映画や音像の数が多作品の場合、制作時の負担が増加するという課題がある。◆提案手法(Fig. 1)では、動画像から推定したオブジェクトの位置に基づいて音響メタデータを自動生成し、音像を構築することで制作者の負担を軽減する。◆主観評価実験では、提案手法を用いて構築した音像の、実映画(Fig. 2)における定位精度を確認した。その結果(Fig. 3)、音像の方向感と動きに関して、従来のステレオ映画と同程度の定位精度であった。

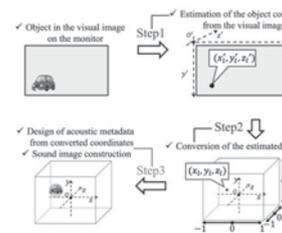


Fig. 1: Overview of the proposed method



Fig. 2: Video images used in the experiment

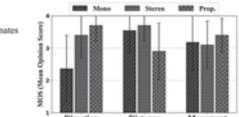


Fig. 3: Experimental results of consistency between sound image and video image

1-1-4

1-1-4

1-1-4 左右音量比特徴量を援用したConv-TasNetによるステレオ音楽分離

Stereo music source separation using Conv-TasNet with inter-channel level difference features ☆加藤大輝, 北村大地(香川高専), 矢田部浩平(農工大)

- ◆ステレオ音楽信号は、各音源が左右チャンネル間の音量差によって特定の方位に定位するようにミキシングされている。➡この音源毎の定位の違いが重要な空間的特徴量となる。◆方位分離で得られた空間的特徴量をFILMジェネレータに通し、DNNに補助的に入力することで分離精度向上を期待(Fig. 1)

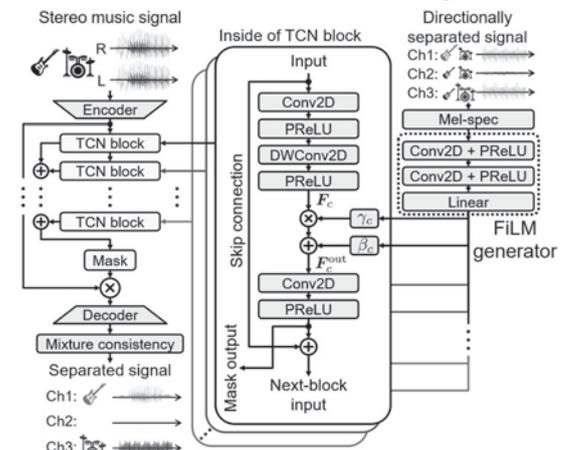


Fig. 1: Overview of the proposed IC Conv-TasNet with FILM-based conditioning.

1-1-5

1-1-5

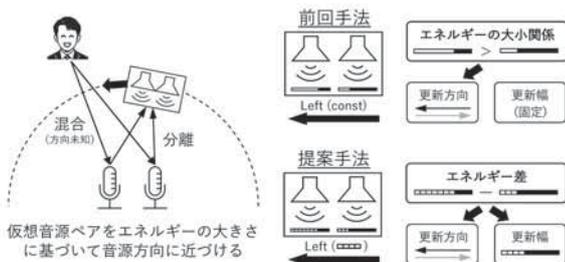
凸最適化に基づく劣決定音源分離における 支配的な音源方向の選択法

Dominant source direction selection method
for underdetermined source separation based on convex optimization

◎ 菅川朋樹, 矢田部浩平 (農工大)

空間辞書を用いたDOA推定と劣決定音源分離

- 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
- エネルギーのピークとなる方向を逐次探索することでDOAを推定, 推定DOAに対応する分離信号から音源を推定する



▶ 適応的な更新により, DOA推定精度と計算コストを改善

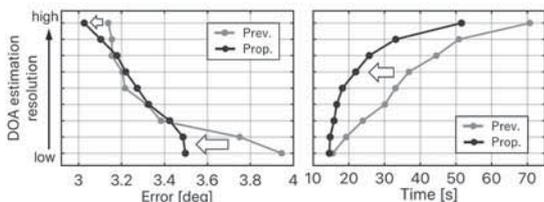


Fig. 1 Comparison of DOA estimation error and processing time

1-1-7

リサンプリングによる音源分離の 性能低下に関する実験的分析

Experimental Analysis of Performance Degradation in
Audio Source Separation Caused by Resampling

◎ 今村 奏海(東大/産総研), 中村 友彦(産総研), 矢田部 浩平(農工大),
猿渡 洋(東大)

◆ 動機: 深層ニューラルネットワークに基づく音源分離では, リサンプリングによる性能低下が確認されているが, その原因は未解明

◆ 仮説: 性能低下に関する2つの仮説を検討

- アップサンプリング後の信号に高周波数成分が含まれないことが性能低下を誘発
- 低周波成分と整合した高周波成分であるかが分離性能に寄与 → ノイズを付加したリサンプリング手法により検証

◆ 結果: 両仮説とも支持

- リサンプリングの補間カーネルにノイズを付加する手法により, 音源分離の性能低下を緩和可能
- 様々な音源分離モデルでも効果を確認

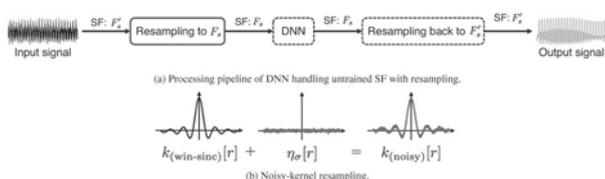


Fig.1: Overview of processing pipeline for proposed resampling methods.

1-1-6

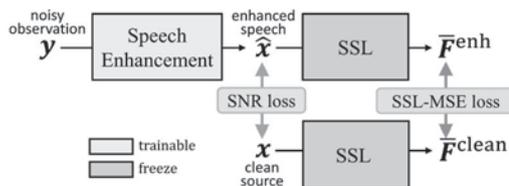
教師なし事前学習モデルを用いた 汎用的な音声強調器の学習

General Speech Enhancement with
Self-Supervised Representation Domain Loss

◎ 佐藤宏, 落合翼, デルクロアマーク, 森谷崇史, 芦原孝典, 増村亮 (NTT)

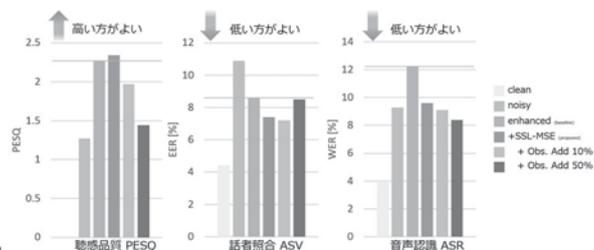
◆ 概要

- ・ 従来シングルチャンネル音声強調は, 各適用先のタスクごとにチューニングが必要で, 未知のタスクへの汎化が困難だった.
- ・ ワンモデルで様々なタスクのフロントエンドとして使用できる汎用的な音声強調器を SSL モデルを使って構築する.



◆ 評価実験 (Average Precision [%])

- ・ 単一の強調モデルで3種類の後段タスクを汎用的に改善



1-1-8

1-1-8

ネコの聴診音信号に含まれる心音成分の強調

Enhancement of heart components in cat auscultatory signals

◎ 岸本龍央, 矢田部浩平, 岩永朋子 (農工大)

背景 ネコの急性心不全による突然死を予防したい

- ▶ 心雑音の有無を聴診によって聞くことが重要



現状 病院では, ネコの緊張による心拍数の上昇や凶暴化がネック

- ▶ 検査に有用な聴診音を採取すること自体が困難となる
- ▶ ネコがリラックスする在宅での聴診が考えられる (代案)
- ▶ 安定した聴診が可能だが, 心臓音が purring 音に埋もれる (嗚鳴らし音, ゴロゴロ音)

提案 畳み込み非負値行列因子分解によって, 聴診音を心臓音と purring 音に分離 (Fig. 1)

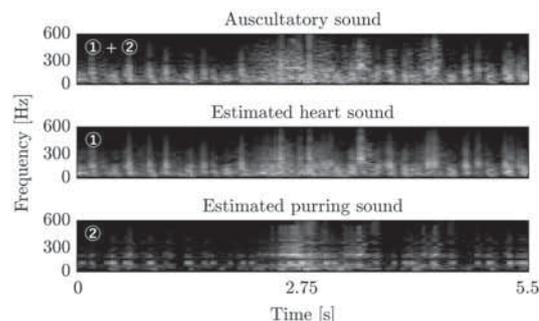


Fig. 1 The heart sound (middle) and the purring sound (bottom) extracted from the auscultatory sound (top)

1-1-9

1-1-9 時間領域ブライント音源分離のための 深層高速フルランク時空間相関分析

Neural Fast Full-rank Covariance Matrix Analysis
for Time-domain Source Separation

☆野崎雄斗, 坂東宜昭 (産総研),
西城耕平 (産総研/早稲田大), △大西正輝 (産総研)

- ◆目的: 周波数相関を活用して深層ブライント音源分離の性能向上
- ◆関連研究: 深層高速フルランク空間相関分析 (Neural FastFCA)
 - フルランク空間共分散行列の同時対角化
 - 高性能だが周波数相関を考慮せず位相情報を活用できていない
- ◆提案法: 深層高速フルランク時空間相関分析
 - 空間相関に加え周波数相関を同時にモデル化する
 - 時間領域の深層音源分離法として解釈できる
- ◆実験結果: Spatialized WSJ0-2mix での分離性能の評価
 - Neural FastFCA と比較して SDR と PESQ で優れた性能を発揮
 - 周波数相関の導入が分離性能と音質向上に寄与することを確認

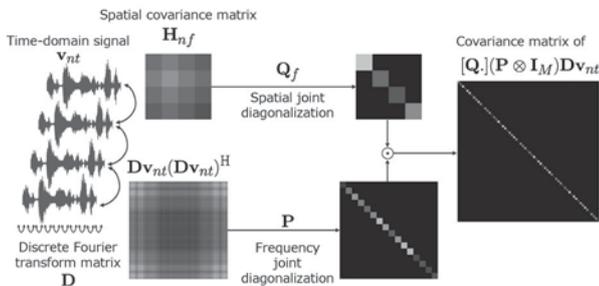


Fig. 1: Overview of the proposed generative model for source signals

1-1-11

1-1-11 キャリア波抑圧型両側帯変調方式と 対数スケール帯域分割に基づく 極小領域オーディオスポットの 音声漏洩低減

Speech leakage reduction in pin-spot audio based on double-sideband modulation suppressed-carrier and logarithmic-scale subband decomposition

☆岩上瑞希 (立命館大院), 歌籠庭, 中山雅人, 西浦敬信 (立命館大)

- ◆パラメトリックスピーカ (Parametric Array Loudspeaker: PAL) を用いた極小領域オーディオスポット形成の音声漏洩を低減するため、キャリア波抑圧型両側帯変調波を対数スケールに基づいて帯域分割する手法を提案する。(Fig.1)
- ◆提案手法では、側帯波を周波数帯域で分割し、分割した帯域を2台のPALに割り当てる。これにより、両PALの波が重なる位置で音声再生され、オーディオスポット外では漏洩音のスペクトルが変化する。
- ◆低親密度語の音声を用い、提案手法で形成した極小領域オーディオスポットと各PALの放射軸上の音声了解度を評価した。その結果、特に男性音声において、オーディオスポットでは高い了解度を維持しつつ、オーディオスポット外では了解度が低下する傾向が確認された。

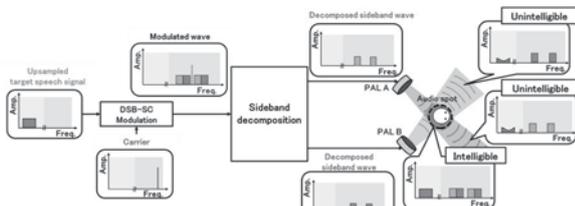


Fig. 1: Overview of the proposed method

1-1-10

1-1-10 重み付き最尤推定法に基づく オンライン型独立低ランク行列分析

Online ILRMA Based on Maximum Weighted Likelihood Estimation

◎石川悠人, 高宗典玄, 中村友彦 (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)

- ◆本稿では新たなリアルタイム音源分離手法として、独立低ランク行列分析 (ILRMA) のオンライン拡張手法である online ILRMA (O-ILRMA) を提案する
- ◆提案法では、まず重み付き最尤推定法 (MWLE) に基づき時間フレーム毎のコスト関数を設計し、計算コスト削減のために一部の中間変数を推定値で近似することでオンライン更新則を導出する
 - さらに、安定した実行のための処理を追加する
- ◆移動音源を含むシミュレーション実験により、O-ILRMA が従来法 (オンライン型独立ベクトル分析: O-IVA-IP/ISS) と比べて高い性能を示し、かつ移動音源に対して頑健に動作することを確認した

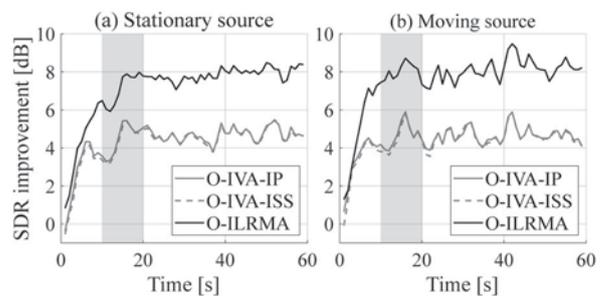


Fig. 1: Average SDR improvements of (a) stationary and (b) moving sources

1-1-12

1-1-12 指向特性を考慮したラメ関数アレイにおける 平面波生成

Plane wave rendering on a Lamé function-based array considering loudspeaker directivity

◎坂口 智弘, 及川 靖広 (早大理工)

- ◆音場合成に関する様々な研究が行われている
 - 実環境への実装には指向性を考慮した音場合成手法が必要
- ◆我々はこれまでラメ関数により設計されたスピーカアレイを用いて平面波生成を行ってきた。
 - 各スピーカの放射に指向性がある場面におけるラメ関数アレイでの平面波生成方法を提案

$$T_L(x; k) = -4\pi \sum_{n=0}^N \sum_{m=-n}^n \sum_{l=1}^L j_n(kx) kr(\alpha, \theta_l, \phi_l) \times e^{ikr(\alpha, \theta_l, \phi_l)} j_n^{(2)}(kr(\alpha, \theta_l, \phi_l)) a_l(k) Y_{nm}^*(\theta_l^r, \phi_l^r) Y_{nm}(x)$$

- ◆指向性を持っているため回転が必要
 - 球面調和関数を回転させることで対応

$$T_{LD}(x; k) = 4\pi \sum_{n=0}^N \sum_{m=-n}^n \sum_{l=1}^L j_n(kx) R_n(kr(\alpha, \theta_l, \phi_l)) \times a_l(k) d_{lm}(k) Y_{nm}^*(\theta_l^r, \phi_l^r) Y_{nm}(x)$$

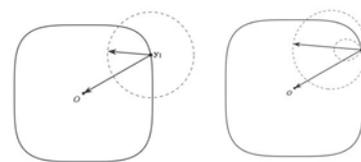


Fig. 1 Location relationship (Left; Loudspeaker has point source, Right: Loudspeaker has directivity)

1-1-13

1-1-13 3次元メッシュRIR測定のための移動式マイクロホンアレイにおける治具反射音の波数領域マスキング

Wavenumber-Domain Masking of Measurement-Induced Reflections in Microphone Array for 3D Mesh RIRs Measurement
○羽田 陽一, 任 逸(電通大)

- ◆アクチュエータで奥行き方向に100 cm, 高さ方向に50 cm 移動可能な48ch 直線マイクロホンアレイを用いて2 cm 間隔の48×51×26 = 63,648個の3次元格子点上の室内インパルス応答(RIR)を測定した。
- ◆アレイを直線に保つために側方と上部に3 cm 角のフレームを設置したが、フレーム治具からの反射・回折音がRIRに混入した。
- ◆特に振幅が大きい直接音に対するフレーム反射音を抑制するため、RIRの初期部においてアレイx軸, z軸に平行入射する治具反射音を波数領域にてマスクする方法を提案する。
- ◆RIRの内挿・外挿等の研究に向け、補正後のRIRデータとアニメをGitHubで公開している。
- ◆<https://yh-audio.github.io/meshgrid-ir.html>

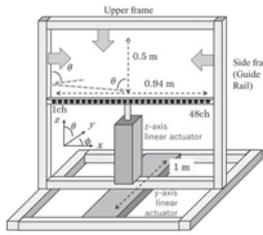


Fig.1: Microphone array system

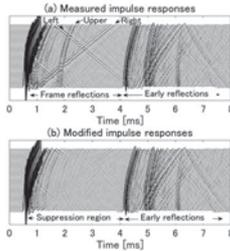


Fig.2: Examples of (a) measured and (b) modified RIRs

1-2-1

1-2-1 インフラ構造物の非破壊検査のためのAIを用いたマルチ打音法の最新動向

AI-based multi-impact acoustics imaging method for non-destructive testing of infrastructure
○歌川紀之(佐藤工業), 黒田千歳(佐藤工業)

- ◆トンネルや橋梁などのコンクリートのインフラ構造物の点検は、構造物の耐久性向上や第三者被害を防止するために重要である。
- ◆打音検査では、ハンマによる打撃音を点検者が聞いて、コンクリート構造物の表面の浮きや剥離などの欠陥の有無を判定する。精度を向上するために、マイクロフォンを用いた打音法の開発を行っている。
- ◆欠陥の有無を判定するために、自己組織化マップ(SOM)とサポートベクトルマシン(SVM)の2種のAIを用いた打音法を用いている。
- ◆ここでは、2種類AIおよび厚さ評価を用いたマルチ打音法により、塩害を受け40年経過したPC桁の打音を行った。その結果、耳で聞く打音検査と同様な結果が得られた。



a) Application area

打音位置	欠陥	健全
1	欠陥	健全
2	欠陥	健全
3	欠陥	健全
4	欠陥	健全
5	欠陥	健全

b) Result of SVM

打音位置	欠陥	健全
1	欠陥	健全
2	欠陥	健全
3	欠陥	健全
4	欠陥	健全
5	欠陥	健全

c) Result of hammering inspection

Fig.1 Result of applying multi-impact acoustic imaging method to PC girder

1-1-14

1-1-14 再生エリアを限定した仮想音源の距離感呈示について

On distance perception of the virtual sound source within the limited reproduction area
☆五島 優太, 羽田 陽一(電通大)

- ◆複数の受聴者に異なる仮想音源を呈示することを目的に、仮想音源の作る音圧分布に窓関数をかけることで再生エリアを制限
- ◆窓関数として、矩形窓(PROP-RECT)と端点が滑らかなハンギング窓(PROP-HANN)を使用
- ◆Fig.1のように再生エリアが異なる複数の移動する仮想音源を同時に呈示し、移動感および再生エリア外に漏出する音のうるさを評価
- ◆Fig.2から、移動感を呈示しつつ、従来法(CONV)よりうるさを抑制できることを示唆

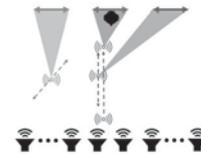
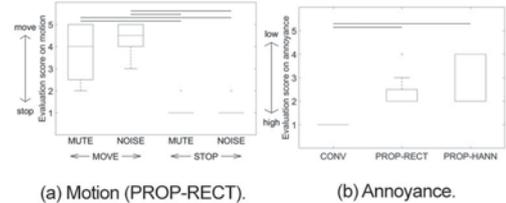


Fig. 1 Arrangement of subjective evaluation.



(a) Motion (PROP-RECT). (b) Annoyance.

Fig. 2 Results of subjective evaluation.

1-2-2

1-2-2 AIが拓く血管・血液のアコースティックイメージング

AI-driven advances in biomedical ultrasound imaging of blood vessel and blood flow
○大村真朗, 高尙策, 長谷川英之(富山大・工)

- ◆医用超音波イメージングのAI活用は、画質向上、信号処理の演算負荷低減、領域分割・分類問題の最適化などが挙げられる。
- ◆血管領域でAIが拓く効果は、より高時間分解能・高S/Nに情報を可視化し、詳細な生体ダイナミクスを解析するためのサポートと考える。血流を強調するためのクラッタフィルタの最適化、ドプラスペクトルをより短い観察窓で忠実に算出するための演算最適化、血管の自動識別などの報告例がある。
- ◆本報告では著者らの取り組み(領域分割、画像再構成)を紹介する。
- ◆血液エコー解析: 敵対的ドメイン適応を組み込んだU-Netを構築し、in silico+血液模擬流体実験の両ドメインで学習することで、血球配向に関わる非スペックル成分を頑健に検出した。
- ◆血流像再構成: 時空間フィルタの計算負荷を軽減するため、フィルタ後出力を教師としてAIで血流を再構成した。フィルタ後と同等のコントラストを得つつ計算負荷低減の可能性を示した。
- ◆血管壁セグメンテーション: 樹状突起構造をU-Net結合層に取り入れることで、高速超音波の平面波イメージングで懸念される内腔と血管壁境界の不連続性・ノイズ頑強性の改善が見られた。
- ◆血管トラッキング: Bモード圧迫法で静脈の潰れをトラッキングし、静脈領域検出・血栓有無の自動判定を目指している。現状は画像処理中心だが、音響学的観点からRF信号処理も検討しつつ、もやもやエコーや血栓特性の抽出を検討していく。

1-2-3

1-2-3 非接触音響探査法における円形空洞欠陥部位のたわみ振動の COMSOL 振動解析 – 共振周波数・欠陥サイズ・深さ・厚みの関係 –

Vibration Analysis of Flexural Vibration of a Circular Cavity Defect in Non-Contact Acoustic Inspection Method using COMSOL - Relationships among Resonance Frequency, Defect Size, Depth, and Thickness -

○杉本和子, 杉本恒美 (桐蔭横浜大院)

- ◆非接触音響探査法は、非接触で遠隔から複合材料の内部欠陥を検出し映像化する手法である。空中音波で対象物を面加振し、その振動速度分布から SSE 解析を行い、欠陥部位の共振周波数を検出する。
- ◆コンクリートに埋め込まれた円形空洞欠陥を本手法で実験した結果、直径 30 cm では、たわみ振動の共振周波数は欠陥深さに比例して増加していた。しかしながら、直径 20 cm ではその傾向が見られなかった。
- ◆この現象の理論的解明のために、COMSOL Multiphysics を用いて、非接触音響探査法の実験を模擬した設定で振動解析を行い、欠陥部位の共振周波数・欠陥サイズ・欠陥深さ・厚みの関係を調べた。

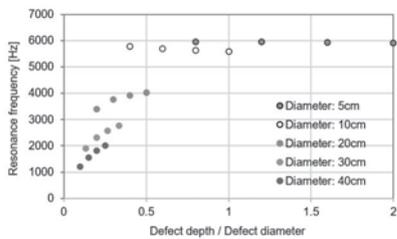


Fig.1: The relationship between "defect depth / defect diameter" and resonance frequency.

1-2-5

1-2-5 レーザ送信マイクロホン受信による弾性波伝搬の映像化に関する検討

Study on visualization of elastic wave propagation using laser transmission and microphone reception

◎清水 鏡介(愛媛大院), △高山 駿介(愛媛大),

△重松 尚太, 中畑 和之(愛媛大院)

- ◆金属薄板に対するレーザー照射によって発生した Lamb 波の伝搬を、薄板近傍に固定設置したマイクロホンで受信することで映像化が可能であるか検討を行った。
- ◆Fig. 1 に映像化結果を示す。Fig. 1 より、異なる波長の波動が伝搬する様子が確認できる。
- ◆Fig.5 に分散曲線を示す。点線線は Rayleigh-Lamb 方程式から計算された A0 モード、実線は音速 340 m/s における空中超音波の理論値を示している。理論値と実験値が一致しており、Lamb 波 A0 モードの伝搬が確認できる。また、Lamb 波と同時に空中超音波の伝搬も映像化されていることも確認された。

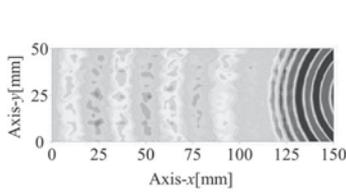


Fig. 1: Visualization results.

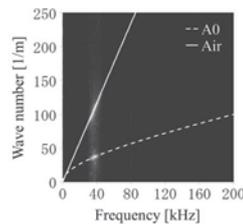


Fig. 2: Dispersion curve.

1-2-4

1-2-4 コンクリート非破壊計測のための非接触音響探査法に関する研究 – 移動計測時の欠陥検出性能の改善に関する検討 2 –

○中川 裕, 杉本恒美, 杉本和子, 上地 樹(桐蔭横浜大院),

△黒田千歳, △歌川紀之(佐藤工業株), △二瓶靖和(富士フィルム株)

- ◆現在、我々は音波照射加振とレーザドップラ振動計(LDV)を用いた構造物の内部に生じる剥離や亀裂欠陥の探査方法について、計測機器を連続的に移動しながら振動計測を行うことができる移動計測を実現するための研究を行っている。
- ◆移動計測時に、光学ノイズの一種であるスペックルノイズが測定波形に記録され、欠陥位置の推定精度が劣化する事が判明した。
- ◆今回はウェーブレット係数の積分値による信号の基準化と時間周波数ゲート(Time and Frequency gate)処理を用いたスペックルノイズ低減方法について検討を行った。
- ◆提案手法を用いる事で、スペックルノイズが低減され欠陥部の応答の判別が容易になった。

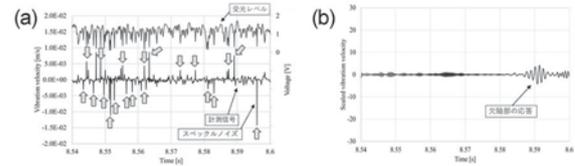


Fig.1 信号処理(基準化、TF ゲート)前後の波形 (a)処理前、(b)処理後

1-2-6

1-2-6 空中非線形超音波励起による金属板を伝搬する波動の位相を利用した欠陥イメージング

Defect Imaging Using Phase of Waves Propagating in Metal Plate by Airborne Nonlinear Ultrasound Excitation

☆石川周男(日大院・理工), 清水鏡介(愛媛大院), 伊藤洋一, 大隅歩(日大・理工)

- ◆強力空中超音波により金属薄板に励起された Lamb 波の高調波成分を用いた微小欠陥の可視化を FEM 解析により検討した。
- ◆波長と欠陥寸法の相似関係に着目し、基本波では検出困難な 10 mm の欠陥に対し、第 3 次高調波の位相分布を用いてイメージングを行った。
- ◆その結果、高調波の位相画像において欠陥形状が明瞭に可視化され、微小欠陥の検出有効性が示された。
- ◆これにより、検査対象の欠陥サイズに応じて使用する次数の成分を選択することで、効率的な欠陥検出が可能となる。

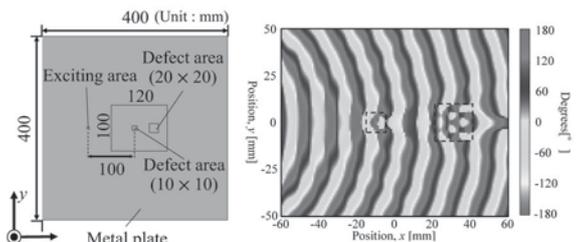


Fig.1 Simulation model (Front side). Fig.2 Phase image (3rd).

1-2-7

1-2-7 空中超音波によるガイド波を用いた非接触厚さ計測

Non-contact thickness measurement using guided waves generated by air-coupled ultrasonic

○田中雄介, △浅川濯 △小倉幸夫(ジャンプローブ)

- ◆空中超音波により板、パイプにガイド波を発生させて厚さを計測した。
- ◆ガイド波の群速度、位相速度が厚さにより変化することを利用した。
- ◆ガイド波は波形のピーク位置が伝搬と共に変化した。
- ◆鋼の厚さを遠距離伝搬時は群速度、近距離伝搬時は位相速度で評価し、200[kHz]では厚くなるほど群速度、位相速度が低下した。
- ◆群速度はガイド波受信波形の包絡線ピーク位置、位相速度は特定ピークの変化を少しずつ追う方法で評価した。

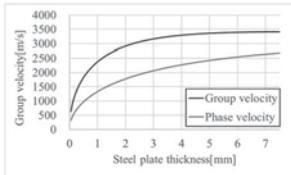


Figure 1 Dispersion curves of phase and group velocity in steel at a frequency of 200 kHz.

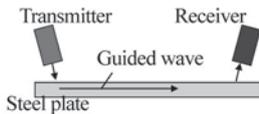


Figure 2 Experimental setup in measuring steel plate.

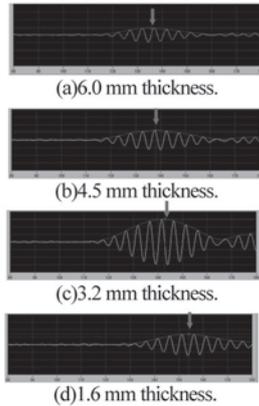


Figure 3 Receiving signals of guided wave at steel plate.

1-2-9

1-2-9 超音波後方散乱特性解析による赤血球集合度の *in vivo* 計測 —伝播減衰に関する検討—*

In vivo measurement of red blood cell aggregation by analyzing ultrasonic backscatter characteristics - Investigation of propagation attenuation -

☆新山晴斗, 荒川元孝, △石井琢郎(東北大).

△八代 諭, △石垣 泰(岩手医科大), △西條芳文(東北大)

- ◆我々は、非侵襲かつ定量的な赤血球集合体サイズ推定法を提案してきた。本手法では、赤血球集合体の後方散乱特性を理論後方散乱特性と整合することでサイズを推定する。後方散乱特性の算出には血液の減衰係数が必要であるが、従来の減衰係数計測法^[1]には精度の点で課題があった。この原因として、計測におけるSN比が十分でなかったためと考える。本報告では、SN比の高い計測法である置換法^[2]の適用について検討した。
- ◆血液模擬ファントムおよび20代健康者1名の手背静脈を対象に、従来法および置換法により減衰係数を算出した上で、サイズを推定した。
- ◆手背静脈に対する *in vivo* 実験の結果を Fig. 1 に示す。(b)置換法では健康者に対する計測として妥当な推定サイズ(5 μm) が得られた。この結果から、置換法の適用によるサイズ推定精度の向上を確認した。

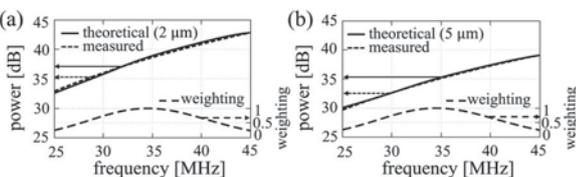


Fig. 1: Measured backscatter characteristics and theoretical backscatter characteristics at the estimated size. (a) previous method^[1]; (b) substitution method^[2].

1-2-8

1-2-8 気泡援用超音波加熱における振幅包絡統計解析による加熱凝固領域推定の基礎的検討

Fundamental Study on Estimating Thermal Coagulation Area using Envelope Statistics in Bubble-Enhanced Ultrasound Heating

☆末田智己(東北大院・医工学研), 森翔平(東北大院・工学研), 吉澤晋(東北大院・工学研/ソニア・セラピューティクス)

- ◆本研究では強力集束超音波(HIFU)治療における加熱凝固領域を、振幅包絡統計解析を用いて定量評価することを目的としている。
- ◆HIFU 照射前後にて取得した加熱凝固領域の超音波イメージングデータに ROI を設定し、二乗平方平均値を用いて規格化した振幅包絡の確率密度分布関数 (PDF) およびモーメントを評価した。
- ◆非スペックル成分を適切に処理した結果、HIFU 照射前後で PDF がシフトし、3次モーメントの値に有意な変化が見られ、加熱凝固による組織性状変化を捉えられたことが示唆された。
- ◆今後の展望は、病理画像との比較を行い、組織性状変化と超音波散乱特性変化の対応関係を明らかにすることである。

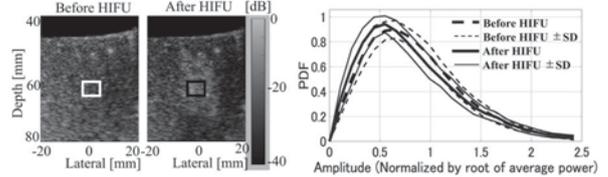


Fig. (Left) B-mode images before and after HIFU exposure (Right) Comparison of envelope PDFs before and after HIFU exposure

1-2-10

1-2-10 試料の減衰を考慮した超音波顕微鏡による生体組織の音響特性計測

小林玲奈, ○荒川元孝(東北大・医工), 小林和人(本多電子), 石井琢郎,

西條芳文(東北大・医工)

- ◆超音波顕微鏡による音響特性(音速、音響インピーダンス、減衰係数)の計測において、試料の減衰を考慮した方法を検討した。
- ◆Fig. 1(a)に従来法と提案法によるヤギ頸動脈切片の二次元音速分布を示す。提案法により、不自然に大きい/小さい点の音速値が改善された。
- ◆Fig. 1(b), 1(c)にそれぞれ音響インピーダンス、減衰係数の測定結果を示す。提案法により、特に減衰係数でばらつきが小さくなった。
- ◆得られた音響特性を用いて振幅スペクトルを計算し、測定値と比較した結果、提案法により精度の高い音響特性が得られていることを確認した。

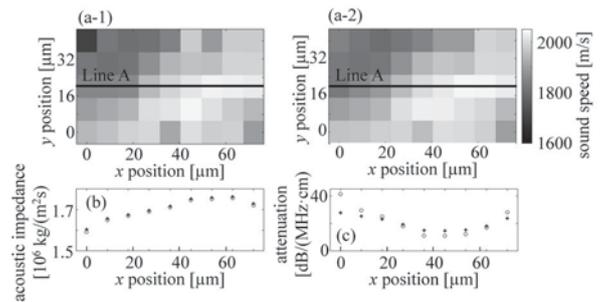


Fig. 1: (a) Two-dimensional sound speed distributions by the (a-1) conventional and (a-2) proposed methods. (b) Acoustic impedances and (c) attenuation coefficients along line A. (C: conventional method, +: proposed method).

1-2-11

1-2-11 光学式固有音響インピーダンス計測法を用いた生体模擬材料の計測

Development of an Optical Intrinsic Acoustic Impedance Measurement Method for Biomimetic Materials

○田村和輝(横浜国大)

- ◆生体を定量評価する物理量の一つに固有音響インピーダンスがある。
- ◆本研究はレーザー超音波法と半透明基板を組み合わせ、非接触で試料の固有音響インピーダンスを推定する。
- ◆試料を保持する半透明基板は、超音波パルスを発生させる527nmのナノ秒パルスレーザー光を吸収する光吸収層と、超音波パルスの反射エコーによって生ずる基板の面外変位を計測する633nmのプロンプレーザー光を反射する光反射層を積層した。
- ◆半透明基板上に生体を模擬した水寒天または20wt%NaCl寒天を置き、基板を機械走査して音響反射率の分布を計測した。
- ◆水とNaCl寒天の反射率は 0.23 ± 0.05 と 0.18 ± 0.05 で、サンプルの固有音響インピーダンスの平均値は1.53 MRayl(誤差+3%)と1.70 MRayl(誤差+17%)と推定された。

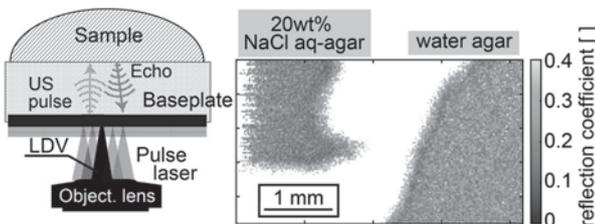


Fig.1: Measurement principle and calculated acoustic reflection coefficient.

1-2-12

1-2-12 培養ディッシュの超音波振動がMC3T3-E1細胞の成長に与える影響

Effects of Ultrasonic Vibration of Culture Dishes on the Growth of MC3T3-E1 Cells

☆早川優太郎, 市川崇崇, 小山大介(同志社大・理工)

- ◆ディッシュに培養した骨芽細胞に分化する細胞であるMC3T3-E1細胞に対して91kHzの超音波たわみ振動を加えることで、超音波が細胞成長に与える影響を検討した。
- ◆細胞の分化後に形成される石灰化基質の形成量を顕微鏡で撮影し、位相差画像の輝度によって評価した。
- ◆細胞の配向に関して有意な変化は観測されなかった。一方、超音波照射により石灰化基質の量は増加し、超音波によって細胞の石灰化が促進する可能性が示唆された。
- ◆今後、形成した石灰化基質の結晶配向および骨密度の評価、異なる条件下における細胞の配向制御について検討する予定である。

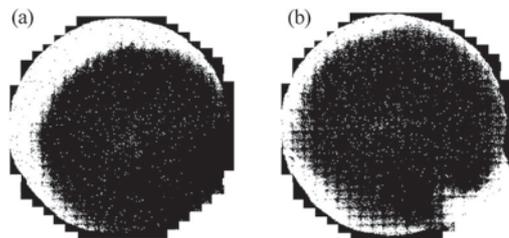


Fig. 1: Binarized images of the cultured cell ((a) Control, (b) under ultrasonication).

1-3-1

1-3-1 (招待講演)九州大学における高大連携事例

(Invited lecture) High school-university collaboration case study at Kyushu University

○(九州大・芸工)

- ◆九州大学の高大連携事例として、「九州大学未来創成科学者育成プロジェクト(QFC-SP)」について概説する。著者自身のQFCリサーチ活動についての概要も述べる。



Fig.1: The web page of the "Kyushu University Future Creators in Science Project."

1-3-2

1-3-2 (招待講演)中高大連携の実践事例報告

(Invited Paper) Case Studies of Secondary-University Collaboration: Practices and Experiences.

○鐘ヶ江暢子(工学院大学附属中学・高等学校)

- ◆中学生・高校生の大学進学において、偏差値を中心とした従来の進路選択は、必ずしも個々の学習意欲や将来の目的と一致しない場合がある。近年では、学びの内容や教育理念を重視した大学選択の重要性が指摘されている。本実践報告では、中学生・高校生が将来学びたい分野を早期に意識し、大学での学びに触れる機会を持つことが、進学後の学習満足度や大学との適合性にどのような影響を与えるのかについて述べる。これにより、中高生と大学とのより良いマッチングを実現する進路選択の在り方について、実践を報告し、進路選択を考える中高生や関係者にとっての一つのきっかけとなることを目指す。



Fig. Examples of opportunities to experience learning at university.

1-3-3

1-3-3 (招待講演)高校生による音声生成の実験を指導した経験を通じて

Best practices for supervising a high school research project in speech production

○荒井隆行(上智大・理工)

- ◆ 昨今、高大連携の動きが強まる中、高校生が大学の教員と共同で研究を進めるケースも増えてきている。
 - ◆ そのような状況の中、私が実際に関わった高校生の研究プロジェクトを指導するという事例を中心に、その意義と課題について議論。
- ↓
- ◆ 本事例は、立命館慶祥高等学校の当時3年生であったOさんとの共同研究。
 - ◆ 最初に問い合わせがあったのは2021年7月、2021年10月に生徒本人、そして当時指導をされていた先生と共にZoomによるミーティング。
 - ◆ 母音を系統的に網羅した測定という点で新規性があるものと判断し、その成果を日本音響学会の研究発表会で発表することに。
 - ◆ 今後、本学会においても高校生や中学生との連携が深まり、その中で培われた優秀な研究については、研究発表会の場などにおいてその成果が発信されることを期待。
 - ◆ 同時に、注意すべき点については引き続き議論を。

1-3-5

1-3-5 研究発表会における高校生参加の現状と展望

Current states and prospects of high school students' participation in ASJ meetings

○水町 光徳(九工大)

- ◆ 日本音響学会研究発表会は、高校生も参加及び研究発表が可能である。
- ◆ 日本音響学会は、高校生が参加登録費免除で研究発表会へ参加できる制度を設けており、これまでの高校生参加の実績をFig. 1に示す。
- ◆ 第152回(2024年秋季)研究発表会では、高校生による研究発表の優遇措置として、高校生と本会正会員との共著発表(4件)を試行した。
- ◆ 今後の展望として、音響学と日本音響学会の発展のための研究発表会における高校生発表の制度化について議論する。

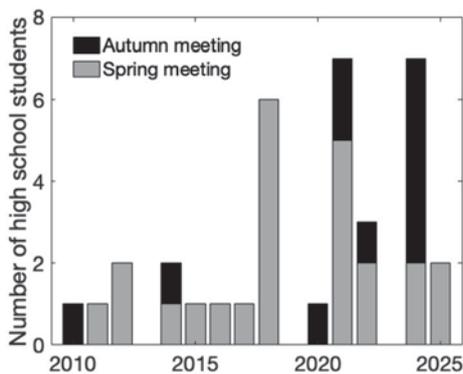


Fig.1: Number of participating high school students. Black and gray bars indicate the number of high school students who participated in autumn and spring meetings, respectively.

1-3-4

1-3-4 音響学会研究発表会において 高校生の楽器音響研究の講演をサポートした事例について

Supporting a high school student's presentation on musical instrument acoustics at the Acoustical Society of Japan Research Presentation

○亀川 徹

- ◆ 2024年秋季研究発表会で実施した高校生による楽器音響研究の講演のサポートに対応した事例について紹介した。
- ◆ 音響学の裾野を広げる意味でも今後大いに取り組む意義がある一方で、下記のような問題点が明確となった。
 - 申し込みをどのように受け付けるか?
 - 研究を進める上での基本知識のチュートリアル
 - 事前の準備や指導にどこまで関与するか?
 - 進路の希望との対応
- ◆ 今後音響学会としてどのように取り組んでいくかといった議論を深めていく必要がある。

1-3-6

1-3-6 音による行動変容と空間デザインの応用に関する研究動向の調査

A survey of research trends on behavioral change by sound and its application to spatial design

◎清家 裕喜子, 石塚 崇(清水建設技研)

- ◆ 音を「ナッジ」として建物の空間デザインに組み込み、空間利用者の行動変容を誘発する手法を検討している。
- ◆ 音による行動変容と空間デザインへの応用に関する先行研究を調査し、研究動向を整理した。国内では2000年代以降に本格化し、近年その研究数は増加傾向にある。
- ◆ 音楽のテンポやジャンルが、消費・購買行動に及ぼす影響の研究が最も多く(Fig.1)、BGMのテンポが滞在時間や売上に寄与することや、音楽ジャンルによって商品選択に影響があることが報告されている。
- ◆ 次いで都市空間等における個人や群衆の歩行・滞在行動を対象とした研究が多く(Fig.1)、歩行速度や滞在時間への影響が報告されている。
- ◆ 実環境における心理プロセスの解明やBGM以外の音刺激の検証、長期的な提示による「慣れ」の影響など、空間デザインとして活用するための検証について、検討の余地がある。

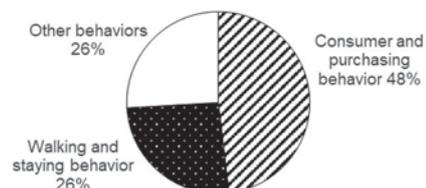


Fig.1: Classification of target behaviors in previous literature.

1-3-7

1-3-7 主観と客観評価に基づく音環境の改善による疲労感の軽減

Reduction of Fatigue Through Improvement of the Sound Environment Based on Subjective and Objective Evaluations

☆鬼頭風(中央大院), 山口雅夫, 戸井武司(中央大)

- ◆ 昨今、様々な空間に応じて適切な音環境を実現する試みが進められているが、疲労負荷に対しての配慮が十分にされていない。そこで本研究では、様々な定常的な音を対象に行う主観および生体情報を用いた客観評価に基づき、音環境の疲労感の評価を検討した。また、音環境の音響特性を変化させ、疲労感の増加の軽減を検討した。
- ◆ 様々な音の主観評価と音響特性から重回帰式を導出し、疲労感に関連すると考えられる集中因子にシャープネスが強く影響することを把握した。
- ◆ Fig.1に示す主観評価の結果より、自動車の定速走行音 C1 の 1.5~3.0 kHz の音圧レベルを 15 dBA 低減し、全体の音圧レベルを 65 dBA に調整することでシャープネスを 1.38 acum から 1.28 acum に低減した C1_Flt は、C1 と比較して集中因子の評価点が向上することを把握した。
- ◆ Table 1 に示す脈波 (LF/HF) による客観評価の抜粋より、LF/HF は時間経過により増加する傾向がみられ、疲労感の増加を確認できる可能性を示した。また、区間前後の LF/HF 差を比較すると、主観評価における集中因子の評価点が高い C1_Flt は疲労感が増加しにくいことが示唆された。

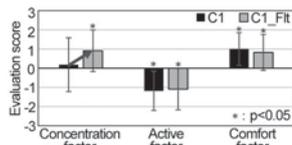


Fig. 1: Subjective evaluation of car sounds

Table 1: Objective evaluation of car sounds

Subjects	Evaluation sounds	LF/HF		
		2~5 min	7.5~10.5 min	Difference
Subject-1	C1	2.50	6.08	3.59
	C1_Flt	0.33	0.69	0.37
Subject-2	C1	1.42	4.57	3.15
	C1_Flt	0.40	0.78	0.38
Subject-3	C1	0.07	1.17	1.10
	C1_Flt	1.84	0.25	-1.58

1-3-8

1-3-8 保育施設周辺の住宅地域における音環境と保育施設に関する意識調査

Questionnaire Survey on Sound Environment and Childcare Facilities in Residential Areas Surrounding Childcare Facilities

☆黒木幸穂(九州大・芸工), 片岡寛子(電通大), 高田正幸(九州大・芸工)

- ◆ 保育施設が新設される際に音をはじめとする環境の変化を懸念した住民が反対し、計画が頓挫した事例が報告されている。保育施設の音は、新設の際に既存の施設でも住民との摩擦の原因になることがある。
- ◆ 先行研究では、保育施設新設への賛否が、施設との関係性や騒音感受性といった個人の特性と関連することが示された。また、首都圏での調査では、居住年数やソーシャルキャピタル(社会関係資本)に関する意識との関連もみられた。本研究では、住民の保育施設新設への賛否と、既存の施設への音に起因する苦情に着目した社会調査を行った。
- ◆ 音環境の実測調査と調査票を用いた住民の意識調査を実施した。調査票には、新設賛否を含む先行研究と同様の質問に加え、保育施設から発生する音に対する苦情行動に関する質問を追加し、保育施設新設への賛否や施設に対する苦情行動と関係する各種要因の検討を行った。
- ◆ 結果として、保育施設新設への賛否には回答者の住宅所有形態、騒音感受性、公開行事の認知、公開行事への参加意思といった個人の特性や保育施設との関係性が関連することがわかった。また、保育施設や行政に対する苦情行動に關係する要因の検討では、騒音感受性、新設への賛否、近隣住民との面識・交流の程度が苦情行動に関連することがわかった(Table1)。

Table1 Results of logistic regression analysis on complaint behavior

	B	p	EXP(B)	95 %CI
Noise sensitivity	1.268	0.003	3.554	[1.531, 8.252]
Attitude toward the establishment	1.218	0.000	3.380	[1.834, 6.229]
Degree of acquaintance and interaction with neighbors	-0.375	0.007	0.687	[0.522, 0.904]

1-3-9

1-3-9 2 音声同時提示における音源位置が聴取能力に及ぼす影響の調査

Impact of Sound Source Location on Listening Performance During Simultaneous Presentation of Two Speech Signals

◎宮本 佳奈, 矢澤 櫻子, 山口 仁, 野口 賢一, 鎌本 優 (NTT)

- ◆ 目的
水平面上の全方位にわたる 2 音声同時提示において、音源位置が聴取能力に及ぼす影響を調査する
- ◆ 実験
1. 8 つのスピーカの任意の 2 点から単語を読み上げた音声を提示し、単語の聞き取り成績から聴取しやすい音源位置を調査 (Fig.1)
2. 聞き取り成績が高い条件群と低い条件群で脳波を解析
- ◆ 結果
1. 90° 以上音源位置が離れている場合に聞き取り成績が高くなった
1'. ただし、同一矢状面からの音源は聞き取り成績が低くなった
2. 聞き取り成績に応じて脳活動に差が生じた (Fig.2)

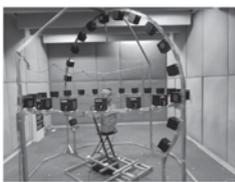


Fig.1: Experimental environment

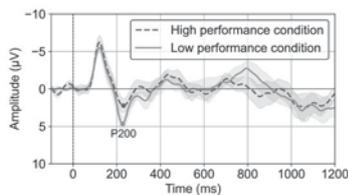


Fig.2: Grand-average ERP waveforms recorded from Pz

1-3-10

1-3-10 音響マルチパラメータを用いたニューラルネットワークによる主観的印象予測モデルの構築

Prediction Model for Subjective Impressions Using Neural Networks with Multiple Acoustic Parameters

☆中江翔飛(中央大院), 田辺総一郎, 山口雅夫, 戸井武司(中央大)

- ◆ 説明変数に音響マルチパラメータ画像、目的変数に SD 法による主観的印象評価得点を採用したニューラルネットワークを用いて、音の印象予測モデルを構築する。次に、作成した学習済みモデルに対して SHAP を適用することにより、音響パラメータごとの予測に対する貢献度を算出し、音の印象形成に寄与する要因を可視化する。
- ◆ 基礎検討として、ブラウンノイズと純音を組み合わせた定常模擬音を対象にモデルを構築し、本モデルが高精度に印象評価得点を予測できることを示した。また、SHAP により印象に寄与する音響パラメータを可視化できることを把握した。
- ◆ 実用検討として、実際の自動車加速音を対象に印象予測モデルを構築し、複雑な非定常音に対しても、Fig.1 のように高い精度で印象評価得点を予測できることが明らかとなった。また、Fig.2 に示す SHAP により、Sedan_3D は音の粗さ感が弱く、穏やかさが高いということがわかった。

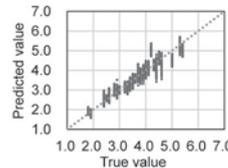


Fig. 1: True vs. predicted calm

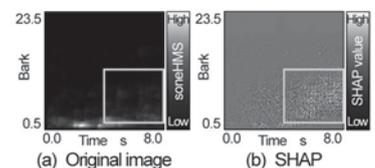


Fig. 2: Roughness of Sedan_3D

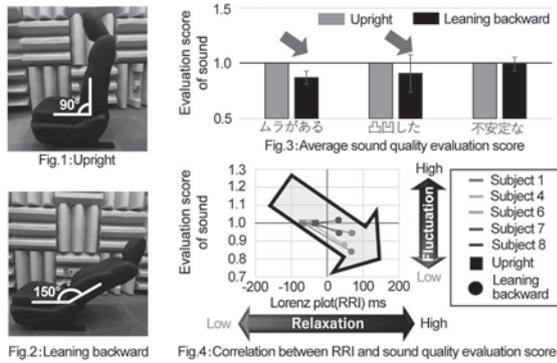
1-3-11

1-3-11 シート姿勢による心理状態が音知覚に及ぼす影響

The Effect of Psychological State due to Seat Posture on Sound Perception

☆大平朋輝(中央大院), △佐藤篤司, 新堀圭一郎(トヨタ紡織), 田辺総一郎, 戸井武司(中央大)

- ◆自動運転化によって多様化する車室内姿勢を想定し、姿勢による心理状態が音知覚に及ぼす影響を検討する。
- ◆Fig.1 およびFig.2に示す、直立座位と後傾座位における心理状態のリラクセス度を、心電計測で評価した。
- ◆Fig.3 より直立座位と後傾座位において、同じ音でも異なって聞こえる傾向が見られた。特に、約5-10 Barkの周波数帯域における音の変動特性が姿勢によって変化する傾向が見られた。
- ◆Fig.4 より、リラックスしている時は音の変動感を小さく感じる傾向が見られた。



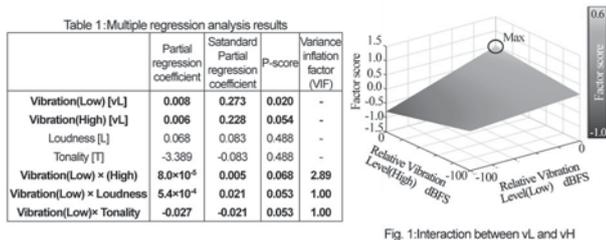
1-3-13

1-3-13 車両振動の有無による感性サウンドマップに基づく自動車走行音の特徴分析

Analysis of Vehicle Noise Characteristics Based on Affective Sound Maps with and without Vehicle Vibration

☆中里峻人(中央大院), 田辺総一郎, 戸井武司(中央大)

- ◆加速音の印象と心理音響メトリクスとの関係を検討し、感性サウンドマップを作成することで、人間の感性の可視化を行う。
- ◆模擬的に作成したEV加速音に対し、振動を付加すると、「躍動」がより想起された。
- ◆EV1 およびEV2 では「躍動」に対して、寄与が高いとされる周波数帯域が異なることが分かった。
- ◆Table 1に示す重回帰分析により、「静穏 - 躍動」軸に対して振動の低周波(1-30 Hz)の主効果だけでなく、先行研究より特徴的な心理音響メトリクスであるラウドネス、トナリティとの交互作用の影響を受けていることが示唆された。
- ◆重回帰分析に基づいて算出した定量式をもとにFig. 1に示す応答曲面グラフを算出し、「静穏 - 躍動」軸に関わる静穏因子が最大となる適切な条件を確認した。



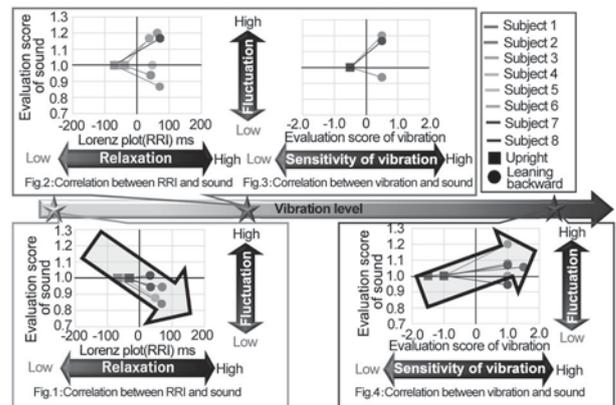
1-3-12

1-3-12 走行振動の有無による心理状態が音知覚に及ぼす影響

The Effect of Psychological State due to The Presence or Absence of Running Vibration on Sound Perception

☆大平朋輝(中央大院), △佐藤篤司, 新堀圭一郎(トヨタ紡織), 田辺総一郎, 戸井武司(中央大)

- ◆自動運転化によって多様化する車室内姿勢を想定し、姿勢による心理状態、および体感振動が音知覚に及ぼす影響を検討する。
- ◆Fig.1 およびFig.2 より、振動がない時はリラックスするほど音の変動感を小さく感じる傾向が見られた。
- ◆直立座位に比べて後傾座位で振動を強く感じる傾向が見られた。
- ◆Fig.3 およびFig.4 より、小さな振動が加わる場合は、体感振動と音の変動感知覚に関係性が見られないが、大きい場合は振動を強く感じるほど音の変動感も大きく感じる傾向が見られた。



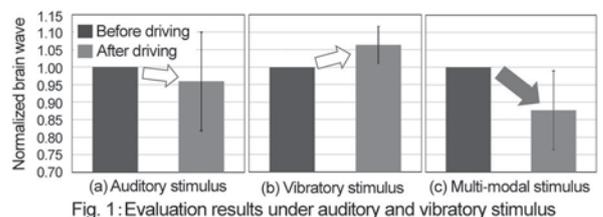
1-3-14

1-3-14 脳波を用いた疲労感軽減のための運転時音環境の提案

Proposal of an in-Vehicle Acoustic Environment to Reduce Driver Fatigue Based on EEG Analysis

☆大久保勇祐(中央大), 加曾利拓真(中央大院), 山口雅夫, 戸井武司(中央大)

- ◆脳波の High α 波(10~13 Hz)に基づく疲労感評価手法は提案されているが、振動環境下での検証はなされていない。本研究では、路面状態に起因する音と振動の違いが疲労感の軽減に及ぼす影響を明らかにすることを目的とし、脳波の High α 波に基づく振動環境下における疲労感評価の有用性を検討した。
- ◆先行研究では、振動が大きい条件ほど疲労感が増加することが報告されている。本研究の実験結果においても、振動が大きい条件で High α 波が減少し疲労感が増加していたことから、振動環境下においても脳波の High α 波を指標として疲労感を評価できる可能性が示唆された。
- ◆Fig. 1に示す音刺激(a)、振動刺激(b)および両者の複合刺激(c)を提示した時の High α 波は、単刺激では疲労感の低下がみられなかったが、複合刺激下で両者の感覚印象に乖離が生じる場合は、疲労感が増大した。そして、複合刺激を考慮した環境設計の重要性が示唆された。



1-3-15

1-3-15 姿勢および聴覚特性を考慮した 在宅医療機器の稼働音の快適性評価

Research on Pleasantness in Home Medical Devices Considering Posture and Auditory Characteristics

☆神山紗和(中央大院), 山口雅夫, 戸井武司(中央大)

- ◆在宅医療機器の稼働音を対象とし、姿勢(座位・仰臥位)および周囲環境の違いが音質に与える影響を、生理的・心理的、聴覚特性および寝具を含めた部屋の音響特性の観点から検討した。
- ◆脈波を用いた生理的および主観的評価により、仰臥位ではリラクセーション度が高くなること示された。
- ◆仰臥位では特定の周波数帯の聴覚特性の影響 (Table 1) により、シャープネスが大きい稼働音 (Fig. 1) ほど、耳障りになり快適性が低下した。また、住環境を想定した評価室では、部屋の反響や寝具の吸音の影響により、無響室とは異なることが示された (Fig. 2)。
- ◆実使用環境を想定した条件下で評価することの重要性が示唆された。

Table 1: Evaluation result of significant differences by test statistics

Frequency [Hz]	125	250	500	1000	2000	4000	8000
Left ear	0.75	0.39	0.67	0.72	0.34	0.07	0.24
Right ear	0.79	0.59	0.75	0.02	0.05	0.72	0.65

∗: P値<0.05, ∗∗: P値<0.01

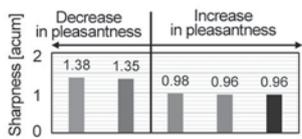


Fig. 1: Sharpness of evaluation sound

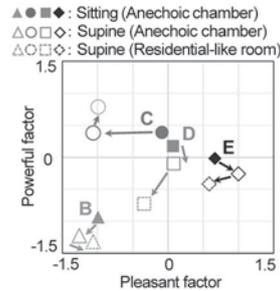


Fig. 2: Factor scatter plot (Posture and room effects)

1-3-17

1-3-17 複数純音を有するEVインバータ音の マスキング効果の検討

Investigation of Masking Effects in EV Inverter Sounds with Multiple Pure Tones

☆加曾利拓真(中央大院), 大久保勇祐, 金堂雅彦, 山口雅夫, 戸井武司(中央大)

- ◆インバータから発生するキャリア周波数およびその側帯波による複数純音が広帯域雑音下で知覚されるメカニズムについて、構成純音が臨界帯域の内外に存在する場合に着目し、そのマスキング効果の違いを明らかにした。
- ◆複数純音が同一の臨界帯域内に存在する場合、最小可聴突出量は、キャリア周波数のLpから、帯域内の純音の総パワー値の増加量を差し引いた値と推定できる。一方、側帯波がキャリア周波数の臨界帯域外に位置する場合は、構成純音のなかで周波数が最も低い側帯波の最小可聴突出量によって推定できることが判明した。
- ◆低速走行時では、暗騒音レベルが低く、キャリア周波数と側帯波が同一の臨界帯域内に収まるため、単一純音の最小可聴突出量よりも数dB低いレベルで知覚される可能性が示された。本成果は、自動車の低速走行時におけるインバータの音質設計の重要性を意味する。

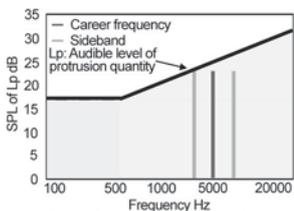


Fig. 1: Pure tones of equal level

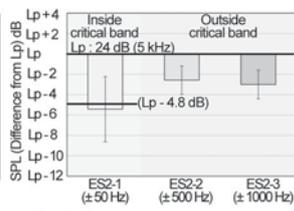


Fig. 2: Evaluation results of multiple

1-3-16

1-3-16 コーンの水分量変化による喫食音が 喫食に与える印象

The Impression of Eating Sounds Caused by Changes in the Moisture Content of Soft Serve Ice Cream Cones

☆水沼紗愛(中央大院)

△木村和紀, △廣田智統(日世株式会社), 戸井武司(中央大)

- ◆コーンの水分量変化から、喫食音が喫食時に与える印象を把握する。
- ◆水分量の増加に対する気導音と骨導音のパワー比の変化は、生地の硬度が高い(Fig.1 B)は気導音の割合が高くなり、硬度が低い(C)は骨導音の割合が高くなること示された。
- ◆喫食音の評価から、水分量の増加により気導音の割合が高くなる(Bs)は明瞭因子が高くなるのに対し、骨導音の割合が高くなる(Cs)は明瞭因子が低くなること示された。
- ◆喫食音の評価(■)と喫食時の評価(□)の比較では、水分量の増加により気導音の割合が高くなる(B)は喫食時の変化量が大きくなるのに対し、骨導音の割合が高くなる(C)は喫食音の変化量が大きくなること示された。

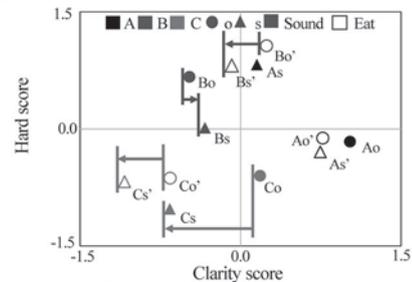


Fig. 1 Clarity and hard plots of sound

1-3-18

1-3-18 製品におけるふさわしい音の 特徴抽出および評価手法の構築

Development of Feature Extraction and Evaluation Methods for Appropriate Product Sounds

☆八木陸斗, △松本和也(中央大), 大平朋輝(中央大院), △山本弘, 田辺総一郎, 戸井武司(中央大)

- ◆従来の製品音評価は、騒音レベルなど物理的特性のみに着目したものが多かったが、近年では、製品の機能感と快適感を両立させた「ふさわしい音」を目指す企業が増えている。そこで本研究では、スティッククリーナの「ふさわしい音」の特徴を明らかにした。
- ◆作成した吸引圧力調整機構を用いて複数の吸引圧力を基準化し、EPA構造に基づいて選定した形容詞対によるSD法で主観評価を行った。快適因子と吸引因子の感性マップを作成し、評価音の心理音響メトリクスと比較することで、「ふさわしい音」の特徴が音圧レベル、シャープネスおよびトナリティで示されることを明らかにした。
- ◆録音したスティッククリーナの動作音に対して、ふさわしい音の特徴に基づき音圧レベルの上昇、シャープネスおよびトナリティの減少により改善した。改善音の主観評価を実施し、快適感、吸引感が向上したことから、抽出した特徴が妥当であることを検証した。



Fig. 1: Tonality of M2



Fig. 2: Tonality of M2'

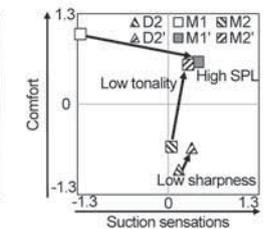


Fig. 3: Sensibility map of improved sounds

1-3-19

1-3-19 袋詰めインタラクションにおける副次的発生音の質感評価 その2

Evaluating the impression of the bagging sound part2

☆佐藤大晟, 黒田隼海(神奈川工科大), 三浦貴大(産総研), 上田麻理(神奈川工科大)

- ◆本研究では、製品を梱包する際に発生する付属音と映像が、購入者の購買意欲および支払意思額(WTP)に与える影響を、仮想評価法(CVM)を用いて明らかにすることを目的とする。
- ◆実験では、お菓子と文房具を対象に、袋の種類や提示条件(映像のみ、音のみ、音と映像)を変化させた動画を用い、評価を行った。その際梱包の意図を説明した教示の有無による効果も検証した。
- ◆分析の結果、適切な教示を与えることで初期提示額に対する購入希望率のオッズ比が2.69倍、WTPの中央値は約2.5倍に向上し、適切な意味付けが金銭的評価に不可欠であることが示唆された。

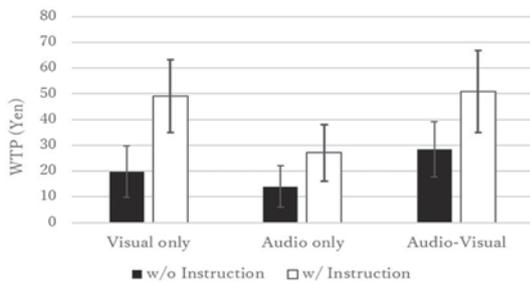


Fig. 1: Differences in evaluation scores by presentation conditions and instructions.

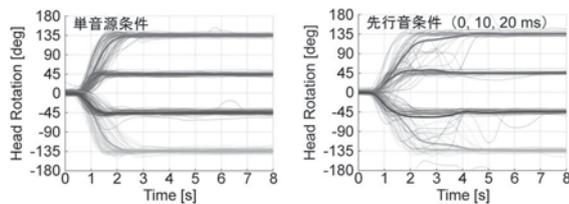
1-4-2

1-4-2 先行音効果を利用した避難誘導時における被験者の頭部挙動の検証

Investigation of Participants' Head Movements During Evacuation Guidance Using the Precedence Effect

☆一色暁光(日大院・理工), 伊藤洋一, 大隅歩(日大・理工)

- ◆音声を用いた避難誘導システムの研究を行っている。これまで、先行音効果を利用した音声避難誘導において、避難者の頭部運動が方向感評価に与える影響は十分に検討されていなかった。
- ◆本研究では、実空間(13×16×3 m)を模擬したVR環境下で、HMDにより被験者の頭部位置・姿勢を記録し、頭部挙動と方向感評価の関係を検証した。
- ◆単音源条件と先行音条件を比較した結果、先行音条件では目標方向への収束が遅れ、目標角通過後に戻ると修正動作が散見された。
- ◆以上より、頭部運動を分析することは、「方向感の明確さ」を客観的に評価する指標として有効となる可能性がある。



(a) Single-source condition. (b) Precedence-effect condition.

Fig. 1: Head movements during evacuation guidance.

1-4-1

1-4-1 パノラマ写真によりVR表示された建築空間の視覚印象に調和する残響時間の提案

Proposal of reverberation time to match the visual aspect of architectural space displayed in VR using panoramic photographs

☆千葉妙, 石川あゆみ(岐阜高専)

- ◆我々の研究室では、人が建築空間の視覚刺激(見た目)から予想した響きを“予想残響時間(以下 ERT)”と定義し、これを被験者実験で測定してきた。
- ◆先行研究では、パノラマ写真によりVR表示された建築空間の視覚刺激に調和する残響時間を求めるモデルとして、実験における対象空間の容積とそのERTの散布図が提案された。
- ◆本報では、既報で得られた知見(空間構造および空間の既知・未知で分けてモデルを検討すべき)を反映した新たなモデルを提案した結果を示す。

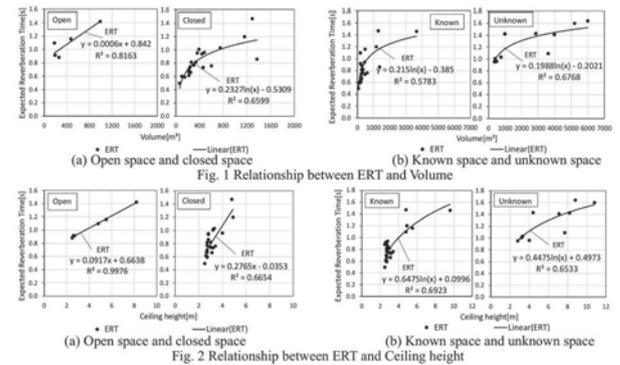


Fig. 1 Relationship between ERT and Volume. Fig. 2 Relationship between ERT and Ceiling height.

1-4-3

1-4-3 室内吸音設計規準に関する騒音下の音声伝達の聴感実験 - 等価床面吸音率と聴き取りにくさ・うるささとの関係 -

Auditory experiment on speech transmission in noise for indoor sound absorption design standards -Relationship of equivalent floor absorption coefficient with listening difficulty and annoyance

◎陳科吉, 佐久間哲哉(東大・工), 佐藤逸人(神戸大・工)

- ◆AIJESとして「建築空間の吸音設計規準・指針」室区分Iの推奨値として等価床面吸音率に関して、幾何音響解析により模擬音場を用いて聴き取りにくさ、うるささの聴感実験を行い、その関係性を検証した。
- ◆ U_{50} では吸音率によらず、初期反射音によって解析値の方が大きく、特に小床面積では顕著である。
- ◆ α_{f0} の低下に伴い、聴き取りにくさ・うるささの上昇が確認できる。床面積の増加に伴い、うるささへの影響は小さいが、聴き取りにくさは明確な上昇が見られる。
- ◆ U_{50} とLDRの回帰曲線により、理論値では床面積の大小により曲線の上方・下方への乖離が顕著である。一方、解析値では低騒音下でややLDRが高い傾向にある。

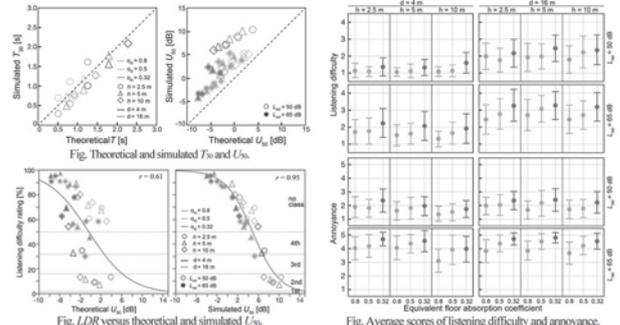


Fig. Theoretical and simulated T_{50} and U_{50} . Fig. LDR versus theoretical and simulated U_{50} . Fig. Average scores of listening difficulty and annoyance.

1-4-4

1-4-4 飲食店を対象とした
会話しやすさの実験的検討
Experimental Investigation of Conversational Ease
in Foodservice establishments

☆中村天馬(神奈川工科大), 池上卓也(USEN), 三浦貴大(産総研),
田中哲雄, 上田麻理(神奈川工科大)

- ◆本研究では、物理的な「静けさ」は必ずしも飲食店の利用者にとって快適とは限らないとの考えのもと、利用者にとって「快適な会話環境」の要件を多角的に検証することを目的とした。
- ◆本研究では物理的条件を統制可能な実験手法を採用し、会話テーマの「秘匿性」と「騒音レベル」が会話しやすさに与える影響を調査した。
- ◆一定の騒音レベルがあったほうが話しやすい環境がある！？

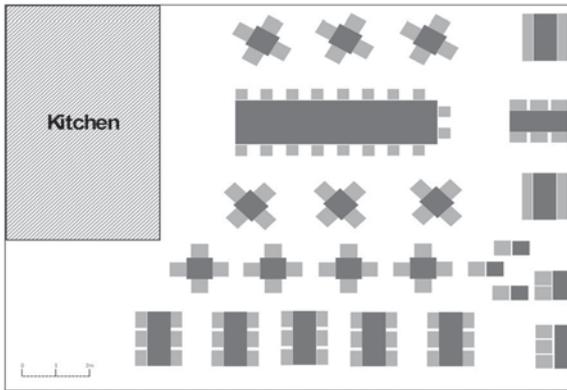


Fig.1: Spatial Configuration of the Cafeteria

1-4-5

1-4-5 大学講義室等における音環境が作業に
及ぼす影響に関する評価実験
A Subjective Evaluation of the Effects of the Acoustic Environment on
Learning Activities in University Classrooms

☆高曾根晴架, 米村美紀(前工大)

- ◆大学教育施設の講義室では講義の聴講だけでなく、個人作業やグループワークなどの授業形態がとられることもあり、音環境が及ぼす影響は活動によって異なる可能性がある。
- ◆本研究では、講義中に想定される様々な活動に対して音環境が及ぼす影響を調べることを目的とし、作業のしやすさなどに関する主観評価実験を行った。
- ◆作業は4名1組で行い、個人作業として文章校正課題と音声聴取課題、複数人作業として2対2のディベートを行わせた。
- ◆作業の種類・提示騒音の種類によって反応が異なり、音声聴取課題が騒音の影響を最も受けやすいことが分かった。また、複数人作業における発話のしやすさには個人差があることが確認された。

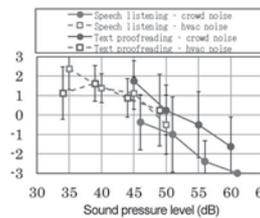


Fig. 1 : Ease of task performance (individual task)

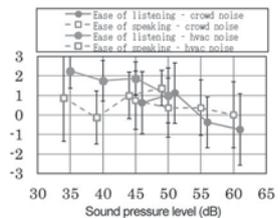


Fig. 2 : Results of the evaluation of group work

1-4-6

1-4-6 2次元理論モデルによるパッチ型吸音構造
の吸音特性に関する研究
Investigation on sound absorption characteristics of patch-type sound
absorbers using a two-dimensional theoretical model

○眞田 明, 藤本 望夢(岡山工技セ),
△萩原 淳之(三菱工業)

- ◆著者らは、これまでに多孔質吸音材料の表面に非通気のパッチを配置することで、低周波数から比較的広い帯域で吸音効果が得られるパッチ型吸音構造を提案している。
- ◆本研究では、モード展開を用いた2次元理論モデル (Fig.1) により垂直入射吸音率を求める式を導出し、吸音特性および吸音原理を明らかにした。
- ◆理論モデルによる吸音率の計算結果と FEM による計算結果が一致することを確認した (Fig.2 : $a=0.1$ m とし開口率を変えた場合)。
- ◆低周波数側のピークは、パッチ下空間の共鳴、高周波数側のピークはパッチ配置の周期性に起因する共鳴 (材料中においてユニットセル幅 L) に1波長入る共鳴) と考えられる。

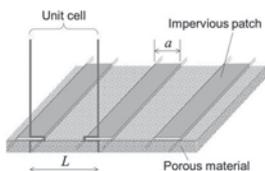


Fig.1: 2-D patch-type sound absorber.

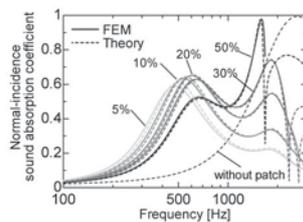


Fig.2: Calculated normal-incidence sound absorption coefficient.

1-4-7

1-4-7 周期分散配置した多孔質材料の
吸音・散乱特性の数値解析
Numerical Analysis of Sound Absorption and Scattering Characteristics of
Periodically Distributed Porous Materials

☆福島悠貴, 井上尚久(九大・芸工)

- ◆室内の残響調整において、吸音材の配置は分散配置により面積効果を利用することが効率的であることが度々言及されるが、その効果が残響室で観測されるものと同様・同程度に室内で観測されるかは、波動論的に自明でない。
- ◆多孔質材料の分散配置問題に対し、Flush model, Ribbed model (Fig.1) の二種を設定し、無限周期構造に対する入射・反射場を仮定して解析を行った。
- ◆面積効果は低・中音域よりも高音域で顕著である。
- ◆ランダム入射吸音率の周波数特性の挙動は、散乱限界周波数 f_{sc} (これは周期のみにより決定する) 前後で大きく変わる。

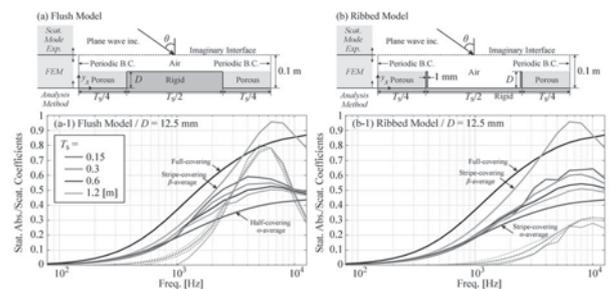


Fig.1: 1/3 octave band statistical characteristics calculated for different models and material thicknesses. Solid and dotted lines are absorption and scattering coefficients, respectively.

1-4-8

1-4-8 スリット構造を内包した多孔質材による低域吸音性能の向上についての検討

Low-frequency sound absorption improvement in porous materials with embedded slit structures.

◎川瀬遠我, 小柳慎一郎 (竹中技研)

- ◆低周波数域における吸音性の向上を目的として、スリット構造を内包した多孔質複合構造を対象とし、等価流体モデルに基づく数値解析により、共鳴周波数の設計手法と吸音特性の関係を検討した。
- ◆従来の共鳴周波数設計式を踏まえ、多孔質材内部の等価音速(あるいは等価波数)を共鳴周波数の設計式へ導入することで、多孔質母材中に形成されるスリット共鳴の設計可能性を評価した。
- ◆多孔質層内にスリットを埋め込むという単純な構成においても、多孔質材の等価音速を設計式に組み込むことにより、低周波数域の共鳴周波数を目標値に合わせて設計し得ることが示された。
- ◆一方で、吸音ピークの高さや帯域幅に本質的な限界が存在し、実用的な高吸音化・広帯域化を実現する必要がある。

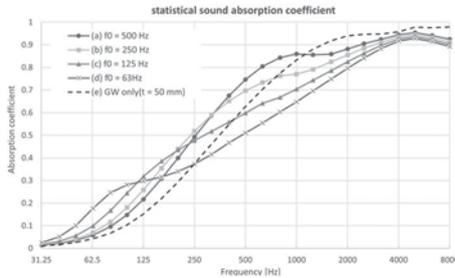


Fig.1: Statistical sound absorption coefficient at each design frequency. (●: f0=500Hz, ■: f0=250Hz, ▲: f0=125Hz, ×: f0=63Hz, ---: GW only)

1-4-10

1-4-10 建築設計図書を用いた吸音検討のケーススタディ

一吸音設計の基礎資料として その3ー Case Study on sound absorption evaluation using architectural drawings. - Fundamentals of Absorption Design, Part 3.

◎西川嘉雄(長野高専)

- ◆長野高専の建築設計図書の仕上げ表と図面から諸室の吸音力を求め、吸音 AIJES の室区分 I の推奨値との対応を確認した。
- ◆図-1a),b)に設計図書の条件での計算である。図-1c),d)は天井の吸音率 $\alpha=0.3, 0.5$ で計算した結果である。一般的な岩綿吸音板 $\alpha=0.5$ 程度で、おおむね I-3 は達成できる。さらに吸音を加えることで、I-2 を達成できる。吸音材設置していない空間は、騒音の喧騒感や近距離の音声伝送性能の確保は難しい。建築設計者は室の使い方を十分検討し仕上げを選択することが大切である。

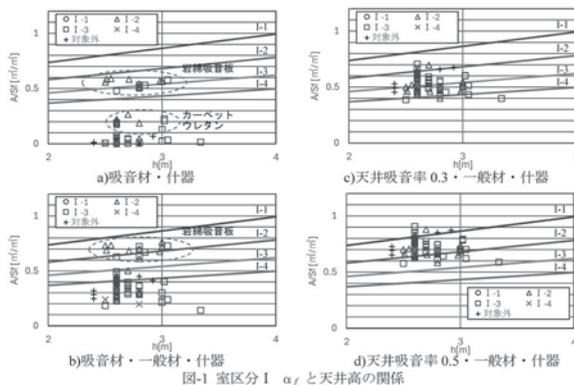


図-1 室区分 I α_f と天井高の関係

1-4-9

1-4-9 アンサンブル平均を用いた吸音特性の現場測定法 —PP 法に関する整理—

An in-situ measurement method using ensemble averaging technique for sound absorption of materials - A further discussion on PP-method -

○大鶴 徹, 富来礼次, 岡本則子 (大分大・理工)

- ◆アンサンブル平均による材の吸音特性の *in-situ* 測定法(EA 法)の測定原理について整理を行った。
- ◆大分大学残響室において、EApp 法、ならびに、EApu 法の 100 Hz 以下の低周波数域への適用可能性に着目した測定を試行した。
- ◆除湿器稼働のもとで実施した今回の実験では、サブウーハーの効果は顕著ではなかった。(Fig.1)
- ◆50 Hz ~ 1250 Hz の周波数域で、EApp 法は EApu 法に比し、安定した吸音率を与えている。
- ◆125 Hz 以下の周波数域で、EApu 法で得られるインピーダンス値は実部・虚部ともに変動が大きい。

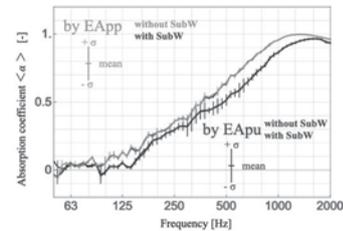


Fig.1: Absorption coefficient (α) values of glass-wool panel measured by both EApp and EApu methods. The effect of using a sub-wwoofer is plotted together.

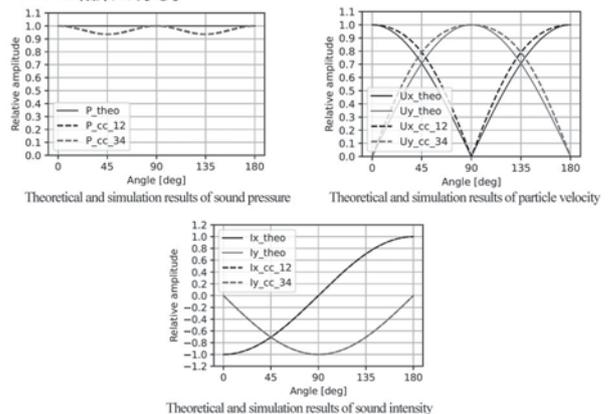
1-4-11

1-4-11 c-c 法に基づくインテンシティプローブの単一平面波音場における精度の検討

Investigation of the accuracy of an intensity probe based on the c-c method using cardioid microphones in a plane-wave sound field

○羽入 敏樹, 星 和磨, 鈴木 諒一(日大・短大)

- ◆筆者らは、カーディオイドマイクロホンを用いて音圧、粒子速度や音響インテンシティを測定する c-c 法を提案している。対向する一対のカーディオイドマイクロホンは c-c 法の最も基本的なプローブ構成である。c-c 法ではマイク間隔は誤差要因であるため、本報ではマイク間隔による誤差について単一平面波音場を想定して検討した。
- ◆その結果、プローブ軸方向の音圧と粒子速度に関しては振幅誤差が生じず、位相誤差のみ生じることがわかった。インテンシティについては角度に依らず誤差が生じない。以下はマイク間隔 10 mm、周波数 8 kHz の結果である。



1-4-12

1-4-12 c-c 法による音圧と粒子速度の絶対値測定 - 音響管内における一次元音場 -

○星 和磨, 羽入 敏樹 (日大・短大)

筆者らは、カーディオイドマイクロホンを音響管内の一次元音場に挿入し、音圧と粒子速度の絶対値測定を試みた。その結果、音圧、粒子速度、ともに既存センサと同等の周波数特性が得られることを確認した。

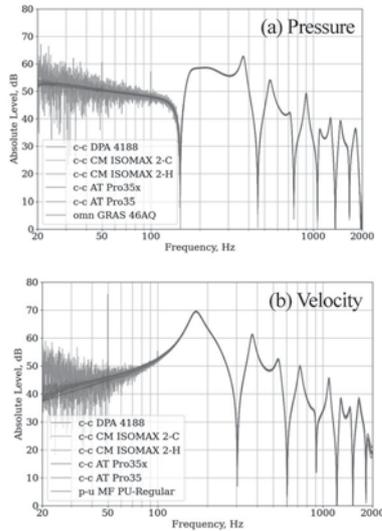


Fig. 1 Pressure and sound velocity measurement results under rigid terminal condition in an impedance tube.

1-4-14

1-4-14 等価曲げ剛性を反映した積層薄板の 偏微分方程式の弱定式化

The weak formulation of a partial differential equation of a laminated thin plate modelled with the equivalent flexural rigidity

○井上尚久(九大・芸工)

- ◆積層板を均質な単板としてモデル化するため、波数依存等価曲げ剛性を微分作用素として表現し、人工的な偏微分方程式を導出してきた。
- ◆積層板 (実際には梁) の微分方程式に対して重み付き残差法を適用することにより弱形式方程式を導出し、導入されたことになる物理量・物理的仮定を考察した。
 - エネルギー量
 - 一般化変位・一般化力とその釣り合い
 - 境界条件 (完全反射境界8種 (Fig.1) および無反射境界)

- ◆さらに、導出した弱形式方程式を数値的に解くための有限要素を定式化した。
- ◆導出した内挿関数により、概ね妥当に積層板の音響透過損失解析が行えるが、要素分割に対する収束性については課題も認められた。

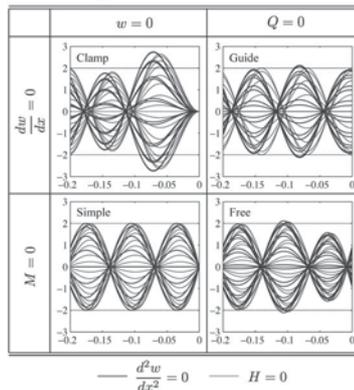


Fig. 1 Perfectly reflective boundary conditions for the laminated plate supported at $x=0$.

1-4-13

定在波の定式化における三角関数の合成 公式の使用に関する一考察

A consideration on the usage of trigonometric addition formulas in the formulation of standing wave

○佐野久, 柏崎絃(小野測器)

平面波において、入射波と反射波によって形成される定在波を定式化するとき、三角関数の合成公式を使用する。多くの文献では、合成後の位相角 ϕ を正接で表している。しかし、逆正接の値域は $-90^\circ < \phi < 90^\circ$ に限られており、 $-180^\circ < \phi \leq 180^\circ$ の範囲ではない。

このため、正接のみを用いると第2、第3象限の ϕ を正しく求めることができず、定在波の正しい時間波形を得ることができない (Fig. 1)。

この問題を解決するため、三角関数の合成公式どおり、 ϕ は正弦と余弦で表し、 $-180^\circ < \phi \leq 180^\circ$ と記述することを提案する。この方法で ϕ を求めると、正しい時間波形を得ることができる (Fig. 2)。

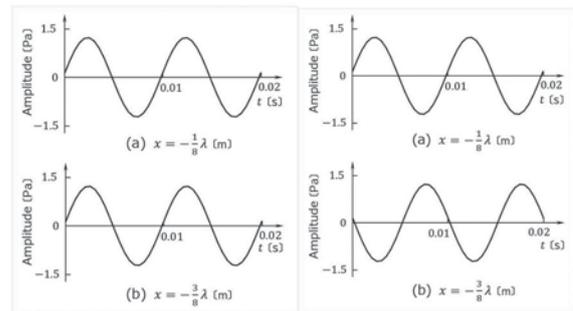


Fig. 1 Standing wave calculated by arctangent

Fig. 2 Standing wave calculated by trigonometric addition formula

1-5-1

1-5-1 ReDialTTS-JP: 音声復元した In-the-Wild 方言音声を用いた日本語多方言音声合成

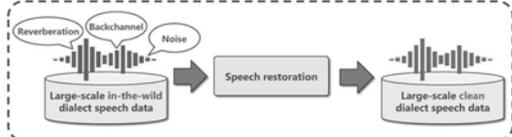
ReDialTTS-JP: Japanese Multi-Dialect Text-to-Speech Synthesis Using Restored In-the-Wild Dialect Speech

◎山内一輝, 中田亘(東大院・情報理工)

齋藤佑樹(東大院・情報理工/産総研), 猿渡洋(東大院・情報理工)

- ◆音声復元を用いた In-the-Wild 方言音声を高品質化し、これを学習データとして活用することで、高性能な日本語多方言 TTS モデル「ReDialTTS-JP」を構築
- ◆入力テキストの形式として、①カタカナ文、②標準語逐語訳、③漢字かな混じり文 (ASR 出力文) の3通りを採用した TTS モデルを比較評価し、日本語方言 TTS における最適なテキスト形式を検討

Stage 1: Cleaning dialect speech by speech restoration



Stage 2: Training an LLM-based multi-dialect TTS model

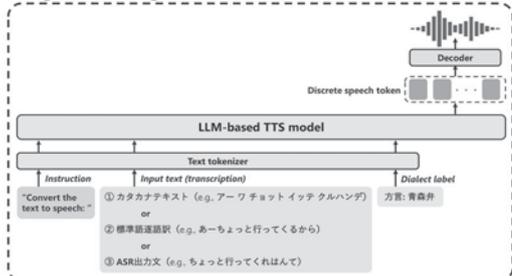


Fig. 1: Two-stage construction pipeline of ReDialTTS-JP.

1-5-2

1-5-2 離散トークンベースの生成的音声強調の信頼度スコアに基づく出力フィルタリング

Confidence-based Output Filtering for Discrete Token-Based Generative Speech Enhancement

◎山内一輝(サイバーエージェント/東大院・情報理工), 村田雅人, 関翔悟(サイバーエージェント)

- ◆ 生成的音声強調 (Generative Speech Enhancement; GSE) は、強調時に音素の置換・削除や、話者の不整合などの「ハルシネーション」を引き起こす

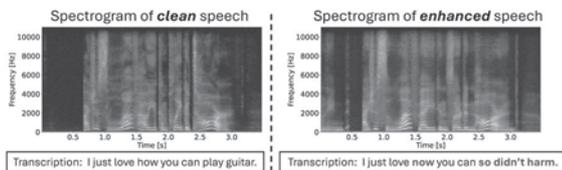


Fig.1: An example of a “hallucination error” caused by GSE.

- ◆ 本研究では、離散トークンベースの GSE モデルを対象に、ハルシネーションを含む出力音声を検知しフィルタリングする手法を提案する
- ◆ 提案手法は離散トークンベースの GSE が算出する「信頼度スコア」を活用し、強調失敗の可能性がある強調音声を除外するための指標として用いる
- ◆ 提案手法の実用性を示すため、In-the-Wild TTS データセットを信頼度スコアに基づくフィルタリングを用いてキュレーションする (Fig.2) ことで、学習される TTS モデルの性能が向上することを検証する

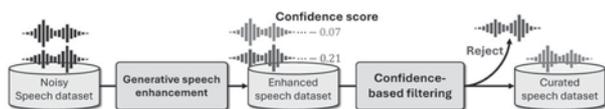


Fig.2: Overview of the speech dataset curation process with our proposed confidence-based filtering.

1-5-4

1-5-4 深層隠れセミマルコフモデルに基づいた End-to-End テキスト音声合成

End-to-End Text-to-Speech Synthesis using deep hidden semi-Markov models

☆浅野秀輝, 橋本佳, 南角吉彦, 徳田恵一(名工大)

- ◆ 研究背景
 - 深層隠れセミマルコフモデル (深層 HSMM) は高品質な音響特徴量の予測が可能
 - 直接波形を生成する End-to-End モデルへの適用は未検討
- ◆ 提案手法: 深層 HSMM と HiFi-GAN を統合した End-to-End 音声合成
 - Parallel Model: 深層 HSMM の潜在変数をメルスペクトログラム生成デコーダと HiFi-GAN へ入力
 - Sequential Model: 深層 HSMM が生成したメルスペクトログラムを HiFi-GAN へ入力
- ◆ 5段階 MOS による主観評価実験 (XIMERA コーパス)
 - 従来モデル (VITS) に対する優位性を確認
 - Parallel Model が高い MOS を達成し、提案手法の有効性を実証

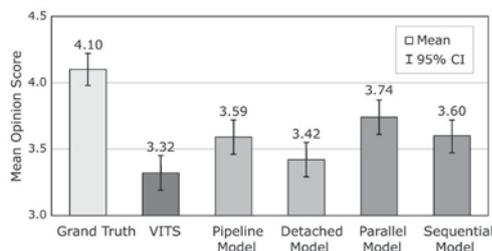


Fig.1: 5-point MOS results for synthesized speech naturalness.

1-5-3

1-5-3 LLM-based 音声合成における連続性を考慮したビームサーチ

Beam Search for LLM-based Text-to-Speech Incorporating Continuous Tokens

◎上乃聖, 李晃伸(名工大)

- ◆ 近年、LLM を活用した音声合成が質の高い音声合成の手法として注目されている。
- ◆ LLM を用いた予測には、Audio Codec を用いて音声を離散トークンとして、通常のテキストと同様に予測を行う。
- ◆ しかし、音声のトークンは時間解像度が細かく、系列長が長くなり、また、同 ID の繰り返しの出現も考えられ、先行研究でも同 ID の繰り返しを抑制した手法が提案されている。
- ◆ そこで本研究では、LLM-based 音声合成のビームサーチにおいてプロンプト音声内の同 ID の繰り返しに対して、スコアにボーナスを追加する。
- ◆ また、時間変化において変化が少ない、沈黙に着目して、沈黙で多く使われている ID 以外のものに対して、過去の候補における同 ID の繰り返しとの積算でペナルティを追加する。
- ◆ 実験結果により、従来の手法と比較して明瞭性が向上し、生成音声の長さに関して先行研究のビームサーチと比較して、長い音声生成できたことを確認した。

Table 1: Word error rate (WER), UTMOS, average duration for generated speech. WER is measured by Whisper large-v3-turbo.

Sampling method	WER (%)	UTMOS	Duration (s)
Repetition aware sampling	2.79	3.57	7.10
Temporal repetition diverse beam search	2.67	3.54	6.79
Proposed beam search	2.22	3.58	7.04

1-5-5

1-5-5 共役事後分布に基づいた深層隠れセミマルコフモデルに基づく音声合成

Deep hidden semi-Markov model-based speech synthesis using conjugate posterior distributions

☆石井信吾, 橋本佳, 南角吉彦, 徳田恵一(名工大)

- ◆ 深層隠れセミマルコフモデル (DHSMM)
 - Seq2Seq モデルに HSMM 構造を組み込んだ深層生成モデル
- ◆ 提案手法
 - 継続長分布に共役性を適用し、VAE のポトルネック表現として HSMM を用いた学習法を提案
 - 従来法よりも推論過程の理論的な見通しや解釈性が向上
- ◆ 結果
 - 既存の DHSMM と同等の品質を持つことを示した

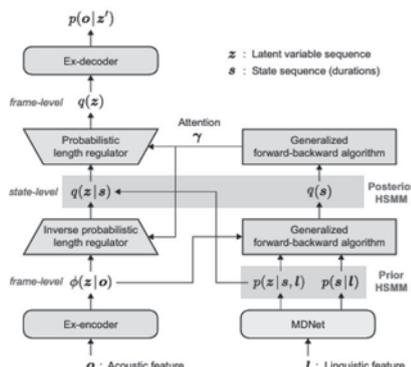


Fig.1: Structure of the proposed model

1-5-6

1-5-6 合成データのみを用いた ASR ドメイン適応のための LLM テキスト拡張と発音リスペリング

Synthetic data domain adaptation for asr via llm-based text and phonetic respelling augmentation

©山下夏生, 永塚光一, 小窪浩明, 土肥宏太, Ho Tuan Vu(日立製作所), 根本藍, 小市良祐(日立アドバンスシステムズ)

- ◆本研究では、合成音声のみを用いた ASR ドメイン適応のために、LLM によるテキスト拡張と発音リスペリング拡張 (PRA) を組み合わせた手法を提案する。
- ◆語彙多様性、パープレキシティ、およびドメイン特有語彙被覆を同時に最適化するフィルタリング戦略と、テキスト段階で発音揺れを注入する PRA により、実運用に近い合成データ生成を実現し、4つのドメイン特化データセットに対する認識精度を大幅に向上した。

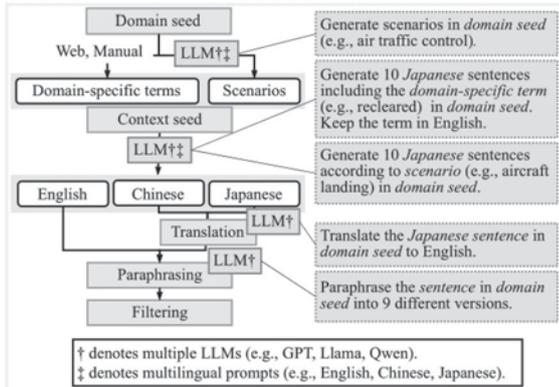


Fig.1: Proposed Text Augmentation Pipeline

1-5-8

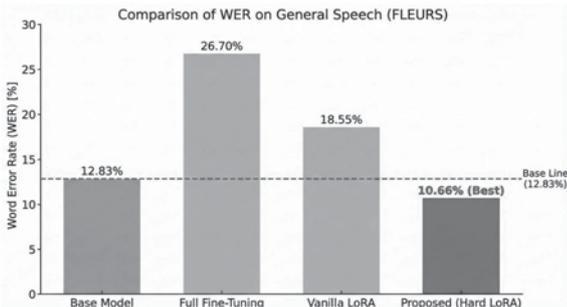
1-5-8 発話内容ドメイン教師あり LoRA による構音障害音声認識のドメイン適応

Domain Adaptation of Dysarthric Speech Recognition Using Utterance-Content-Domain-Supervised LoRA

☆小笠原朝陽(岩手大)* 高道慎之介(慶大/東大)*

末永剛(クリエイター), 談宜育(岩手大)

- ◆単一話者日本語構音障害音声コーパス (SS-JDSC: 15h) を活用し、発話ドメインごとの音響的変異が音声認識モデルに与える影響の分析を可能に。(basic,hard,dailyの3ドメインサブセット)
- ◆構音障害音声を音響的特徴でドメイン分離し、ドメイン別 LoRA (expert) により特定の変異を独立して学習させる手法により分析。言語尤度への過度な依存を抑制する。
- ◆歪みの激しい hard ドメインの学習が、一般音声 (Google FLEURS) の WER を 12.83% から 10.66% へ改善 (相対エラー率約 17% 減少) させる、クロスドメインでの正の転移を発見。



(表1: 各LoRAの計測結果と、whisper-large_v3のWERをグラフで示す。)

1-5-7

1-5-7 複数音声認識モデルの統合による未知ドメインへの頑健な適応手法

Robust adaptation to unknown domains via integration of multiple ASR models

☆北條圭悟, 若林佑幸(豊橋技科大), 太田健吾(阿南高専), 小川厚徳(NTT), 北岡教英(豊橋技科大)

- ◆複数の音声認識モデルから得られる音響情報とターゲットドメイン言語情報を統合する音声認識システムを提案
- ◆DNN-HMM 音声認識 : 複数システムの統合で認識精度を改善
- ◆End-to-end 音声認識 : 言語モデルの併用が精度向上に効果的
 - Density ratio approach (DRA) : 音声認識モデル内部の言語情報を除去することで、ターゲットドメインへの適応精度が向上
- ◆複数システムの統合と DRA を応用した音声認識システムを構築
 1. DRA : 言語情報の除去による音響情報の抽出
 2. システム統合 : 抽出した複数ドメインの音響情報を統合
- ◆提案手法は従来のシステム・言語モデル統合手法の精度を上回る
 - 追加の学習無しで未知ドメインに頑健な認識が可能

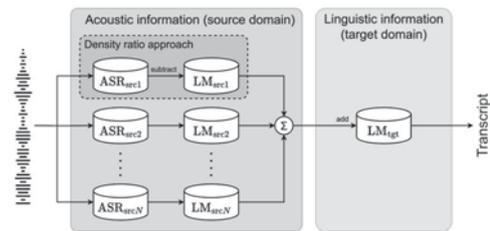


Fig.1: Overview of proposed method.

1-5-9

1-5-9 ILM-CTC: 内部言語モデルを明示的に学習可能な CTC 音声認識

Internal Language Modeling for CTC based Speech Recognition

©高城 巽成, 若林 佑幸 (豊橋技科大), 小川 厚徳 (NTT), 北岡 教英 (豊橋技科大)

- CTC ベースの ASR も学習データの言語情報 (ILM) を学習している [1].
- 構造的に明示的な ILM をモデリングするのが困難だった。

ILM-CTC:

- ILM を明示的に学習可能に拡張
- 推論時は任意で LM ブランチを停止して ILM を除去
- 従来の CTC よりも高性能なドメイン適応性能

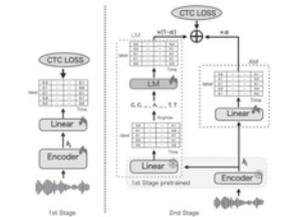


Fig. 1 Overview of the ILM-CTC architec-

Table 1 Domain adaptation performance when the source domain is LibriSpeech, and the target domain is GigaSpeech. (WER %)

Model	Branch	SubLM	AddLM	Subsets of GigaSpeech set M					
				audiobook	P-news	Y-people	Y-news	Y-science	Y-edu
Conformer-CTC	-	N/A	N/A	3.69	11.6	15.4	16.4	16.3	14.6
		N/A	4-gram	3.53	10.7	14.4	15.4	14.7	13.6
		1-gram	4-gram	3.36	10.0	13.8	14.8	14.1	13.1
ILM-CTC	AM+LM	N/A	N/A	3.69	11.6	15.4	16.4	16.3	14.6
		N/A	4-gram	3.43	10.4	14.1	15.0	14.3	13.4
		1-gram	4-gram	3.36	10.1	13.9	14.9	14.2	13.3
	AM only	N/A	N/A	3.65	11.6	15.4	16.4	16.2	14.6
		N/A	4-gram	3.42	10.4	14.0	14.9	14.2	13.3
		1-gram	4-gram	3.35	10.0	13.6	14.5	13.7	13.0

[1] N. Das et al., "Mask The Bias: Improving Domain-Adaptive Generalization of CTC-based ASR with Internal Language Model Estimation", ICASSP, 2023.

1-5-10

1-5-10 クープマンモード分解による音声分析の一検討

A study of speech analysis based on Koopman mode decomposition

○舟木慶一(琉球大学)

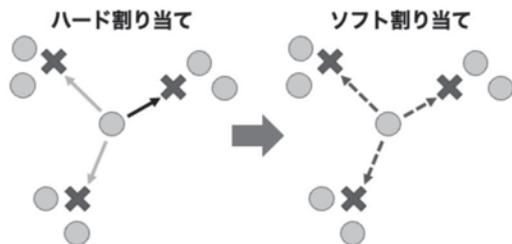
Koopman 作用素は、非線形力学系を線形作用素の枠組みで解析する強力な手段である。1931年にBernard Koopmanによって導入されたこの理論 [1]は、近年、データ駆動型のシステム解析やモデル予測において再び注目を集めている [2]。本稿では Koopman 作用素の離散スペクトルに基づく時系列データ解析手法である Koopman モード分解 (KMD) を用いた音声分析法の検討を行う。具体的には KMD を動的モード展開 (Dynamic Mode Decomposition: DMD)[3][4] または拡張動的モード展開 (Extended Dynamic Mode Decomposition: EDMD)[5] で再構成し、再構成された信号に対し、LPC 分析ならびに TV-CAR 分析を行う。

1-5-12

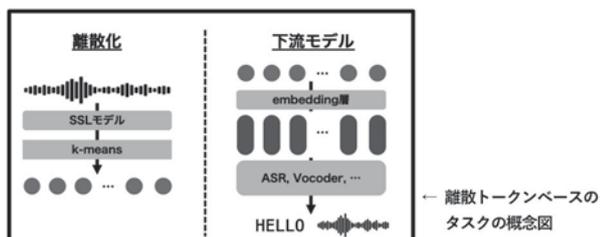
1-5-12 下流タスク推論時における SSL モデル由来の音声離散トークンの確率分布的利用

Probabilistic Use of Discrete Speech Tokens Derived from SSL Models for Downstream Task Inference

◎ 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)



- 離散トークンベースのタスクの推論時に、トークンのソフト割り当てを適用
- 各セントロイドへの距離に基づき、embedding の重み付き和を計算
- 学習時には従来通りのハード割り当てを維持 = 学習時の効率性は維持
- ASR, 再合成タスクで性能向上
- 話者情報は捨象しつつ、言語情報、韻律情報の正確性向上



1-5-11

1-5-11 時間領域と周波数領域のスパース性を自己調整する LPC

Linear predictive coding with adaptive adjustment of sparsity between time and frequency domains

◎ 鳥居美月, 矢田部浩平 (農工大)

〈背景〉ピッチの高い音声に対応するため、時間と周波数の各領域におけるスパース成分を個別に扱いたい

〈提案〉inifmal convolution を用いて各領域でのスパース成分の和として予測残差をモデル化

▶ 領域ごとに対応する成分を個別に評価

〈結果〉提案手法 (Prop) は従来手法 (Conv) より、分析結果が時間的に安定した (Fig. 1 枠内)

▶ フォルマント周波数をより正確に推定できる可能性が示唆された

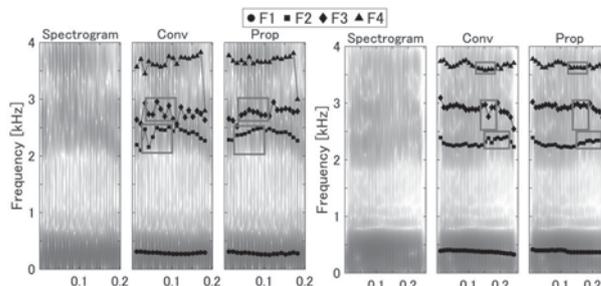


Fig. 1: Spectrograms of speech and estimated formant frequencies for male (left) and female (right) speakers

1-5-13

1-5-13 Leveraging wav2vec 2.0 and MFCC for Articulation rtMRI Landmark Generation in Visual Pronunciation Learning

☆ Mushaffa Rasyid Ridha, Faisal Mehmood, Sakriani Sakti

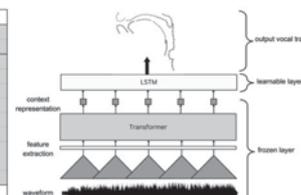
◆ This paper addresses the limitations of current Computer-Assisted Pronunciation Training (CAPT) systems, which typically provide proficiency scores without actionable guidance on how to physically produce sounds (vocal tract trajectories)

◆ The video demonstrations are available at <https://zenodo.org/records/13756435>



◆ This study demonstrates that fine-tuned wav2vec 2.0 embeddings significantly outperform MFCC and phonemes in reconstructing vocal tract trajectories from speech (from rtMRI). By capturing complex articulatory dynamics, this approach provides a robust foundation for real-time visual feedback.

Articulatory	Pho	MFCC	w2v	
Lip Protrusion	Up	6.85%	72.54%	77.02%
	Down	47.72%	43.22%	64.91%
Aperture	L	43.89%	46.71%	67.75%
	R	48.27%	43.72%	65.92%
TAD		47.22%	44.48%	68.75%
TD		47.01%	41.78%	61.81%
TPD		55.09%	54.13%	64.05%
TU		54.45%	48.92%	70.43%
UP		57.89%	66.25%	84.61%
TP		63.72%	56.46%	68.46%
E		37.57%	35.04%	40.06%
G		50.08%	50.53%	37.57%



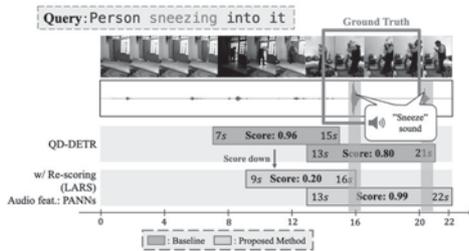
1-5-14

1-5-14 動画内区間検索における関連度値校正のための音響・映像特徴量統合

Audio-Visual Feature Integration for Calibrating Relevance Scores of Video Moment Retrieval

☆今村剛大(名古屋大, LINE ヤフー), 小松達也(LINE ヤフー), 宗像北斗(LINE ヤフー), 戸田智基(名古屋大)

- ◆動画内区間検索 (VMR) とは
 - 検索文と関連する動画区間を出力するタスク
- ◆従来手法: QD-DETR
 - 物体検出モデル DETR (物体位置特定 & 物体カテゴリ識別) を VMR (区間位置特定 & 最適区間識別=検索文関連度値推定) に応用
 - DETR に内在する関連度値推定性能の低さが課題
- ◆本研究では, Late-Fusion Re-scoring Module (LARS)を提案
 - 音響特徴量を DETR 以後の関連度値校正に応用
 - DETR に内在する課題を回避して関連度値を再検討
- ◆実験的評価により, 3つのデータセットで有効性を確認



Example of search result improvement with LARS

1-5-16

1-5-16 Neural Audio Codec を用いた自己教師あり学習モデルにおける事前学習データの言語が下流タスクに与える影響の検証

Investigation of the Impact of Pre-training Data Language on Downstream Tasks in Self-Supervised Learning Models Using Neural Audio Codecs

○瀧澤 大吾¹, 中村 友彦¹, Samuele Cornell², William Chen², 深山 覚¹, 渡部 晋治²(¹産総研, ²CMU)

- ◆NAC HuBERT の事前学習に用いる言語と NAC の学習言語が下流タスクに与える影響を, 英語・日本語・中国語の音声を用いて検証した
- ◆言語的タスクである音声認識と, パラ言語的タスクである音声感情認識を用いることで事前学習言語の影響を異なる側面から評価した
- ◆下流タスクと同じ言語で事前学習した NAC HuBERT を用いた場合に, すべての評価言語において最も高い性能を示した
- ◆一方, NAC の学習に用いる言語の違いによる性能差は小さく, 言語が性能に与える影響は限定的であった

Table 1 事前学習言語と音声認識評価結果

評価言語	NAC HuBERT 事前学習言語			NAC 学習言語		
	英語	日本語	中国語	英語	日本語	中国語
英語 ¹	10.4	21.5	24.4	10.4	14.5	14.5
日本語 ²	30.5	14.5	20.9	10.7	14.6	14.4
中国語 ³	31.0	19.5	14.5	10.4	14.5	14.4

本要旨では, 言語依存性が最も顕著に表れる音声認識の一部結果を代表例として示す。音声感情認識の結果については本文で詳述する
¹ Libri-Light (test_clean) ² LaboroTVSpeech ³ WenetSpeech

1-5-15

1-5-15 Neural Audio Codec を用いた低コストな自己教師あり学習モデルの検討

A Study of Low-Cost Self-Supervised Learning Models Using Neural Audio Codecs

○瀧澤 大吾¹, 中村 友彦¹, Samuele Cornell², William Chen², 深山 覚¹, 渡部 晋治²(¹産総研, ²CMU)

- ◆大規模な自己教師あり学習 (SSL) モデルの事前学習には, **大容量のストレージ**およびデータ処理に要する**膨大な時間**が必要である
- ◆本研究では, 音声を効率的に圧縮する手法である Neural Audio Codec (NAC) を用い, **ストレージと計算コストを低減**した SSL モデルの構築法を提案する
- ◆事前学習において, **学習データサイズを 90% 以上削減**するとともに, **学習時間および GPU 使用量を低減**した
- ◆音声認識・音声感情認識タスクにおいて, 音声波形を入力とする **HuBERT と同程度の性能**を示した
- ◆一方, NAC による音声圧縮に起因すると考えられる性能の**わずかな低下も確認**された

Table 1 NAC HuBERT の事前学習コスト削減と下流タスク性能

モデル	事前学習			下流タスク性能	
	データサイズ	学習時間	GPU 使用量	ASR ¹ (WER %)	SER ² (ACC %)
HuBERT	100%	100%	100%	10.7 / 18.6	65.61
NAC HuBERT	10%	81%	87%	10.3 / 20.8	64.57

¹ Libri-Light (test_clean / test_other) ² IEMOCAP

1-5-17

1-5-17 話者・言語情報の構造的分離を可能にするニューラル音声コーデックの検討

Exploring Disentangled Neural Speech Codecs from Self-Supervised Representations

©相原 龍(三菱電機), 升山 義紀

ウィシャーン ゴードン, ジェルマン フランソワ, ルルー ジョナトン(MERL)

- ◆ニューラル音声コーデックとは:
 - Neural Nets を用いて, 音声のコンパクトな表現を推定
 - 推定した表現から音声へ復元できる。
- ◆離散コーデックは生成モデルで扱いやすいので嬉しい。
- ◆音声には音素, 韻律, 話者の 3 つの情報がたとえられる: これらの情報が**構造的に分離**された離散トークンを検討した。

手法:

- ◆自己教師ありモデル WavLM の中間表現を利用
 - 話者や韻律と強い相関がある層が存在
- ◆中間表現の時間方向の平均と分散を話者情報とみなせる
- ◆中間表現と話者表現の残差を韻律情報とみなせる

結果:

- ◆非分離手法とほぼ同等の分解再合成精度を得た
- ◆構造的分解能力を One-shot 声質変換で示した

知見:

- ◆話者表現の離散化は品質劣化に直結。十分な情報量が必要。

1-5-18

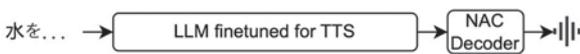
1-5-18 大規模言語モデルに基づく日本語音声合成の品質及び安定性評価 Quality and Stability Evaluation of Japanese Speech Synthesis based on Large Language Models

◎中田 亘, 齋藤 佑樹(東大院・情報理工/産総研), 猿渡 洋(東大院・情報理工)

- LLM で日本語音声合成しました
- G2P なしでも動きました
- 400 発話で特定話者に適応できました
- 100 発話だと合成が不安定になりました



(a) 従来の日本語 TTS. 図ではアクセントラベル付き日本語音素列を使用



(b) 本研究で使用する LLM に基づく日本語音声合成. G2P を使用せず, かな漢字混じり文から直接合成.

Fig. 1: 従来の日本語 TTS と, 本研究で採用する LLM に基づく日本語音声合成の比較.

1-6-2

1-6-2 軟骨伝導を用いた中耳手術での術中の聴力評価に関する研究

Intraoperative hearing assessment using cartilage conduction in middle ear surgery

○西村忠己(奈良医大), 細井裕司(奈良医大), △北原 紘(奈良医大)

- ◆中耳手術では術中の聴力評価が困難であり, 軟骨伝導が新たな評価法として有望である。
- ◆滅菌可能な軟骨伝導振動子を試作し, 過酸化水素ガスプラズマ滅菌で検証した。
- ◆15 回までの滅菌で大きな破損はほぼなく, 感度変化も小さかった。
- ◆試作振動子は術中モニタリングへの応用が期待でき, 今後は臨床での検証を進める。



Fig. 1 Prototype cartilage conduction vibrator

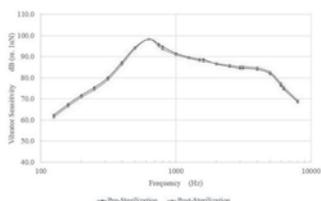


Fig. 2 Changes in vibrator sensitivity after sterilization

1-6-1

1-6-1 80 歳以上から始める補聴 個人用軟骨伝導聴覚補助イヤホンの新たな可能性

Hearing Compensation for the Over-80s: The New Potential of Personal Cartilage Conduction Hearing Support Earphones

○下野真理子(国立長寿医療研究センター), △内田育恵(愛知医大), 細井裕司(奈良医大)

- ◆難聴高齢者に対する適時適切な補聴は健康寿命の延伸に寄与すると考えられるが, 認知機能や手指の巧緻性の低下, 経済的負担により, 高齢期からの補聴器導入が敬遠される事例は少なくない。そこで我々は新たな選択肢として個人用軟骨伝導聴覚補助イヤホンに着目した。
- ◆気導補聴器と比較した本機器の特徴を Table 1 に示す。本機器は, 気導補聴器の適合や管理が困難な症例においても外耳道を開放した状態で装用可能であり, 耳の閉塞感や不快感が少なく受容されやすい。さらに, 音響性聴覚障害のリスクが低いとされる軟骨伝導経路を活用しており, 安全性も高い。
- ◆以上より, 個人用軟骨伝導聴覚補助イヤホンは補聴器導入に抵抗がある症例への初期導入手段, もしくは代替手段として活用が期待される。

Table 1: Comparison Between Personal Cartilage Conduction Hearing Support Earphones and Air Conduction Hearing Aids

	Personal Cartilage Conduction Hearing Support Earphones	Air Conduction Hearing Aids
Legal Classification	General audio equipment	Controlled medical device
Frequency-specific gain adjustment according to hearing levels	No	Yes
Individual Fitting	No	Yes
Ear Attachment Features	Hole-free solid vibrator that prevents earwax buildup and ensures easy cleaning	Eartips/Earmolds with holes for sound delivery
Pricing	Market price: Approximately ¥30,000 per unit (Binaural)	Average market price: Approximately ¥100,000 - ¥300,000 per unit (Single ear)

1-6-3

1-6-3 Transformer-OTGSM を用いた骨伝導音から音声への音質改善

A Transformer with Optimal Transport Guided Gain Smoothing for Bone-Conducted-to-Speech Enhancement

○中山 仁史(広島市大院)

- ◆骨伝導音から音声への明瞭化のために Transformer-OTGSM を提案した。
- ◆OT により高帯域不足や過補償に由来する帯域分布の乖離を補正し, GSM でフレーム間の不連続を抑える設計とした。
- ◆東北大松下単語データベース (男声 5 名×3 セット) を対象にモデルの構築を行い, 有声区間のみ・遅延補正後の LSD で評価した。
- ◆結果として Transformer+OT が最良で, 骨伝導音の LSD である 17.71 dB から 12.55 dB へと改善が可能なことを確認した。

Table. 1 : Improvement performance of each method

Method	LSD [dB]	ΔLSD [dB]
BCS	17.71	—
MMSE-LSA-SPP	22.86	-5.15
MedianMap (local)	16.89	0.82
MedianMap (global)	24.98	-7.28
MedianMap (multi-band)	24.86	-7.16
Multi-band + OT shaping	14.74	2.97
Multi-band + harmonic assist	24.83	-7.12
Transformer	24.83	-7.12
Transformer + OT	12.55	5.16
Transformer (alt.)	24.86	-7.16
Transformer + OT (alt.)	12.58	5.13

1-6-4

1-6-4 骨導・軟骨導音提示による気導音抑圧に最適な位相条件の検討

Preliminary study on the optimal phase condition of bone-conduction stimulation for suppression of air-conducted sound

鳥谷輝樹(山梨大院・総研), 三木歩(山梨大・工), 小澤賢司(山梨大院・総研)

- ◆背景: 気導・骨導純音相殺実験 (Stenfelt, 2007) に着想を得て, 骨導・軟骨導音提示による気導音抑圧法を検討中.
→ 純音を 8 dB 以上抑圧できたが, 複合音の抑圧に課題あり.
- ◆目的: 250~4000 Hz を対象に, 気導音抑圧に最適な骨導・軟骨導提示音の位相条件を検討し, 複合音の抑圧効果を確認すること.
- ◆方法: (1) 気導音 (250, 500, 1000, 2000, 4000 Hz 純音) と等価なラウドネスとなるよう振幅調整した骨導・軟骨導音 (二次音) を同時提示し, ラウドネスが最小となる二次音の位相条件を調整法で決定. (2) 得られた最適位相条件をもとに設計した二次音調整フィルタを適用 → 上記 5 周波数の純音・複合音の抑圧量を評価.
- ◆結果: (1) 骨導提示条件では特に 1000 Hz 以下で, 軟骨導提示条件では全周波数で, 参加者に依らず一貫した位相特性が得られた. (2) フィルタ適用時の抑圧量は最適位相条件下の抑圧量に対し -3 dB 以内 (Figure 1). 複合音では最大約 10 dB の抑圧効果あり.

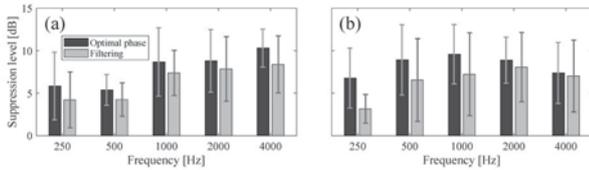


Figure 1: Mean suppression level of pure tones across participants using (a) bone-conduction and (b) cartilage-conduction transducers.

1-6-6

1-6-6 軟骨伝導能動騒音制御におけるヴァーチャルセンシングに関する基礎検討

A preliminary study on virtual sensing method for cartilage conduction active noise control

☆佐能春嘉, 笹岡直人(鳥取大・工), 下倉良太(大阪大・基礎工)

- ◆軟骨能動騒音制御(CC-ANC)の制御フィルタを構成するための非侵襲的な手段としてヴァーチャルセンシング手法を提案する.
- ◆既存の手法では, 制御フィルタを構成するために耳内にマイクを設置する必要があり, リスクが伴う. 提案手法では, 耳の外側で得られる外耳道内反射を含む経路情報から制御フィルタを推定する(Fig.1).
- ◆無響室にて人間被験者の1次経路, 2次経路, および反射を含んだ経路のインパルス応答を測定し, それらを用いてシミュレーションを行った.
- ◆提案手法の基本構成では騒音低減が確認されなかったが, 補償を行う適応フィルタを追加した場合には騒音低減が確認された(Fig.2). 今後ヴァーチャル化に向けた検討を行う.

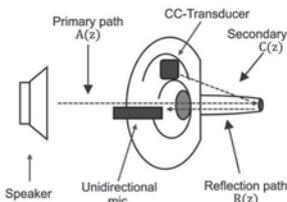


Fig.1: System overview.

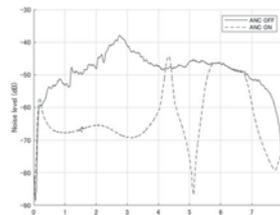


Fig.2: Noise level before and after control.

1-6-5

1-6-5 DSP と能動的騒音制御を用いた軟骨伝導音によるリアルタイム狭帯域雑音除去

Real-time active noise control targeting narrowband noises by using cartilage conduction and DSP

○下倉良太, 加藤良樹(大阪大), 笹岡直人(鳥取大)

- ◆軟骨伝導は耳を開放したまま音聴取できることが特徴である.
- ◆この特徴を生かし, 我々は軟骨伝導音を二次音源とする能動的騒音制御(ANC)に挑戦している. この技術により耳を開けたまま外部騒音を遮断することができる.
- ◆さらに騒音と音声と同時に到来する環境で, 音声だけを鼓膜に到達させる「選択的消音」にも取り組んでいる.
- ◆本研究は, 適応バンドパスフィルタ(ABF)で狭帯域雑音のみを抽出し, 音声と分離して雑音のみを Filtered-X Least Means Square (FxLMS) で除去するシステムを DSP に実装し, 実環境において消音が可能か実験を行った.
- ◆その結果, 提案法である ABF-FxLMS は音声を残したまま狭帯域雑音を 9 dB 低減させることに成功し, 選択的消音の可能性を示すことができた.

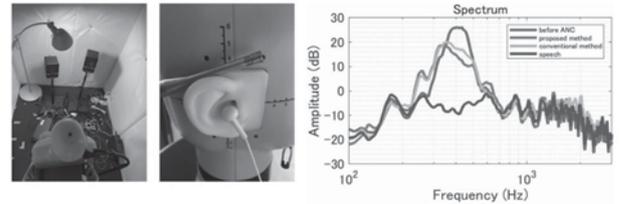


Fig. 1: (Left) Photos of experimental condition: the white ball is a cartilage conduction transducer, (Right) error signals for narrowband noise with center frequency 400 Hz.

1-6-7

1-6-7 (招待講演) 音声品質評価における分野間連携の意義と展開可能性

—非常用屋外拡声システムを事例として—

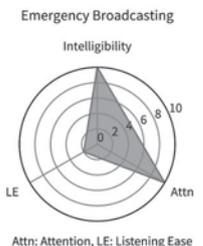
Significance and prospects of interdisciplinary collaboration in speech quality evaluation: Focused on emergency outdoor public address systems

○佐藤逸人(神戸大), 栗栖清浩(TOA), 小林洋介(室蘭工大)

- ◆音声品質の評価は分野横断型のテーマであり, 多種多様なアプリケーションごとに要求される音声品質も異なるが, 本稿では異なる分野間の連携を視野に入れ, 了解性, 注意性, 平易性の3軸(下表参照)による共通の音声品質評価の枠組みを提案した.
- ◆非常用屋外拡声システムは, 短時間での確実な情報伝達が必要とされる一方で, 正常性バイアスと外乱要因による妨害効果大きい. したがって, 上述の評価の枠組みでは, 平易性よりも了解性と注意性が大幅に重視されるアプリケーションと特徴付けられる(下図参照).
- ◆上述の特徴を踏まえ, 日本音響学会非常用屋外拡声システム調査研究委員会では, 了解性を高めるだけでなく, 緊迫感のようなパラ言語情報にも着目して注意性も高めることで「聞く気にさせる」放送の実現を目指して, 技術規準や音源作成ガイドラインの作成を行っている.

Factor	Value
Intelligibility	Accurate delivery of information
Attention	Inducing the intent to listen
Listening Ease	Stress-free sustained listening

Fig.1: Key factors of speech quality evaluation and their value (Above table) and weighting of the factors for emergency broadcasting (Right radar chart).



1-6-8

要旨

(招待講演) ビジネス・教育・メディアの立場から見た「命を救うコンテンツ」の過去・現在・未来

○三上昌也 (株式会社ユニシー/岩手大学)

日本は世界有数の災害多発国であり、防災行政無線やアラートに代表される拡声音は、災害時に住民へ危険を伝達する重要な社会インフラとして機能してきた。しかし、東日本大震災以降の各種調査では、警報情報を認知していたにもかかわらず避難行動に至らなかった事例が多数報告されている。背景には、異常事態に直面した際に現状を過小評価してしまう正常性バイアスの存在が指摘されており、従来の拡声音が必ずしも人間の行動変容を十分に促してこなかった可能性が示唆される。

拡声音の設計はこれまで、音圧や明瞭度といった工学的指標を中心に発展してきた。一方で、人間の認知や感情に働きかけ、具体的な行動を誘発することを主目的とした設計や評価は限定的であった。本講演では、筆者が携わってきた音声合成技術(TTS)開発の実務経験、映像・音響を扱うメディア制作の知見、ならびに大学における分野横断型教育(PBL)の実践を踏まえ、拡声音を「命を救うコンテンツ」として再定義する視点を提示する。

近年、深層学習に基づくTTS技術の進展により、話速や抑揚、声の切迫感を段階的に制御することが可能となり、音声を単なる読み上げ手段ではなく、人の行動に影響を与える設計対象として再検討できる条件が整いつつある。また、メディア分野では、音声表現や演出が感情や判断に影響を与える知見が蓄積されてきたが、防災分野への体系的応用は十分とは言えない。さらに、工学・心理・メディア・行政といった分野のサイロ化が、行動変容に結びつかない構造的要因となっている点にも着目する。

本講演では、拡声音の評価軸を「情報の正確性」から「行動変容の実効性」へと転換する必要性を論じ、心理・生理・行動を含めた多角的分析と分野横断的デザインの重要性を示す。拡声音を人間中心の「命を救うコンテンツ」として捉え直すことが、防災DXが進展する現在において喫緊の課題であることを提言する。

1-6-10

1-6-10 “聞こえる”だけじゃ足りない”

- 補聴器の音質評価の課題 -

Hearing Alone Isn't Enough

- Issues in Hearing Aid Sound Quality Evaluation -

○藤坂洋一 (RION)

◆本報告では、難聴者の主観的音声明瞭度評価と、広く用いられる客観的評価指標との関係について体系的なレビューを提供することを目的とする。これらの指標の基礎となる仮定や報告されている知覚相関を比較することで、それぞれの長所と限界を示す。最後に、補聴器信号処理における難聴者の知覚体験をより正確に反映する客観的評価指標のフレームワークの開発に向けた今後の方向性について議論したい。

◆今後の方向性として、以下のように考えている。多くの研究者の協力による開発が行われることを期待したい。

◆健聴者では判断できない難聴者個々における処理音の優劣を予測することが可能な客観指標を開発したい。フレームワークとしては、「HASQI/HASPIからの発展」を念頭に、被験者間分散や主観評価のばらつきを説明できない限界を補うため、多くの被験者属性や音響特徴量を入力とした深層学習を用いた個人適応型知覚モデルによって、主観評価を多次的に予測したい。これにより同一聴力レベルでも異なる主観評価を示す現象をモデル化し、補聴器処理の個別最適化や臨床評価への実用的応用を可能にできると考える。

1-6-9

(招待講演)

防災放送のためのパラ言語情報を含む音声の合成と評価技術の動向

Trends in synthesis and evaluation technologies of speech with paralinguistic information for disaster-prevention broadcasting

○能勢 隆 (東北大学/株式会社ユニシー)

◆防災放送は災害時に住民の生命を守る重要な情報インフラであり、発話内容の正確な伝達に加え、危険の切迫度を直感的に伝える音声表現が求められる。

◆従来研究では信号処理による可聴性改善が中心であったが、放送の自動化・安定化の観点では音声合成(Text-to-Speech, TTS)の利用も有効である。

◆一方で、最近のニューラル TTS は自然性や表現力が高い反面、言語情報の可聴性・理解性や実装制約が十分に考慮されない場合があり、防災用途では慎重な設計が必要である。

◆本稿では、防災放送に求められる音声の要件を整理した上で、緊迫感などのパラ言語情報を含む音声合成・スタイル制御技術や近年のTTSの動向を概観する。

◆また、合成音声の品質評価および緊急度・情動の度合を客観的に評価する技術として合成音声の自動品質評価手法、および自己教師あり学習に基づく感情推定手法などを概説するとともに、今後の展望について述べる。

1-6-11

1-6-11 工場における電気音響設備を用いた非常放送の音声明瞭度に関する検討

Study on the intelligibility of emergency announcements using electroacoustic equipment in factories

◎浅井拓朗, 富高隆, 浜田由記子, 田中ひかり(大成建設)

◆工場では生産・空調設備の稼働だけでなく、吸音材の使用に制約がある場合もある。そのため、暗騒音が大きく残響が長くなり、電気音響設備による明瞭性の確保が難しい。一方で非常放送の達成目標(評価基準)が整理されていない。

◆実際の工場を対象に音響測定を実施した。騒音レベル70~80 dB(暗騒音)、残響時間3.5秒、音声明瞭度STIは0.26~0.29程度であったが、放送内容は聴取可能だった。

◆工場の暗騒音・残響を模擬した試験音(STI 0.09~0.37)を用いて聴感実験(10名)を行った。

◆単語正答率・聴感印象はいずれもSTIと正の相関($r=0.65, 0.74$)を示し、聴取可能な最低限の目安としてSTI 0.26~0.29が示唆された。

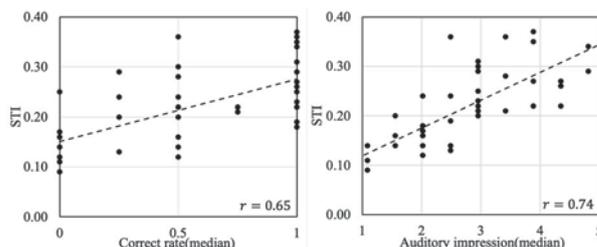


Fig.1: Relationship between STI and correct rate, auditory impression.

1-6-12

1-6-12 鉄道駅の音響設備設計における検討項目の明確化

Clarification of considerations for designing acoustical equipment for railway stations

○深美和樹(電気技術開発)

- ◆鉄道のプラットフォーム上に指向性強いラインアレイスピーカを設置したが、聴感印象がシミュレーションによる評価とは異なる結果になり、その原因を明確にするために実測を行い、対応策を報告した。
- ◆シミュレーション結果の内容を分析した。
- ◆現行のスピーカと新たに設置したラインアレイスピーカの拡声音の測定を行い、比較検討を行ったところ、ラインアレイスピーカからの拡声音が十分聞き取れない原因として、暗騒音と拡声音における周波数毎のSN比に起因していることが判明し、対応策を報告した。(Fig.1)
- ◆測定・比較検討で得られた知見を元に、鉄道駅の音響設備設計における検討項目を整理し、施工後、適切に拡声音がいき届いているかの検証の必要性を提案する。

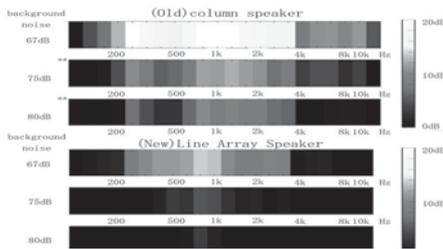


Fig.1: Signal-to-noise ratio at each frequency

1-7-1

1-7-1 楽器を使った「音空間デザイン」の提案と実践

—岐阜高専図書館における実践—*

Proposal and practices of "Sound Space Design" using musical instrumentals - Practice at GNCT library -

☆水口聖菜 (岐阜高専専攻科), 石川あゆみ(岐阜高専)

- ◆我々は、空間に音を加えることで新たな魅力を生み出そうとするデザインを「音空間デザイン」、それによって生み出された空間(既存の空間および音環境に音空間デザインで加えた音を合わせた場)を「音空間」と呼び、その実践を行ってきた。
- ◆先行研究では、音空間デザインでの使用を想定した楽器およびベース音に関する主観評価実験を実施した。
- ◆得られた知見をもとに、本研究では、岐阜高専図書館の利用者が参加できる楽器を使った音空間デザインの実践を行う。
- ◆実践の目的を、本校図書館の利用者を増やすきっかけを作ること、学生と市民のコミュニケーションを促すこととする。
- ◆多くの市民が訪れる本校の学校行事において、二回にわたって音空間デザインの実践を行った。



Fig. 6 Scenes from the Technical College

1-6-13

1-6-13 (招待講演)受聴側視点の音声明瞭性評価

Evaluation of speech clarity from the listener's perspective

◎山本克彦(サイバーエージェント)

- ◆緊急性の高い状況下での音声情報の伝達
 - 音声の明瞭性(了解度・明瞭度)を評価することが重要
 - 『受聴者の聴覚特性』も考慮する必要性
- ◆本発表: 受聴側視点に基づく音声分析や了解度予測 (Fig. 1 中の I_4)
 - 聴覚特性に基づく音声分析
 - ◇ 周波数領域: 聴覚フィルタバンク
 - ◇ 時間領域: 変調フィルタバンク
 - ◇ スペクトル時間変調領域: 周波数変調×時間変調
 - 音声了解度の予測モデル
 - ◇ 古典的な手法: SII, STI, STOI, sEPSM, HASPI
 - ◇ 深層学習に基づく最先端の手法: SFMs とその改良モデル
- ◆音声情報の伝達評価の課題と将来性についても議論
 - 個人性の考慮(オーディオグラム, スペクトル時間変調表現)
 - 多面的な観点からの評価指標の開発 [→関連発表: 3-Q-34]

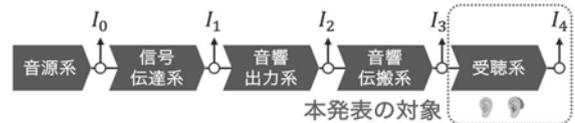


Figure 1: Evaluation points in the sound transmission system.

1-7-2

1-7-2 陰陽・八卦体系を応用した音響生成手法に関する研究*

Interactive Sound Generation Based on the Yin-Yang and Eight Trigrams (Bagua) Framework

☆シヨウメイテツ(日本大学大学院芸術学研究科), 川上央,

三戸勇氣(日大・芸術)

本研究は、易経における陰陽・八卦体系の構造を音響表現へと転換し、身体動作を介したインタラクティブな音響生成手法を提案する。六爻構造を陰陽の二値情報として抽象化し、参加者の動作を契機とするイベント駆動型モデルによって音響パラメータに対応付けた。卦変を連続的な変化ではなく離散的な変化として扱うことで、動作と音の関係を明確化し、音響生成の過程そのものを体験可能な時間構造として提示する。本手法により、易経における「変化」の概念を、身体の動きと音の関係を通して体験的に捉え直すことができる。

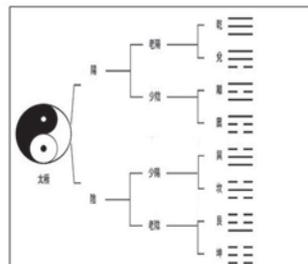


Fig.1 How the Eight Trigrams are generated

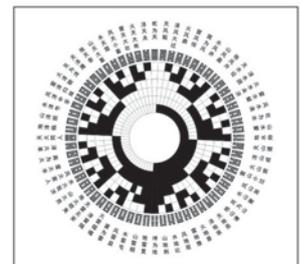


Fig.2 The sixty-four hexagrams

1-7-3

1-7-3 BGMの主観的印象と呈示タイミングの違いが知的作業効率に与える影響

Effects of differences in subjective impression and presentation timing of background music on intellectual work efficiency

☆菅原直也、星野博之(愛知工業大)

◆本研究では、RWC 研究用音楽データベース中の主観的印象(速さ)が異なるクラシック 3 曲を用いて呈示タイミング変化による作業効率向上における有効性調査を行った。

◆実験1: 暗算課題による呈示タイミング変化実験

クラシックのサビ部分の主要評価因子に基づく選曲で、速さ因子が遅い(s)・普通(n)・速い(f)楽曲をそれぞれ課題開始(first)、開始から1分30秒後(center)、開始から3分後(end)の3種のタイミングで5分間の暗算課題中に2分間のみ呈示した結果、遅い曲では「end」に対して「first」で有意(p<0.05)に正答数が多くなった(Fig.1)。

◆実験2: 反応課題による呈示タイミング変化実験

実験1と同じ選曲で、速さ因子が遅い・速い楽曲に対して、実験1と同様に呈示タイミングを変化させ反応課題を行った結果、速い曲では「end」に対して「first」で有意(p<0.05)に反応速度が速くなった。

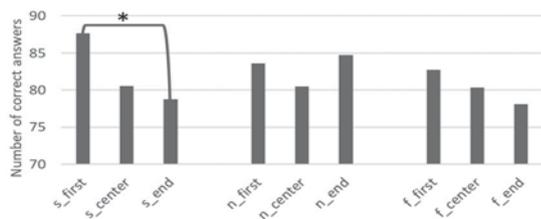


Fig. 1 Results of arithmetic task (Number of correct answers, *: p<0.05)

1-7-5

1-7-5 歌声の映し絵: f_0 高低反転歌声から見えるもの

The portrait of the inverted vocal: What do we see in pitch-reversed voices

○内田照久(大学入試センター・研究開発部)

f_0 高低反転歌声とは

既存の歌声の基本周波数の推移パターンを、対数周波数軸上で高低反転させた f_0 パターンを用いて再合成して生成した歌声

未知の旋律の生成法としての f_0 高低反転歌声

原曲の構成を保持しつつ“未知の旋律”を創出

錯聴作品の制作への展開

f_0 高低反転歌声の特徴を生かした錯聴作品

『歌声の映し絵: 《甲州三坂水面》の調べ』を制作 (Fig.1)

楽譜上の作曲技法としての反行進行との相違点を明示

歌唱時の音高推移の特徴的な現象: オーバーシュートも検討

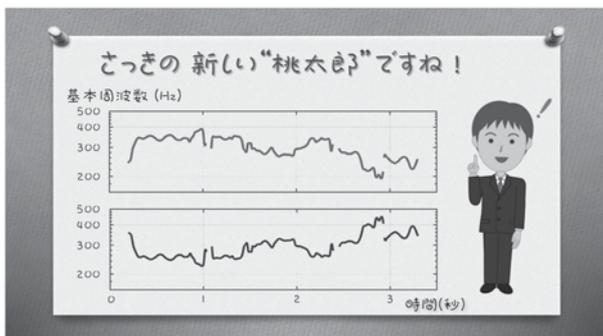


Fig. 1 Screenshot of an auditory illusion artwork

1-7-4

1-7-4 ピアノ楽曲の「ジャズらしさ」に関する知覚的因子分析と推定モデル構築

Perceptual Factor Analysis and Estimation Model Development on "Jazziness" in Piano Music

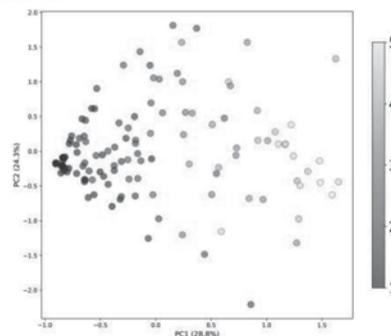
☆KIM Minseok, HUANG Wen-Chin, 戸田智基, (名古屋大学)

◆本研究では、ピアノ楽曲を対象に「ジャズらしさ」の主観評価データを収集した。

◆既存の音楽ジャンル分類器の出力に基づいて楽曲を選定したにもかかわらず、主観評価スコアは低い値の領域に偏った分布を示した。

◆この主観評価行列に対して、欠損値を考慮した非負制約付き行列分解を適用し、楽曲方向の潜在表現を推定した。

◆その結果、潜在表現の主成分と平均主観ジャズらしさスコアとの間に強い相関が観察された。



1-7-6

1-7-6 リミックスモデルを用いた自動コード認識の精度向上

Accuracy improvement of automatic chord recognition using remix models

○三苦歩夢, 古家賢一(大分大)

◆楽曲からコード進行(時間の経過に伴う和音の並び)を認識させることで、楽曲の特徴を捉えることができるため、カバーソングの検索やキー推定などの音楽タスクに応用されている。

◆研究目的: 様々な音楽タスクを高いレベルで行うために、基礎となるコード認識の認識精度を向上させること。

◆精度の向上に向けて、楽器ごとの音量調節を行う深層学習モデルであるリミックスモデルを導入する。Fig. 1に具体的な処理の流れを示す。楽曲を音源分離によって4つの楽器に分離し、これらを入力として各楽器の周波数ごとの音量調節およびリミックスを行う深層学習モデルをコード認識の前処理として導入した。

◆その結果、tetrads(四和音の精度)が58.24%から60.10%へと1.86%向上することが確認できた。

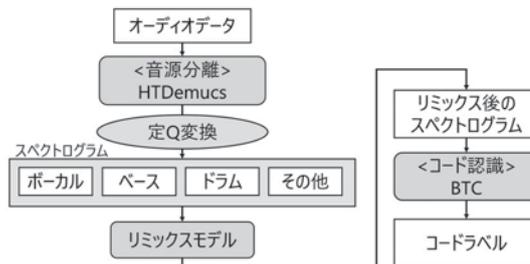


Fig.1: Workflow of chord recognition for the proposed method

1-7-7

1-7-7 敵対的生成ネットワーク型画像スタイル変換に基づくアコースティックシミュレータ

Acoustic Simulator based on Image Style Transfer using Generative Adversarial Network.

☆齋藤太陽, 及川靖広(早大理工)

- ◆エレクトリックギター演奏音からアコースティックギター演奏音への変換について、生成 AI を用いた音声変換技術による新たな手法を提案する。
- ◆MaskCycleGAN-VC を参考に、メルスペクトログラムメルバンド数変更および波形復元過程における Griffin-Lim アルゴリズム導入により、ギター音変換へ適したシステムを構築した。
- ◆システム実装の結果、エレクトリックギター音はアコースティックギター音へ近づいたが、過学習や基本周波数不安定化などの課題が残った。
- ◆基本周波数や特定帯域を重視した新たなロス関数の導入を検討中である。

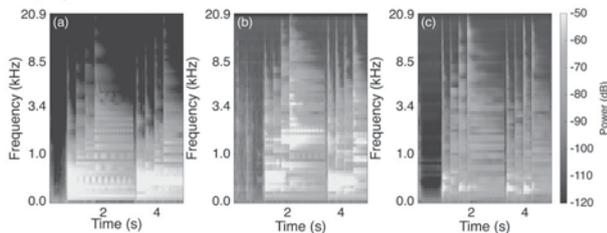


Fig1. (a)Electric Guitar sound, (b)Acoustic Guitar sound, (c)Converted Electric Guitar sound (500 Epoch) Mel-Spectrograms

1-7-9

1-7-9 深層学習を用いたピアノ演奏の自動採譜における強弱記号推定の時系列化

CNN-LSTM: Dynamics Estimation of Piano Music Using Deep Learning for Automatic Piano Transcription

☆文野太陽(愛知工科大学院), 實廣貴敏(愛知工科大学)

- ◆音楽情報検索 (MIR: Music Information Retrieval) における自動採譜研究の予測ターゲットは音楽における基礎的な構成要素のみであり、楽譜上における p (piano, soft), f (forte, loud) など音楽において特に重要な表現を広げる要素であるダイナミクスの変化はあまり検討されていない。
- ◆既発表では、楽曲全体に対し単一の強弱記号を推定する手法を検討したが、実際の演奏における強弱はフレーズや小節ごとに経時的に変化するため、演奏表現の推定としては不十分であった。
- ◆本研究では、2層の CNN と LSTM を組み合わせた CNN-LSTM によるピアノ演奏音源に対して一定区間ごとに強弱記号の推定を行い、時系列情報を含んだ曲中の動的なダイナミクス推定を提案する。(Figure 1)
- ◆正規化した RMS と MIDI Velocity の加重和による強弱記号データセットを用いての実験の結果、CNN-LSTM は一定のダイナミクス変化の推定が可能なが示された。しかし、ff, pp などの音圧が極端なデータに対しては極めて低いスコアとなった。

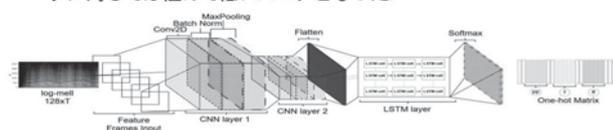


Figure 1 CNN-LSTM model Architecture

1-7-8

1-7-8 ピアノ譜から弦楽四重奏譜への深層学習を用いた自動編曲

Deep Learning-Based Automatic Arrangement from Piano Score to String Quartet Score

☆篠塚颯斗, 小坂哲夫(山形大院・理工学研)

- ◆深層学習モデル「U-net」を用いた自動編曲システムでピアノ譜から弦楽四重奏譜への自動編曲実験を行った。
- ◆自動編曲システムにおける生成結果では、楽器の音域に合わせた音符が生成されている。
- ◆各パートは音高順にある程度分類され、パートの特徴を捉えた編曲が可能。
- ◆自動編曲システムの客観評価実験では、メロディーを担う第1ヴァイオリンとベースラインを担うチェロの評価は高い結果であった (Table 1)。
- ◆しかし、楽曲ごとの客観評価の差が大きい問題があり、その問題に対し、入力楽譜と目標楽譜の類似度と F 値の関係を分析した。

Table 1 Evaluation results for 5 evaluation data

	適合率	再現率	F 値
第1 ヴァイオリン	0.833	0.883	0.777
第2 ヴァイオリン	0.669	0.518	0.440
ヴィオラ	0.759	0.425	0.381
チェロ	0.743	0.718	0.643
全パート	0.751	0.636	0.561

1-7-10

講演取消

1-7-11

1-7-11

デフォーカスを用いた弦振動二次元変位測定法の光学モデル

Optical model for two-dimensional displacement measurement of string vibration by defocusing

☆ 新美智也, 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

課題 ● 絞り値や仰角など撮影条件の設定によって観測されるデフォーカス像の形状が変化する
▷ 撮影条件と像形状の対応関係が不明瞭

今回 ● 幾何光学によりデフォーカス像をモデル化
▷ 観測画像と提案モデル画像の比較を行い提案モデルの妥当性を確認

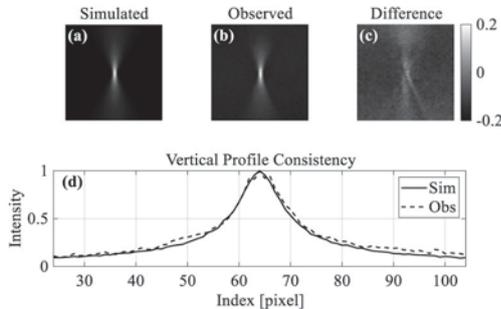


Fig. (a) Simulated image, (b) observed image, (c) peak-normalized difference, and (d) vertical profiles through the peak

1-7-13

1-7-13

ドラムの境界条件を考慮した位相の凸最適化による振動計測

Vibration measurement via convex optimization of phase considering boundary condition of drums

© 古平真理子, 赤石夏輝, 矢田部浩平 (農工大)

最適化に基づく FTP

- FTP は縞状の構造化照明を用いた三次元形状計測手法
- 前回, 最適化問題として解釈することで, 計測対象の領域を自由に設定できる手法を提案した
- この手法をドラムの振動計測に適用したとき, その縁の部分において誤差が残った

提案手法

- ドラムの縁における変位は常に 0 という境界条件を正則化項として前回手法に導入し, 再定式化
- 縁での位相推定精度が向上し, 誤差がさらに低減した

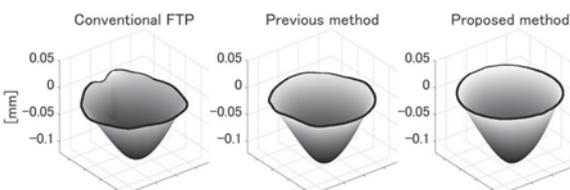


Fig.1 Visualization of the snare drum displacement at a given time

1-7-12

1-7-12 音叉の形状パラメータが振動継続時間に与える影響の検討

Effects of shape parameter on the vibration duration of tuning forks

☆ 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。

◆ 本研究では, 音叉の幅, 厚さ, 左右脚部の間隔の精密測定を実施し, 形状パラメータが振動の持続時間に与える影響について検討を行った。また, 左右の各脚部間の幅・厚さの寸法差を算出し, 脚部の対称性が振動の持続時間に与える影響についても検討した。

◆ 音叉脚部の厚さと振動継続時間には正相関が確認された(Fig. 1)。厚さの増加により曲げひずみが低下し, ひずみ依存の内部損失が相対的に小さくなったことが原因と考えられる。

◆ 左右の各脚部間の幅・厚さの寸法差が大きいほど, 振動継続時間は短縮傾向にあった(Fig. 2)。周波数調整のための研磨工程は専ら職人の感覚に依存しており, 結果として微妙な寸法の左右差が生じると考えられる。研磨工程の最適化によって, 振動継続時間の個体差を低減できる可能性が考えられる。

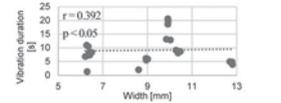


Fig.1 Relationships between the width and the vibration duration.

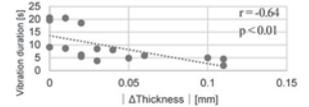
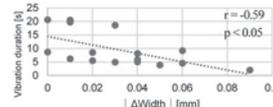


Fig.2 Relationships between the dimensional differences in width (left) and thickness (right) between the left and right tines and the vibration duration.

1-7-14

1-7-14 オーボエの数値シミュレータにおける設計パラメータが演奏音へ及ぼす影響

The influence of design parameters in an oboe numerical simulator on the produced sound

☆ 黒澤文彦 (筑波大・院・シス情工), 若槻尚斗, 海老原格 (筑波大・シス情系)

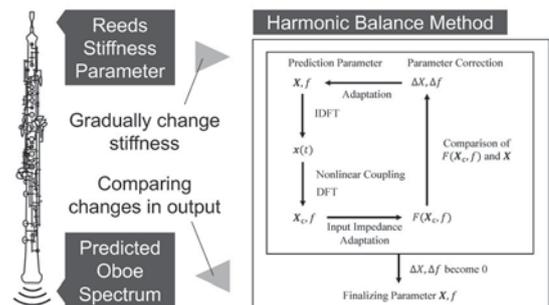


Fig.1: Research Overview (Numerical Simulation Using the Harmonic Balance Method, Stiffness Parameters, and Predicted Spectra)

◆ 本研究は調和平衡法によって構築したオーボエの数値シミュレータにおいて, リードの剛性パラメータを変化させた時の予測演奏音の変化を評価することを目的としている (Fig.1)。

◆ 数値シミュレーションの結果として, リードの剛性が増加するにつれてスペクトル重心が低くなる, RMS が大きくなるなどの音響特徴に違いがみられた。

◆ 数値シミュレーションを利用して定量的にリードの剛性と演奏音の相関を確認できることが示唆され, リードの設計においてこのような手法を活用する足掛かりとして, 有用性が示された。

1-7-15

1-7-15 人工吹鳴による声道音響特性が
トランペットの発音に及ぼす影響の計測
Measurement of vocal tract acoustic effects
on trumpet sound production by artificial blowing

◎白井浩史, 猪俣優響, 若槻尚斗, 善甫啓一, △前田祐佳 (筑波大)

- ◆人工吹鳴実験により、声道の音響特性がトランペット高音域の発振モード選択に及ぼす影響を検証した。
- ◆可変長円筒共鳴器を用いた簡易声道モデルを人工唇の上流に付加し、吹鳴圧力と共鳴胴長さを制御して発振特性を測定した。

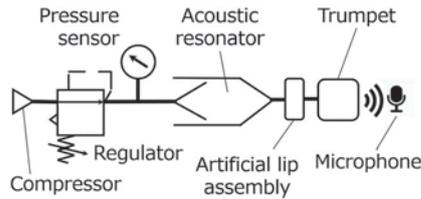


Fig. 1: Artificial blowing test device

- ◆シミュレーションと理論解析により、声道音響が吹鳴システムの利得を変化させ、モード7 (A^b5)と8 (B^b5)の発振の選択を支配する機構を説明できた。

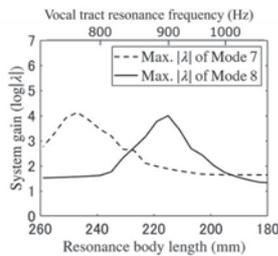


Fig. 2: Blowing sound system gain

1-7-17

1-7-17 フルートのエッジ形状が流れと音に及ぼす影響
Effects of edge shape on flow and sound in flute blowing

◎小野木君枝, △横山博史, 飯田明由 (豊橋技科大)

- ◆エッジ内壁の傾きが異なる2種類のフルート頭部管を対象に空力音直接計算を行った。
- ◆呼気ジェットがエッジ壁面と干渉することで、エッジ付近が高圧となる。
- ◆吹鳴条件は一定であっても、エッジ内壁の傾きを変えることで、エッジ付近の圧力分布が変化し、ジェット進行方向が変化することで、倍音構造が変化すると考えられる。

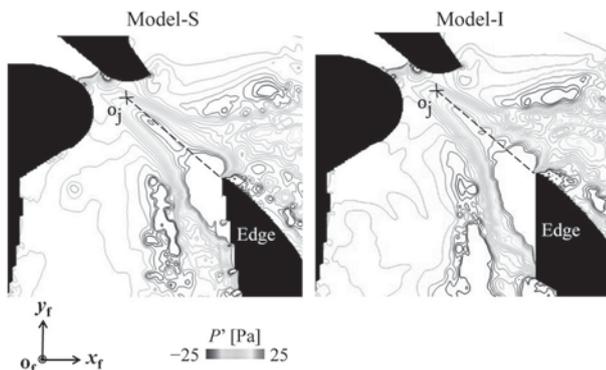


Fig. 1: Time-averaged pressure distribution.

1-7-16

1-7-16 深層学習を用いる金管楽器
入力インピーダンスに基づく
断面積プロファイル推定

Deep learning-based estimation of cross-sectional profiles for brass instruments based on input impedance

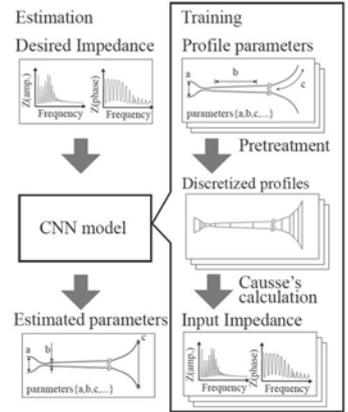
◎猪俣優響 (筑波大・院), 若槻 尚斗, 海老原格, 前田祐佳 (筑波大・シス情系)

<研究背景> 金管楽器の音響特性から管体形状を決定する逆問題よ、非線形性が強く解決が困難である。本研究では1次元量のみミニニューラルネットワークを用い、入力インピーダンスから形状を規定する断面径パラメータを推定する手法を検討した。

<方法> 断面径に関する5つの幾何学的パラメータを推定対象とした。入力データには伝達行列法より算出したインピーダンスの振幅および位相特性を用い、CNNによりこれら音響特性と形状パラメータ間の相関を学習させた。

<結果と考察> 未知の検証データに対し各部位で良好な推定値が得られた。すべての変数でほぼ誤差は同水準であり、全項目で94.5%以上の誤差減少率を記録した。少数パラメータによる検出ではあるが本手法は音響特性から形状を同定する手法として一定の有効性を示した。

<結論> 1D-CNNによる断面径推定の有効性を検証した。今後は推定対象の拡張や実データへの適用が課題である。



1-8-1

1-8-1 音声コーパス研究の流れ
Trends in Speech Corpus Research

◎板橋秀一 (筑波大), 大須賀智子 (NII)

- ◆音声研究を進めるために音声データが必要なことは言うまでもない。音声研究に使用する音声データを体系的に集めたものを音声コーパスと呼んでいる。音声研究の進展により、各種の分析・認識の手法等を客観的に評価することが求められるようになってきたが、それを行うためには、共通の音声データを分析・処理してその結果を比較することが唯一の方法である。そのため共通に使える音声コーパスが必要とされるようになってきた。
- ◆このように、音声コーパスは音声研究のためのデータとして使うことのほかに、研究過程における各種の手法の評価のために必要である。さらには日々失われていく危機言語や方言の保存という使命もある。これに関連して1970年代の後半から、音声認識装置の性能評価法についての論文が発表されるようになった。
- ◆このような背景のもと、1990年代になると音声データを集積・配布・普及するための組織として、アメリカではLDC、ヨーロッパではELRAが活動を開始し、音声研究を促進する力となった。やがて日本でもGSK(言語資源協会)が組織され、後に国立情報学研究所により音声資源コンソーシアム(NII-SRC)が活動を開始した。これら音声資源の集積・配布・普及等の活動は音声研究の推進に貢献してきた。2010年にはNII情報学研究データリポジトリ(IDR)が設置され2026年度からSRCの業務を引き継ぐ予定である。音声コーパスに係わる世界の動きと日本の活動を振り返り、今後の活動の方向性を探る。

1-8-2

1-8-2 音声資源コンソーシアム(NII-SRC)の20年と今後

Twenty Years of the Speech Resources Consortium (NII-SRC) and the Road Ahead

○大須賀智子(国立情報学研究所)

- ◆国立情報学研究所(NII)に2006年に設置された「音声資源コンソーシアム(SRC)」は、音声コーパスの収集と研究目的での配布を行う事業を推進してきたが、20年の節目に本事業を同じNIIに設置された「情報学研究データリポジトリ(IDR)」に引き継ぐこととなった。
- ◆この20年の間に、取扱う音声コーパスは55種にのぼり、2025年末時点のユニーク利用者数は2,240研究室、配布したコーパス数は5,761件である。申込みの内訳をFig.1に、配布コーパス数の推移をFig.2に示す。
- ◆工学分野では、大規模言語モデルの進展により今後はウェブ上から収集した大規模なデータの利用が主流になると思われるが、研究者が構築する高品質なコーパスも追加学習や評価実験において需要が見込まれるし、教育利用や、特に音声コミュニケーション研究においては変わらず必須の研究資料となる。引き続きIDRにて音声コーパスの提供、利用を受け付けるので、是非ご活用いただきたい。

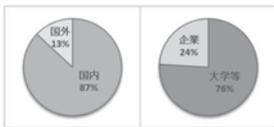


Fig. 1 Breakdown of Speech Corpus Applications to NII-SRC (Per Laboratory)

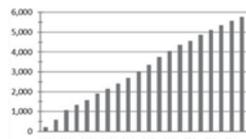


Fig. 2 Number of Distributed Speech Corpora from NII-SRC (Per Laboratory)

1-8-4

1-8-4 感情評定値付きオンラインゲーム音声チャットコーパス(OGVC)の構築と公開およびその展開

Construction, release, and applications of online gaming voice chat corpus with emotional label (OGVC).

○有本奏子(千葉工大)

- ◆2013年にOGVCを国立情報学研究所音声資源コンソーシアム(NII-SRC)から公開
- ◆OGVC構築の狙い
対話中に自然に発した音声から感情を認識するための音声資料を作成する
- ◆OGVC公開の狙い
日本語の自発対話音声に含まれる真の感情を対象に感情音声研究に挑戦する契機を作る
- ◆OGVCによる研究の展開
配布音声に新たなラベルを施すことで、コーパスの活用可能性を拡大
- ◆本稿では、感情音声コーパスであるOGVCを利用した研究の展開について概観する



Fig. 1: CD Labels for OGVC.

1-8-3

1-8-3 山梨県奈良田方言音声データベースの構築

Design and development of a speech database for Narada Japanese

○小西いづみ(東大院・人文社会)

- ◆この発表では、現在構築中の奈良田方言音声データベースの設計および現状と展望について述べる。奈良田は山梨県早川町の北端の集落で、奈良田方言は上昇の有無・位置によるアクセント体系を持つ「言語の島」である。現在の流暢な話者は高年齢者しかおらず若い世代に継承されていないが、言語に対する居住者の関心は高い。このデータベースを活用して文法書、辞書、学習テキストを編み、学術研究利用だけでなく、奈良田方言を学習できる環境を作ることを目指している。
- ◆データベースには次の種類の発話が含まれる。
A. 単語・語句の発話, B. 短文・短文連続の発話, C. 談話
Cには、次のものがある。
C1. 自然談話, C2. ロールプレイ, C3. 物語(独話)
- ◆各レコードには、音声ファイル、その文字化テキストとアノテーション(形態素へ分析、品詞(形態素クラス)情報、意味・機能のグロス)、共通語訳が含まれる。形態素にIDを与えて辞書で管理し、自作の簡易な形態素解析プログラムで自動的にアノテーションを付与した後、人手で修正している。形態素クラスを2段階に分けることで、形態素どうしの接続規則を簡略に記述できている。
- ◆公表済データは動詞・形容詞の活用形とそのアクセントを扱うものだけで数千件で、最終的には短文だけで万単位になる。このデータベースとアクセント情報を含む辞書が拡充されれば、自動音声認識・音声生成も可能だと思われる。

1-10-1

1-10-1 頭部伝達関数の卓越帯域強調による騒音下正中面音像定位精度の向上

Improvement of median plane sound localization accuracy under noisy conditions by emphasizing dominant bands of head-related transfer function

○中村風香(千葉工大) 渡邊悠希 加古達也(NTT) 飯田一博(千葉工大)

- 妨害音が存在する音環境における正中面音像定位の精度向上を目的として、頭部伝達関数の卓越帯域を強調する方法を提案した。上半球正中面7方向での音像定位実験の結果、卓越帯域の強調の効果として以下の結論を得た。
- 全ての上昇角において、強調量の増加に伴って音像方向が音源方向に収束する傾向が観察された(Fig.1).
 - 2) 全ての上昇角において平均上昇角誤差が減少した。しかし、今回の強調量の範囲では、特に上方では妨害音なし条件と比較すると依然として大きい。妨害音による上昇角知覚の手がかりのマスクングが完全に解除されたとは言えない。
 - 60°を除く全ての上昇角で前後誤判定率が減少した。上昇角60°の音源に対しては90°を少し超えた上昇角への回答が増加したため、前後誤判定率は減少しなかった。

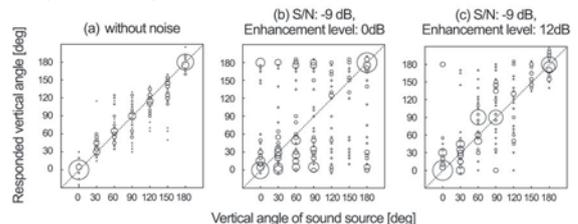


Fig. 1. Responded vertical angles for a male voice: (a) noise-free condition; (b) S/N = -9 dB with no enhancement; (c) S/N = -9 dB with a 12-dB enhancement of the dominant band.

1-10-2

1-10-2 卓越帯域の導入による Parametric Notch-Peak HRTF model の拡張

Extension of the Parametric Notch-Peak HRTF Model by introducing dominant bands

○中村風香 飯田一博(千葉工大)

- 上昇角知覚の手がかりとして貢献する HRTF の卓越帯域を導入した拡張 PNP HRTF モデルを構築した(Fig. 1). さらにこれを用いて全天空の個人化 HRTF を生成するアルゴリズムとアプリケーションを開発した.
- 音像定位実験により最適な卓越帯域レベルを求めた結果, 前方・上方・後方(0, 90, 180°)では比較的大きく, 他の方向では小さい傾向があった(Table 1).
- 最適な卓越帯域レベルを用いた拡張 PNP モデルで生成した個人化 HRTF により, 概ね目標の上昇角に音像を知覚したが(Fig. 2), 今後, ナイーブな被験者を用いた検証が必要である.

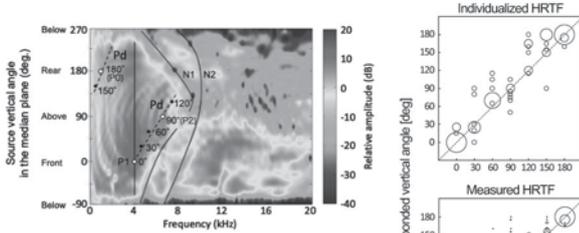


Fig. 1 Relationships between N1, N2, and P1 (solid lines), and Pd (dashed lines) for a representative subject in the median plane.

Table 1 Optimal Pd level for seven vertical angles in the upper median plane [dB].

Vertical angle [deg]	
0	30
6	4
8	8
18	4
2	14

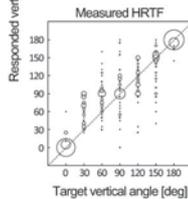


Fig. 2 Responded vertical angle in the median plane.

1-10-4

1-10-4 高齢者の音声了解度に対する末梢系 聴力損失とそれ以降の影響の切り分け

Separating the effects of peripheral hearing loss and higher-level processes on speech intelligibility in older adults

○ 入野 俊夫 (和歌山大/電通大) 山本絢子 宮崎美紀(和歌山大院)

- ◆ 音声知覚において, 聴覚末梢系の聴力損失とそれ以降の影響について議論するための新しい手法を提案した.
- ◆ 模擬難聴システム WHIS を用いて, 1 人の高齢者の聴力レベルに対応する模擬難聴音を若年健聴者 14 名に聴取させて音声了解度を得た. 高齢者の了解度の方が高いことがわかったが, 1 人しか評価できない.
- ◆ 音声了解度客観指標 GESI を用いて, 若年健聴者のデータから決定したパラメータを用いて, 高齢者の了解度を予測した(図 1). この結果, 数名の高齢者で, 平均的な若年健聴者よりも了解度が高く, 1 名だけ低かった. GESI はボトムアップ処理だけで構成されるので, この差分は高次処理の影響を反映していると考えられる.

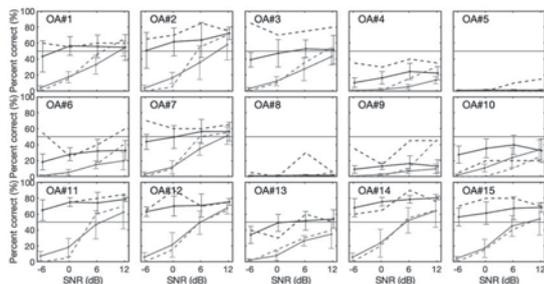


Fig. 1: Subjective speech intelligibility (SI) scores of individual older adults (OAs) and GESI predictions using parameters determined in the SI scores of young normal hearing listeners.

1-10-3

1-10-3 水平面の音像定位を活用した認知性向上 -第2報 聴覚定位感度に基づく情報音の提示-

Cognitive Enhancement Using Sound Image Localization on Horizontal Surface - Part 2: Presentation of Informational Sounds Based on Auditory Localization Sensitivity -

☆菅田康平(中央大院), △松本和也(中央大)

小暮知史, △丹野慶太(アルプスアルパイン), 戸井武司(中央大)

- ◆ 聴覚情報による運転手の安全性や快適性の向上を目標として, Fig. 1 に示す高さの異なる水平面における音像の活用を検討する.
- ◆ 仰角 30° における上水平面 H_U と基準水平面 H_S, 仰角 -30° における下水平面 H_L の 3 つで方位角 60° 毎の 6 方向の頭部伝達関数を取得した.
- ◆ 水平面毎の平均誤差角度の大きさは多くの条件下で下方の方向認知精度が高い傾向となるが, 被験者や音源によって傾向が変化することが示された.
- ◆ 被験者と水平面を考慮した音源によって, Fig. 2 のように方向認知精度が向上することを把握した.

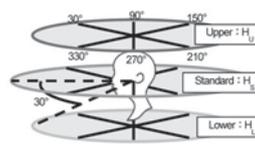


Fig. 1 : Definition of a horizontal plane

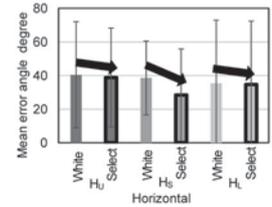


Fig. 2 : Change mean error angle of horizontal

1-10-5

1-10-5 Quality of an extremely low-bit-rate speech coding: 1-bit vocoding and the number of frequency bands

○ Kazuo UEDA (Kyushu Univ.), Ikuo MATSUO (Tohoku Gakuin Univ.), Yoshitaka NAKAJIMA, Gerard B. REMIJN, and Emi HASUO (Kyushu Univ.)

- ◆ An extremely low-bit-rate speech coding technique without using predictions, so-called 1-bit vocoding, has been proposed, achieving 80% intelligibility with only 2,400 bits per second (bps).
- ◆ It is based on infinitely peak-clipped speech signals and on speech amplitude envelopes in four frequency bands that are ubiquitous across eight languages/dialects.
- ◆ To our knowledge, however, the effect of the number of frequency bands on the quality of 1-bit vocoded speech has not been examined, and therefore, we examined it.
- ◆ Five steps of the number of frequency bands (1, 2, 4, 8, and 16 bands) were combined with 10 Japanese sentences spoken by a female and a male talker, and presented to seven listeners through headphones.
- ◆ The listeners were instructed to respond verbally with any number other than zero or a negative value (magnitude estimation) to the quality of the synthesized speech.
- ◆ The results can be approximated with Stevens' power law.
- ◆ Further investigations are warranted to dissociate the potential influence of intelligibility from perceived speech quality.

1-10-6

1-10-6 宣伝音声の話者と聴取者の年齢が媒介モデルに及ぼす影響:多項式媒介分析による検証

Effects of speaker and listener age on a mediation model of advertising speech: A polynomial mediation analysis

○長野瑞生, 井島勇祐, 廣谷定男(NTT)

- ◆話者と聴取者の年齢類似性が購買意欲に与える影響は一貫しない
 - 広告に客層と年齢の近い俳優を起用することで購買意欲が高まる
 - 年齢類似性の高さが必ずしも購買意欲に影響するわけではない
- ◆音声広告では平均F0や話速等の音声特徴から話者年齢を推定
 - 音声特徴の知覚は聴取者年齢によって異なるため、年齢類似性の影響も聴取者年齢によって変わる可能性がある
- ◆音声から生起する購買意欲における感情・音声印象の媒介効果に対する話者・聴取者年齢の非線形な交互作用を多項式調整媒介分析で調査
 - 聴取者年齢が高いほど快感情の媒介効果は高く、覚醒の媒介効果は低い
 - 話者年齢が上がるにつれ快感情・覚醒の媒介効果が高まる
 - 快感情・覚醒・声の明るさ・温かさで、話者年齢と聴取者年齢の三者交互作用が有意。いずれも媒介効果が高い要素であり、購買意欲が話者年齢・聴取者年齢の両方の影響によって左右される可能性を示唆

Table1: Three-way interaction effects on the X→M and M→Y paths through perceived emotions and voice impressions, involving listener's and speaker's age.

Path Mediator	Speech features → Mediators (X → M)			Mediators → PI (M → Y) Purchase intention
	Mean F0	Speech rate	SD F0	
(P)Unpleasant-Pleasant	0.014	0.0013	0.0005	-0.040**
(A)Relaxed-Aroused	0.030**	0.035**	0.002	0.017
(D)Submissive-Dominant	-	-	-	-
(a)High-pitched-Low pitched	-0.005	-0.003	-0.011	-0.010
(b)Hoarse-Clear	-0.008	-0.008	0.009	-0.004
(c)Calm-Unstable	-0.003	-0.012	0.007	-0.001
(d)Powerful-Weak	-0.003	-0.006	-0.006	0.002
(e)Bright-Dark	0.001	-0.032**	-0.006	0.010
(f)Cold-Warm	-0.003	0.024	-0.027*	0.003
(g)Youthful-Elderly	-	-	-	-

*p < .05, **p < .01

1-P-2

講演取消

1-P-1

1-P-1 雑音環境下における蟬音源推定アプリの時間窓長の検討

A study on window length for smartphone application to estimate cicada sound source direction in noisy environments

☆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、ミンミンゼミと指向性狭帯域ノイズを同時に設置したときの方向推定を、推定窓長を変化させてシミュレーション上で行った。
- ◆推定窓長が、鳴き声の「ミ」よりも短いとき、安定した方向推定の結果を得ることができた。

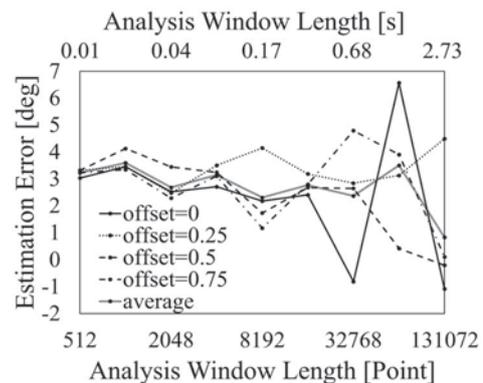


Fig.1: Estimation results for each offset relative to the window length

1-P-3

1-P-3 振り子型マイクロホンによる単チャンネル音源方向定位

Single-channel sound source localization using pendulum microphone

☆高橋祐芽, 赤石夏輝, 矢田部浩平(農工大)

- <背景> 音源方向定位で一般的に用いられるマイクロホンアレイを使用せず、単一のマイクロホンで音源方向定位を行いたい。
- <提案> 振り子型マイクロホンを用いて、単チャンネルによる音源方向定位を行う。音源方向の情報はドブラ効果による観測信号の周波数変調に含まれることを利用し、瞬時周波数の推定により方位角を導く。
- <実験> Fig. 1に示す測定系で、方位角θを0°, 30°, 60°の3条件に設定し、マイクロホンを振り子のように振りながら録音した。
- <結果> 瞬時周波数のフィッティングを全探索と勾配法で行い、方位角の推定を行った。0°, 30°, 60°いずれも高精度に推定できた。

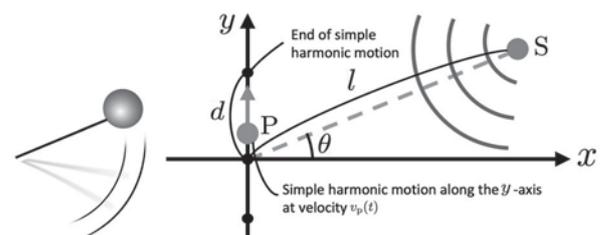


Fig. 1: Schematic diagram of a measurement system using a pendulum-type microphone. P is the observation point and S is the sound source. Sound waves emitted from the source propagate as spherical waves.

1-P-4

1-P-4 周波数重み付け RLS アルゴリズムを用いた GRODMIC による音源定位

Sound Source Direction Estimation Using GRODMIC with Frequency-Weighted RLS Algorithm

☆橋本唯利, 志賀芳則(東京電機大院・工学研)

- ◆音源定位手法として GRODMIC が提案されている。これまで GRODMIC に対して、フィルタバンクを使用し、有音帯域の取捨選択による精度改善を行った。より詳細な有音帯域の判別のためにフィルタバンクの帯域幅を狭めると精度劣化が見られた。これはフィルタ更新を行う RLS アルゴリズムの周波数分解能の不足が要因である。そのため、近似フィルタのフィルタ次数を維持し、狭帯域信号でも高い近似精度を持つアルゴリズムが必要であると主張に至った。
- ◆提案法では、有音帯域が既知である点とアルゴリズムの更新に誤差成分のみが利用される点に着目し誤差信号に周波数重み付けを行う。
- ◆提案法の妥当性を示すため、実環境実験で従来の GRODMIC、フィルタバンク使用 GRODMIC と比較を行った。Fig.1 に推定誤差を示す。
- ◆Fig.1 より提案法による性能改善が確認でき、主張の妥当性が示された。

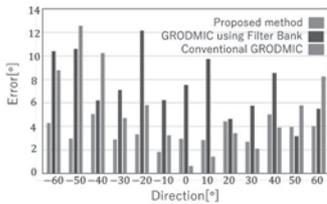


Fig.1: Errors of estimation using conventional GRODMIC, GRODMIC using Filter Bank and proposed method

1-P-6

1-P-6 時間周波数解析の手法を複素スペクトログラムに適用した音声ノイズ除去

Application of time-frequency analysis methods to complex spectrograms for audio noise reduction

☆船波由梨(東京理科大), 鈴木俊夫(神奈川大), △石渡恵美子(東京理科大)

- ◆従来法 [2015, 千葉, 矢田部, 及川]: カーブレット変換による手法
 - 2次元の複素スペクトログラムにカーブレット変換を適用して、閾値処理によりノイズ成分を取り除く手法を提案 (Fig.1)
 - カーブレット変換は画像の曲線的なエッジを抽出し、元の音声を持しやすいが、計算量が非常に多く、実行時間は増大
- ◆本研究: カーブレット変換を他の時間周波数解析の手法に変更
 - 計算量が少ない2次元離散ウェーブレット変換やシアレット変換により (Fig.1), ノイズ除去の性能を変えずに実行時間を短縮
 - 2次元離散ウェーブレット変換による手法は、ノイズをほとんど除去しつつ、非常に高速
 - シアレット変換による手法は高速で、カーブレット変換と同程度に、振幅スペクトログラム上のノイズの除去が可能

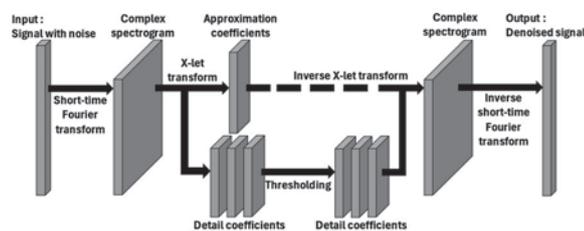


Fig.1: Denoising with complex spectrograms and time-frequency methods

1-P-5

1-P-5 音源推定における非等間隔分散型マイクロホンアレイの構築

Construction of non-equidistant distributed microphone array 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。
- ◆2個, 3個, 3個の3群構成と、新たに構築した非等間隔分散構成の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、7-8 kHzの帯域制限した白色雑音を再生した小型スピーカを20箇所に配置し、音源推定の誤差を求めた。
- ◆非等間隔分散構成は3群構成と比較して、特定の位置で生じる大きな誤差を抑制し、音源推定精度を向上させた。

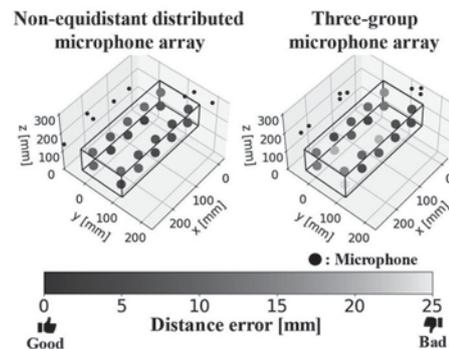


Fig.1. Accuracy of sound source estimation using non-equidistant distributed microphone array and three-group microphone array

1-P-7

1-P-7 ディリクレ非負値行列因子分解のハウリング抑圧への応用

Application of Dirichlet Nonnegative Matrix Factorization to Howling Suppression

◎小川遼, 北村大地(香川高専), 綾野翔馬(科学大)

- ◆我々の提案しているディリクレ事前分布に基づく非負値行列因子分解(ディリクレ NMF)の応用例として、ハウリングの抑圧問題に適用
- ◆ディリクレ NMF は **one-hot な基底ベクトル**(単一周波数の正弦波信号のスペクトル基底)を誘導可能
 - ➡ 「ピー！」というハウリング音をモデル化できる!
- ◆より強力な one-hot 基底ベクトルへの誘導(正則化)として、ディリクレ分布の制約を超えた**拡張ディリクレ NMF**を新たに提案
- ◆音声・音楽信号におけるハウリングの抑圧実験
 - 信号対歪み比の改善量: 音声 7.3 dB・音楽 2.1 dB

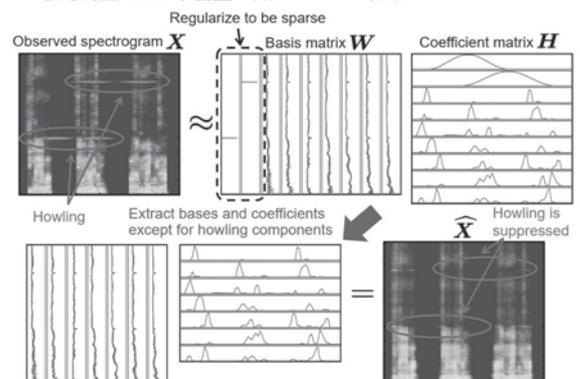


Fig.1: Howling suppression framework based on NMF.

1-P-8

1-P-8 位相を含む時空間スペクトル画像に 対する欠損修復に基づく雑音抑圧の改良

Improved Noise Suppression Based on Inpainting 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: 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の雑音抑圧手法では雑音の完全な除去が難しく、高周波帯域に雑音が残留しやすい。
- ◆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と複数話者学習により、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2D FFTを行い、振幅、 \cos (相対位相)、 \sin (相対位相)を3チャンネル画像とし、画像修復型の拡散モデルに入力する。これにより、目的音が局在する空間直流成分を修復し、得られた結果から目的音を抽出する。
- ◆結果: 女性および男性話者条件において、従来法に比べ雑音抑圧量(NSL)が向上し、条件間の性能差が縮小した(Fig. 1)。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。

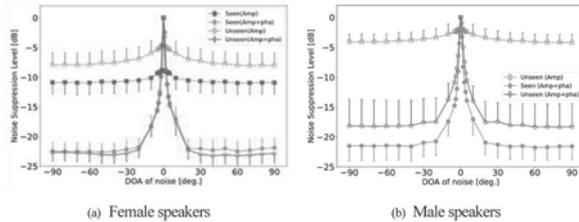


Fig. 1 Performance evaluation based on noise suppression level (NSL)

1-P-10

1-P-10 時空間音圧分布画像を用いた 音源分離にマイクロホンアレイの 位置ずれが及ぼす影響

The effect of microphone array misalignment on the performance of sound source separation using spatio-temporal sound pressure distribution image

☆富田浩音, 坂本修一(東北大通研/院情科研), 鳥谷輝樹, 小澤賢司(山梨大院総合研究部)

- ◆本研究では、時空間音圧分布画像 (Spatio-temporal sound pressure distribution image: STSPD image) を用いた音源分離法におけるマイクロホンアレイの位置ずれの影響を評価した。また、従来のビームフォーミングによる方法と位置ずれに対する感度の違いを比較した。
- ◆時空間音圧分布画像を用いた音源分離法はビームフォーミングによる方法よりも位置ずれに対して頑健であり、位置ずれが生じた場合でも良好な音源分離性能を有することが示された。

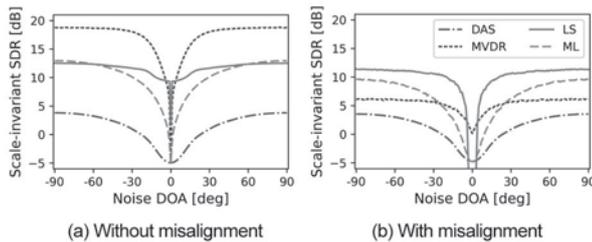


Fig. 1: Scale-invariant Signal-to-distortion ratio (SDR) as a function of noise direction of arrival (DOA), calculated by STSPD image-based methods (LS, ML), delay-and-sum (DAS) beamformer, and minimum variance distortionless response (MVDR) beamformer.

1-P-9

1-P-9 Lipschitz 連続性を保証した 時間周波数マスキング型 DNN

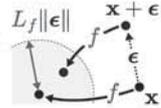
Deep neural networks based on time-frequency masking
with guaranteed Lipschitz-continuity

◎松本和樹, 内田蓮, 矢田部浩平(農工大)

背景: Lipschitz 連続な振幅操作型 DNN

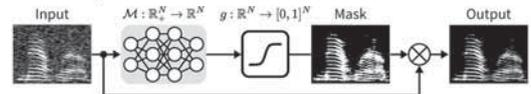
- Lipschitz 連続性は DNN の挙動に対する理論保証に有用 e.g., 最適化アルゴリズムの収束性保証

- ▶ LipsAM (Lipschitz-continuous Amplitude Modifier) 我々は複素スペクトログラムの振幅を処理する Lipschitz 連続な DNN を提案



本研究: LipsAM-SM (Signal Masker)

- 既存の LipsAM はマスキングに対応していなかった



- Lipschitz 連続性を保証するための M と g の条件を導出
- Lipschitz 連続なマスキングを実現するための機構を提案



$$\text{LipsAM-SM} : \mathcal{D}_{(\mathcal{M}, g, \alpha, \beta)}^{(\text{Lips})} = \text{sigmoid}(\max(\mathcal{M}(|z|), \alpha|z| + \beta)) \odot z$$

※ $\mathcal{M} : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ は Lipschitz 連続, $\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}$

1-P-11

1-P-11 重複発話区間に限定した音声分離

Speech separation restricted to overlapping speech regions

◎諸橋 博昭, 相原 龍, 三井 祥幹, 伊佐野 勝人, 齊藤 辰彦(三菱電機)

本発表では音声認識フロントエンドとして、音声分離を重複発話区間に限定して適用することで、処理の高速化と音声認識精度の向上を実現した結果について報告する。

発話全体にシングルチャネル音声分離を適用した場合と、音声分離を重複発話区間検知に限定する提案手法を比較した結果、音声認識率の向上と処理の高速化を実現した。提案手法のイメージ図を図1に示す。

音声認識率と動作速度の評価実験を行った結果、音声分離を発話全体に適用した場合と比べ、提案手法では音声認識率が7.9ポイント改善し、動作速度が36%改善した。

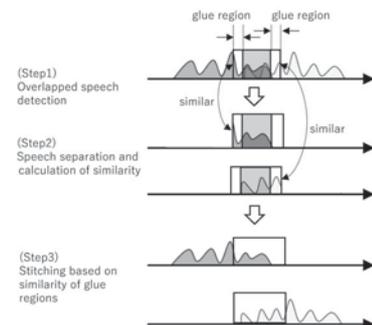


Fig. 1: Schematic illustration of the proposed method.

1-P-12

1-P-12

ブラインド音源分離における 2×2 分離行列の自由度について

On the degree of freedom of 2×2 separation matrices in blind source separation

☆ 下川 凜, 松本和樹, 矢田部浩平 (農工大)



一般の分離行列

- 2音源2マイクブラインド音源分離
- マイク1を基準とする平面波モデルを仮定

混合行列

$$H = \begin{bmatrix} 1 & 1 \\ a_1 e^{i\phi_1} & a_2 e^{i\phi_2} \end{bmatrix}$$

逆行列

分離行列

$$W = \frac{1}{a_1 e^{i\phi_1} - a_2 e^{i\phi_2}} \begin{bmatrix} -a_2 e^{i\phi_2} & 1 \\ a_1 e^{i\phi_1} & -1 \end{bmatrix}$$

行列の自由度 **4**

振幅の制約を追加することで、自由度を削減

制約導入後の分離行列

振幅を共通にする

$$a_1 = a_2$$

$$\frac{1}{a e^{i\phi_1} - a e^{i\phi_2}} \begin{bmatrix} -a e^{i\phi_2} & 1 \\ a e^{i\phi_1} & -1 \end{bmatrix}$$

3

行列の自由度

振幅を1に固定

$$a_1 = a_2 = 1$$

$$\frac{1}{e^{i\phi_1} - e^{i\phi_2}} \begin{bmatrix} -e^{i\phi_2} & 1 \\ e^{i\phi_1} & -1 \end{bmatrix}$$

2

実験結果

- 自由度多 ▶ 性能高
- 残響や窓長の影響大 ▶ 自由度による性能差小

1-P-14

アクティブ聴覚マスキングを用いた オーディオスポット形成における 時間・スペクトル包絡平坦化に基づく 音声漏洩抑制

Speech leakage suppression in audio-spot formation using active auditory masking based on temporal-spectral envelope-flattening

☆ 橋田朋希 (立命館大), 歌航庭, 中山雅人, 西浦敬信 (立命館大)

- ◆ 本研究では、音声漏洩を抑制したオーディオスポットを形成することを目的とする。
- ◆ 提案手法では、マスク信号とアンチマスク信号を用いて聴覚マスキングを能動的に制御するアクティブ聴覚マスキングに、時間包絡とスペクトル包絡の平坦化に基づく目的音声の分割手法を組み合わせる。
- ◆ 音声了解度に関する評価実験により、目的音声分割を用いた場合の音声了解度を評価し、目的音声分割を組み合わせることで、オーディオスポット形成に必要なマスクの音圧を 15 dB 以上低減可能であることを確認した。

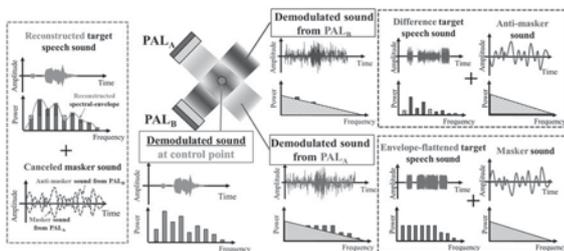


Fig.1: Overview of the proposed method

1-P-13

1-P-13 パラメトリック+フルレンジスピーカを備えた 害獣排除装置の製品化

Commercialization of pest control device equipped with parametric and full-range speakers

○ 小川涼雅, 小川鉄夫, 伊東五朗, 神山史進, 藤原奨(テクノクラート)

- ◆ 近年、異常気象に伴う山間部での作物減少や山間部近くまで拓かれた宅地化などの影響により、野生鳥獣(害獣)による農作物被害の被害や住宅地への動物侵入などによる様々な問題が増えている。
- ◆ 既に音放射を利用した排除装置は市場導入されている。装置設置初期は排除効果が見られるものの、動物側の「慣れ」や装置の故障に伴う稼働停止により、排除効果が薄れる問題を生じていた。
- ◆ 動物行動と屋外設置への耐環境性も考慮し、パラメトリック+フルレンジスピーカによる害獣排除製品の開発と商品化を進めた(図1)。

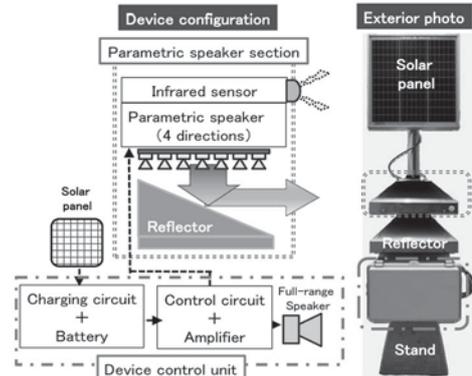


Fig.1: Example of the inside and outside

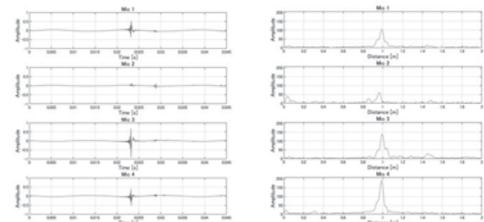
1-P-15

1-P-15 アンビソニックマイク構造4素子系 による音響測距法の基礎的検討

A fundamental study of acoustic distance measurement using a four-element ambisonic microphone system.

○ 中道昇, △清原心(近畿大), 中山雅人(立命館大), 上保徹志(近畿大)

- ◆ 本稿では、位相干渉に基づく音響測距に着目し、アンビソニックマイククロホンを用いた距離推定手法について、基礎的な観点から理論的および実験的に検討する。
- ◆ 今回は初期的な試みとして、アンビソニックマイクロホンの4素子構造に着目し、立体的な音場生成の活用に関しては次回以降に報告する。
- ◆ Fig.1(a), (b) はアンビソニックマイクロホンの4素子における観測信号と距離スペクトルをそれぞれ示す。Fig.1(a)より、音源・対象物の幾何学的配置に応じて各素子が受ける観測信号の強度には差が生じる。しかし、Fig.1(b)を見ると、複数素子の観測結果を総合的に評価することで、より安定した距離推定が可能であることが分かる。



(a) Observed signals. (b) Range spectra. Fig.1: Distance measurement using ambisonic microphone.

1-P-16

1-P-16 嚥下音における音響特徴量に基づく識別性能と汎化能力の検討

Evaluation of Classification Performance and Generalization Ability Based on Acoustic Characteristics of Swallowing Sounds

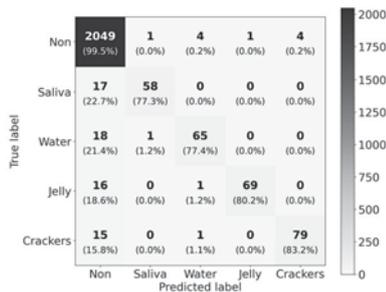
○藤島悟志(国際高専), 南出章幸(金沢工大), △高橋千晶, △経田香織(金沢医科大学), △小木美恵子, △竹俣一也(金沢工大)

◆本研究では、嚥下音に対する複数の音響特徴量を用いた、**嚥下音認識モデルの構築を試行**

【嚥下音: ①空嚥下, ②水 5cc, ③ゼリー 10g, ④クラッカー 3g】

- > 市販の接触型マイクを頸部側面の耳下部付近に装着し収録
- > 特に、正常嚥下における対象物ごとの音響的差異を反映した特徴量と分類器の組み合わせによる性能を評価

◆STFT および Mel 特徴量を用いた SVM モデルでは、10 回交差検証法において**再現率や F1 スコアが 96%以上**であり、**テストデータに対しても 96%以上の精度で予測 (Fig.1)**



◆最良モデルの臨床検査データへの適用では、

Fig. 1 Summary of prediction results of test data using the optimal SVM model with Mel features

一部ではあるが嚥下動作を正しく認識でき、被験者分離における汎化性能の可能性を確認

◀ 比較的安価なマイクを使用した環境下でも嚥下物の物性を反映した明確な音響的特徴が内在していることを示唆 ▶

1-P-18

1-P-18

観測信号の周波数特性に基づく時変計量を用いたブロックベース適応フィルタ

Block-based adaptive filtering with time-varying metrics based on the spectral characteristics of observations

☆村田佳斗, 松本和樹, 内田通, 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)

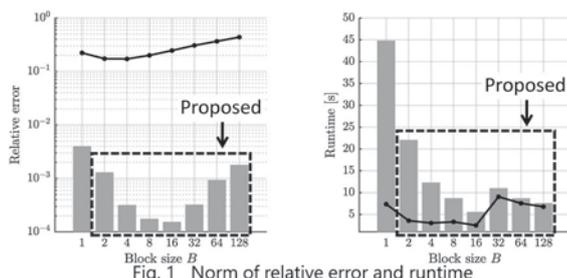


前回 | 観測信号の周波数特性に基づく時変計量を用いたサンプルバイサンプルの適応フィルタ

課題 | 反復毎のフーリエ変換により実行時間が長い

提案 | ブロックベース処理に拡張

結果 | 前回手法・NLMS法より **高い推定性能 (Fig. 1 左)**
前回手法より **高い計算効率 (Fig. 1 右)**



1-P-17

豚呼吸器疾患検知に向けた Transformerアーキテクチャに基づくオートエンコーダの個体適応性の評価

Evaluation of Individual Adaptability of a Transformer-Based Autoencoder for Acoustic Detection of Porcine Respiratory Diseases

☆巽亮太, 神谷太星, 石光俊介, △湯本誠司(広島市大), △田中大介(東北大) △三上修, △小川洋介, △井上寛暁, △石田三佳(農研機構)

◆本研究では、呼吸器疾患罹患豚と健康豚の識別を行うオートエンコーダモデルの構築を目的とし、耳標センサを用いて収録した体内伝導音を用い、個体適応性の評価のためのスペクトログラムの調整と閾値の検討を行った。

◆全ての日程の罹患データと、早期発見のために接種0日から3日目までのデータを評価した AUC_all と AUC_0-3 において、窓長 128、オーバーラップ 0.94 の周波数分解能の高いモデルが概ね最良の分離性能を示した。また、各豚最良のモデルにおいて、閾値を決める係数である k を調整し全頭へ適応すると、病変の程度の低い豚 D で決めた閾値が日程の影響なく安定した Accuracy (Acc.) を示した。今後は、この結果をもとに他の豚のデータを用いた最終的な評価学習を行う。

Table 1 Classification accuracy across all pigs using the optimal model and threshold determined for each pig (A-D: diseased pigs; α: healthy pig).

AUC_all								
pig	BestModel	Best-k	A-Acc.	B-Acc.	C-Acc.	D-Acc.	α-Acc.	Acc._mean
A	128-0.94	0	90.706	90.177	96.53	94.703	22.784	78.98
B	128-0.90	0.5	86.16	91.177	97.481	95.662	39.194	82.055
C	128-0.94	1.1	81.488	90.288	98.191	96.934	53.773	84.123
D	128-0.94	1.1	81.488	90.288	98.191	96.934	53.773	84.123

AUC_0-3								
pig	BestModel	Best-k	A-Acc.	B-Acc.	C-Acc.	D-Acc.	α-Acc.	Acc._mean
A	128-0.94	0	88.579	89.613	95.821	92.382	26.078	78.498
B	128-0.94	0.65	84.999	91.359	97.407	94.691	46.976	83.086
C	128-0.94	1.55	75.739	84.883	99.007	95.141	68.631	84.68
D	128-0.94	1.1	80.932	89.667	98.552	95.591	58.427	84.634

1-P-19

2ヶ所の観測信号を1つの誤差信号としたダクト適応騒音制御における各観測点の収束特性

Convergence characteristics at each observation point in a duct for active noise control applying the addition method of two observed errors

☆柴崎高徳, 武藤憲司(芝工大)

◆2ヶ所の観測信号を1つの誤差信号とした適応騒音制御 (the Addition of Two Observed Errors; A2E を適用した FF-ANC) を提案する。

◆本研究では、ダクト内の定在波の節や腹に限定されない位置で観測される2つの信号を用いて適応騒音制御を行った。

◆実験結果は以下の通りである。

- 400 Hz の正弦波 44.6 s (92.9%) の短縮
- 450 Hz の正弦波 34.1 s (85.9%) の短縮
- 500 Hz の正弦波 2.6 s (21.0%) の短縮

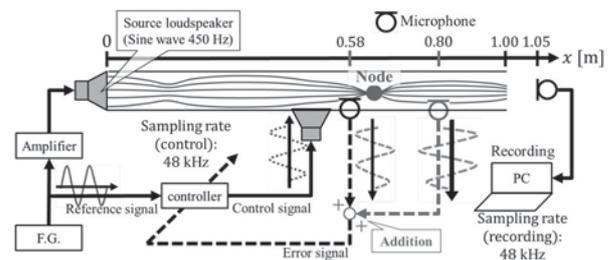


Fig. 1: FF-ANC employing A2E. (450 Hz)

1-P-20

1-P-20

複数の複素スペクトログラムの偏導関数の 零点に基づく正弦波成分の推定

Estimation of sinusoidal components using zeros of partial time and frequency derivatives of multiple complex spectrograms

☆丸林諒, 赤石夏輝, 矢田部浩平(農工大)

背景

音響信号分析において正弦波成分の推定は重要

前回の提案

零点とピークの位置関係から正弦波度合いを推定

課題

周波数の近い正弦波成分では、推定精度が低下する

提案：窓幅を変えた際の零点の挙動に着目する

複数窓幅の零点をフィッティングし周波数を推定

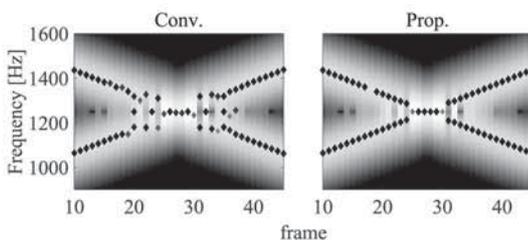


Fig. 1 Spectrogram of two intersecting sinusoidal components. Estimated sinusoidal components are indicated by dots. Left: conventional method. Right: proposed method.

1-P-22

1-P-22 IRMOS による推定インパルス応答を用いた 音響測距における反射素材による 違いの一検討

An investigation on developing Characteristics Analysis Method for Acoustic Reflection board applying Distance Measurement Method with estimated Impulse Response using IRMOS

☆徳富 響(NBU), 副島 翔太(NBU), 森本 祥平 (NBU),

近藤 善隆(ジェイテック), 福島 学(NBU), 兪窪田 泰也(A-Lab.), 柳川 博文(arsl)

- ◆目的：室内音環境を調整するための反射素材の評価
- ◆手法：音響測距法は「送信信号と反射信号の干渉」に着目
「素材への入射信号と反射信号の干渉」が素材の特徴
- ◆実施：反射素材からの反射のみを抽出し音響測距法で分析
IRMOS ならば有効サンプルのみでの分析が可能
- ◆結果：入射角および素材のモデル化の可能性が確認された

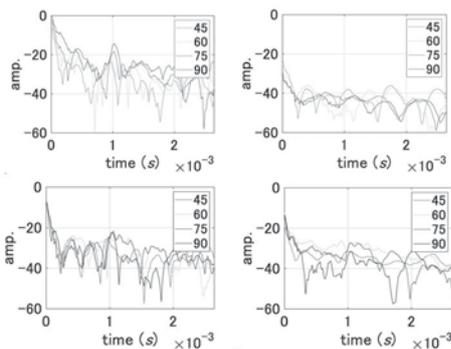


Fig. The Relation among Distance Spectrum
Upper L: Wood Board, Upper R: Glass Wool 40 (mm)

Lower L: Ref.A (interior panel), Lower R: Ref.B (limited-combustible panel)

1-P-21

1-P-21

非定常ガボール変換を用いて 打撃音劣化を低減する タイムストレッチングの実験的評価

Experimental evaluation of percussion-preserving time stretching using non-stationary Gabor transform

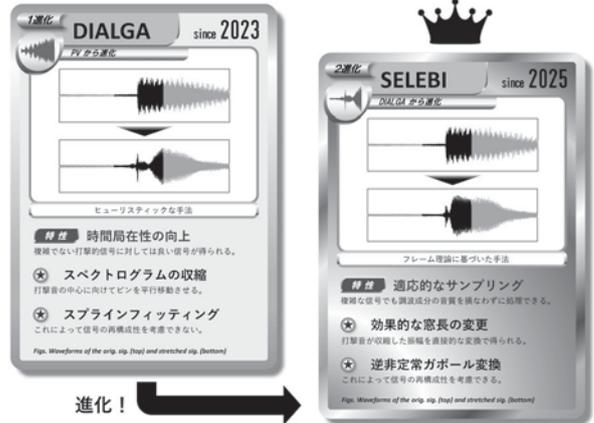
◎赤石夏輝(農工大), Nicki Holighaus (ÖAW), 矢田部浩平(農工大)

背景

位相コーダに基づくタイムストレッチングは振幅と位相の不整合が原因で打撃音が劣化してしまう

動機

振幅の処理によって不整合を軽減したい
振幅を時間方向に縮める手法 (DIALGA) を提案 (2023年)



進化!

前回

非定常ガボール変換を用いた手法 (SELEBI) を提案
打撃音をより良く保持できることを客観評価で示した

今回

客観評価を追加 & 主観評価実験で性能を評価した

1-P-23

1-P-23 発話音声の狭帯域包絡線間相関係数行列 (NCM)を用いた声質評価の一検討

Detection of Physical Condition Changes Using a Narrowband Envelope Correlation Coefficient Matrix (NCM) Derived from Speech Waveforms

☆副島 翔太(NBU), 森本 祥平 (NBU), 徳富 響(NBU),
近藤 善隆(株)ジェイテック, 福島 学(NBU), 柳川 博文(arsl)

- ◆声質評価：健康時と罹患時の声質を評価
- ◆物理変化：鼻声は副鼻腔内の共振信号の放射を妨げることに着目
- ◆検出手法：狭帯域包絡線間相関係数行列(NCM)
話者識別の特徴量を発話音声から導出するために提案
小堀川美共, 末廣一美, 高岡創, 高山泰典, 福島学, 岡本壽夫, 柳川博文, "狭帯域包絡線間相関を用いた話者識別における帯域と識別率の検討", 電子情報通信学会応用音響研究会, pp.13-18, EA2007-82, 2007
そこで、個人の罹患で生じる鼻声の特徴が得られるかを調査
- ◆結果：1) 健康(Healthy)と罹患(Disease)の識別可能性を確認
2) 疑似罹患(Feigning)との識別可能性を確認

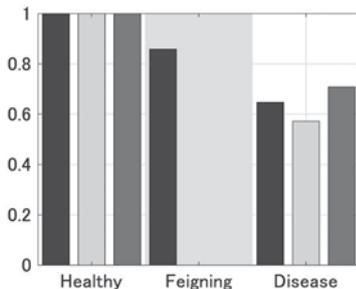


Fig. Relation among average of No.37 of NCM

State Left: Healthy, Center: Feigning, Right: Disease

■: Subject A, □: Subject B, ▨: Subject C

1-P-24

1-P-24 レーザスペckルを利用したリード振動計測の検討

Study of reed vibration measurement using laser speckle pattern

☆内川耕輔, 新美智也, 矢田部浩平(農工大)

背景

- 人の演奏時のリード振動計測手法は難しい

提案手法

- リード表面のスペckルを高速度カメラで撮影
→ 光は非接触なため人の演奏に近い状態で計測可能
スペckルパターンの変位より振動を検討

計測結果

- スペckルパターンの時間変位と音の周波数成分で8倍音までパワーのピークが対応することを確認

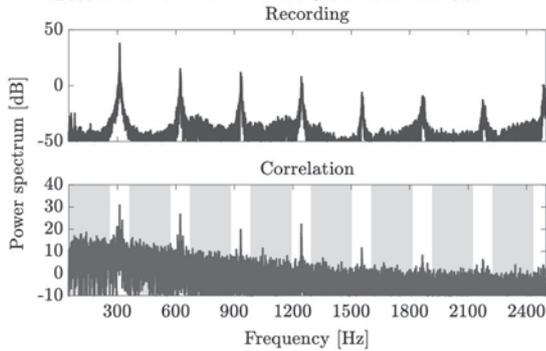


Fig. 1: Power spectrums of laser Speckle displacement

1-P-26

1-P-26 頭部伝達関数のパラメトリックモデルに基づく周波数帯域別ノッチフィルタ設計と定位性能の主観評価

Design of subband notch filters based on parametric models of head-related transfer functions and subjective evaluation of localization performance.

☆近成幸輝(大分大・工学研), 加古達也(NTT), 古家賢一(大分大・理工)

- ◆ 有限要素法による HRTF 計算コストを削減するための補間手法を提案し, 提案手法によって再構成した HRTF の定位性能について, 主観評価実験により検証した.
- ◆ KEMAR HRTF に適合した被験者 (上位 50%, 8 名) を対象とした分析の結果, 提案手法は従来手法より定位誤差が小さく, 提示角度ごとの平均回答角度についても, より高分解能 HRTF と近い定位傾向を示すことが確認された.

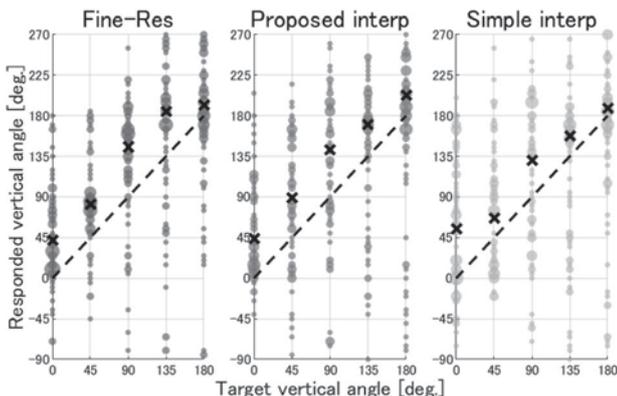


Fig. 1: Responded vertical angle in the median plane (x: Circular mean of responded vertical angles)

1-P-25

1-P-25 光計測による音場投影データを活用した高分解能音響ベクトル場推定の検討

Estimating high-spatial-resolution acoustic vector field based on sound projection data measured by acousto-optic sensing

○谷川理佐子, 石川憲治, 原田登(NTT)

◆ 背景: 粒子速度ベクトルの計測

- 粒子速度センサ, 2 マイクロホン法など
- 計測点の解像度に制約あり → 局所的な音場の評価に課題

◆ 提案手法

- 非接触音響計測が可能な光学的音響計測による音圧線積分データから平面波展開により音場・粒子速度場を高分解能に推定
- 計測条件に応じた波数ベクトルの有効範囲を定式化

◆ 結果

- 数値実験により音場・粒子速度場の推定が可能なことを確認

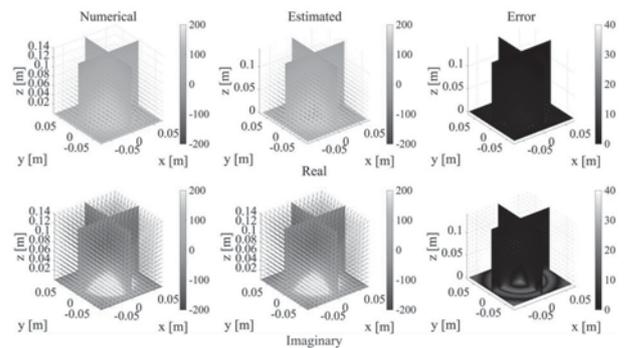


Fig. 1: Estimated results. Slices represent sound pressure and vectors represent particle velocity. (left: Numerical, middle: estimated, right: error / top: real, bottom: imaginary)

1-P-27

1-P-27 頭部運動パターンの差異が生活環境音源聴取時の高次感性評価に及ぼす影響

Effects of differences in head movement patterns on high-order perception of reality during listening to environmental sounds.

☆岡田拓真, 山高正烈 (愛知工科大院)

- ◆ 本研究では, 生活環境音を対象とし, 頭部運動の有無や運動量の違いが, 臨場感や立体感といった高次感性評価, および 3 つの音イベントの探索容易性にどのような影響を及ぼすかを検討した.
- ◆ 頭部運動条件は, 頭部運動なし (Fixed Head: FH), 頭部運動あり (Active Exploration: AE), 頭部自然条件 (Natural Listening: NL) の 3 条件を採用した.
- ◆ 臨場感評価では, Mono および Car 再生方式の場合に AE 条件の評価が低下した. 一方, 包囲感評価では FoE および FOA 再生方式の評価が向上する傾向が見られた (Fig.1).
- ◆ 頭部運動量の結果から, Mono および FOA 再生方式は, ステレオ再生方式と比較して頭部運動量が大きくなる傾向が確認された. これは, 再生方式の音場構成の違いが, 聴取時の探索行動に影響を与える可能性を示唆している (Fig.2).

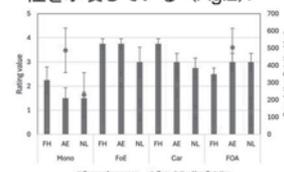


Fig. 1: Subjective ratings of sense of presence and cumulative head rotation angle.

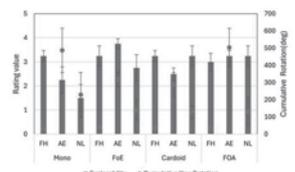


Fig. 2: Subjective ratings of explorability and cumulative head rotation angle.

1-P-28

1-P-28 オーディオARを活用したスポーツ観戦支援システムの構築と実証評価

Development and Evaluation of a Sports Spectator Support System Using Audio Augmented Reality.

○加古達也, 渡邊悠希, 千葉大将, 野口賢一(NTT)
中村風香, 飯田一博(千葉工大)

- ◆ 音響透過性の高いオープンイヤーイヤホンと PNP HRTF を用いたオーディオAR スポーツ観戦支援システムを構築
 - スタジアムにてオープンイヤーイヤホンを装着し, 臨場感を損なわずに実況解説の音声を後方に定位させスタジアム内に配信
 - 複数のPNP HRTF モデルから自身に適合するモデルが選択可能
 - 実際のサッカーの試合において観客45名に対し, スマートフォン, オープンイヤーイヤホンを貸し出し実証実験
- ◆ 実証実験による主観評価結果と有効性の検証
 - 歓声と解説の両立を64%の参加者が高く評価(10段階中7以上)
 - PNP HRTFにより意図した音像定位が実現され, 臨場感を損なわない観戦体験を確認

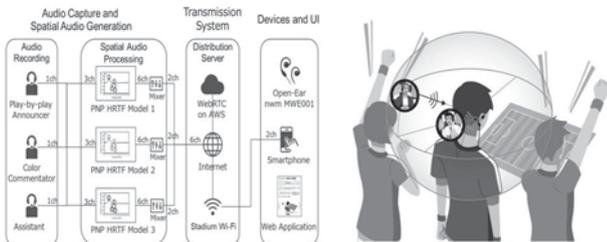


Fig.1 Configuration of the Audio AR system.



Fig.2 Play-by-play and commentary are presented from the rear.

1-P-30

1-P-30 球面スピーカアレイを用いた羽ばたき動作に対する仮想羽音再生について

Spatial reproduction of virtual wing sounds synchronized with human flapping motion using a spherical loudspeaker array

☆藤原那七星, 羽田陽一(電通大)

- ◆ 羽ばたき動作として腕を前後に開閉する際, 動作に同期して身体付近を移動する仮想的な翼の音像を提示した。
- ◆ リアルタイムで移動する音像の提示には, 球面スピーカアレイとモーションキャプチャ, IIR フィルタによる NFC-HOA を用いた。
- ◆ 背中から翼が生える場合と, 腕を翼とする場合の提示方法を検討した。
- ◆ 再生音は, 風切り音や動作音としてのノイズ音と羽音で構成した。
- ◆ 異なる再生手法, 再生音の条件間で翼の所有感を比較した。
- ◆ 結果より, 羽ばたき動作に伴って既存の腕の位置に仮想的な翼の音像を提示するだけで, 腕が翼に変化する感覚が生じることが示唆された。

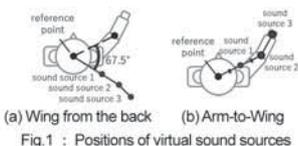


Fig. 1 : Positions of virtual sound sources

Table 1 : Presentation conditions

Condition	Reproduction method	Reproduced sound
NFC_NF	NFC-HOA	Noise+Flapping
NFC_N	NFC-HOA	Noise
HOA_NF	HOA	Noise+Flapping
Stereo_NF	Stereo	Noise+Flapping

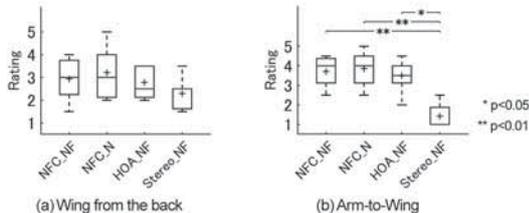


Fig.2 : Subjective ratings of the sense of wing ownership

1-P-29

1-P-29 球面スピーカアレイを用いた残響場再現による音像定位と距離感提示に関する検討

A study on sound image localization and distance perception based on reverberant field reproduction using a spherical loudspeaker array

☆赤井たろ生, 羽田陽一(電通大)

- ◆ 球面マイクロホンアレイで収録した室内インパルス応答(RIR)に基づく HOA 音場再現では環境雑音や測定誤差の影響により, 意図しない残響感が発生し, 音像定位に影響を及ぼすことが懸念される。
- ◆ 測定データに基づく音場合成として NFC-HOA(Data-based)と聴覚的なパラメータに基づく HO-SIRR を利用する。これらの手法と測定を必要としないモデルベースの残響場合成(Model-based)による残響場合成の音像定位性能および空間品質の再現性にどのような違いが生じるかを比較するため VR ヘッドマウントディスプレイと Fig.1 に示す球面スピーカアレイを用いた主観評価実験を行った。
- ◆ モデルベースの残響場合成は測定ベースの 2 手法と比較して知覚距離は表示距離よりも過小評価される傾向があったが, 方向定位について高い精度を示した。一方で高さ方向の定位精度については測定ベースとモデルベースのいずれでも誤差がみられた。



Fig.1: 42ch spherical loudspeaker array.

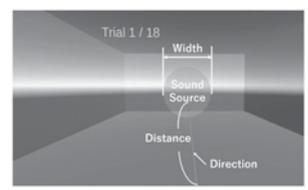


Fig.2: Virtual environment for evaluation.

1-P-31

1-P-31 マルチスピーカ音響再生における直接音と残響音との再生位置の違いが聴こえに及ぼす影響とその要因について

The Effect of Reproduced Positions between Direct and Reverberant Sound in Multichannel Sound Reproduction on Impression

☆真武永遠, 脇田由実(大阪工大),

入交英雄(入交イマーシブオーディオ研究所)

- ◆ 直接音と残響音との最適なミキシングレベルが, 各再生スピーカ角度によりどう変化するかを調べるため, 無響室で録音された曲に残響音を加えたものを基準とし, それと同じ印象に聞こえる残響ミキシングレベルを, 残響音スピーカの位置を変えて測定した。
- ◆ 聴取実験の結果, 角度が広がるほどミキシングレベルが減少する傾向にあった。多重比較検定によると, スピーカ間角度が0度と, ± 30 度・ ± 60 度・ ± 90 度の間には有意差があったが, それ以外の組み合わせは有意差がなかった。(Fig.1)
- ◆ 一方, 音の到来方向とラウドネスとの関係を調べた結果, 角度が0→30→60→90度と正面から離れる毎に, 音量調整値が減少した。この現象は上記ミキシングレベルが減少する要因の一つと説明できる現象であったが, Fig.1に見られる0→30度に至る急激な減少は見られなかった。ミキシングレベルの減少要因が, 上記ラウドネスの現象だけでなく, 他要因(両耳聴による方向性マスキング解除など)も関係していると考えられる。

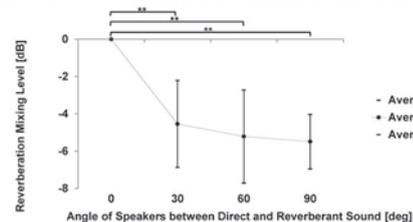


Fig. 1 The Relation between Speaker angle and Reverberation Mixing Level.

1-P-32

1-P-32 AR環境を用いた音像定位評価システム 開発:音像定位精度の検証

Development of a Sound Localization Evaluation System Using AR Environments. Verification of Sound Localization Accuracy

☆花城 清月,ヌルファキーハ ピンティルスリ,浅川 香(東京電機大学システムデザイン工学部),村山 好孝(シーイヤー(株)),渡邊 祐子(東京電機大学システムデザイン工学部)

- ◆AR環境を用いた音像定位評価システムを開発した
- ◆従来の図示型回答方法との比較実験を通して、音像定位精度の観点からAR手法の有効性について検討した。
- ◆方位角においては提示角度と回答角度が概ね一致した。仰角に関しては同様の傾向が確認された。
- ◆回答距離については、従来法に比較してAR法は、ばらつきは大きいものの奥行方向の回答が提示距離に近づくことが分かった。

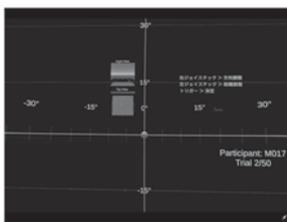


図1 AR空間内の球座標系

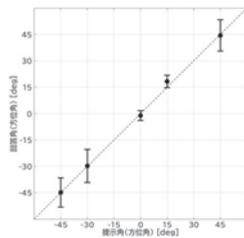


図2 AR法の回答結果(方位角)

1-P-36

1-P-36 超音波トランスデューサにおける 傾斜音響インピーダンス整合層の検討

Simulation on graded index matching layer for ultrasonic transducers

☆Li Yuxing, 和田有司, 中村健太郎(東京科学大)

- ◆非破壊検査や医用診断のためのMHz帯超音波トランスデューサについて、音響インピーダンスが連続的に変化する整合層の特性を数値計算で検討した。
- ◆傾斜音響インピーダンス材料(GIM)の音響エネルギーを最適化するインピーダンス分布を求めた。
- ◆圧電素子(PZT-5H)、GIM および水から構成されるモデルを対象として、有限要素法により広帯域透過特性を検討した。
- ◆指数関数分布は他の分布形状より良好な透過特性を示し、特にGIM厚さ0.4mm以上で3~10MHzにおける透過率が90%を超えた。
- ◆実際の作製を考慮し、離散型モデル(5層)について等厚分割法(CT)と等インピーダンス差分割法(CI)を検討した。
- ◆CTのほうがCIよりも高い透過率特性を示した。

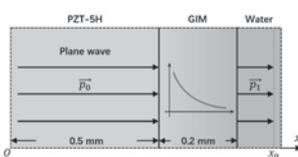


Fig.1: Calculation model.

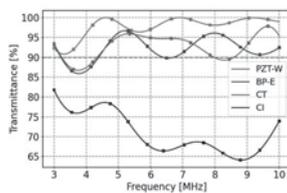


Fig.2: Transmittance for continuous and stepwise acoustic impedance profiles.

1-P-35

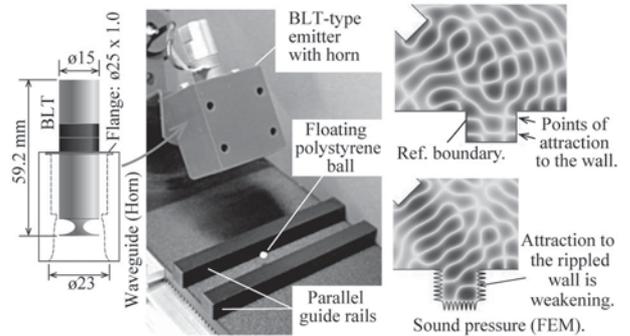
1-P-35 空中超音波の波長よりも小さな構造を 定在波の反射壁に用いた作用について

The effect of using structures smaller than the wavelength of aerial ultrasound as reflecting walls for standing waves

○田村 英樹(東北工大), 笠島 崇, 伊藤 伸介, 辻井 明日香(日本特殊陶業), 青柳 学(室蘭工大)

強力空中超音波を用いた非接触浮揚の一方法として、下図のように並行レール間やコの字形空隙に対して放射した音波が狭空間で定在波を形成してそのノード部でトラップさせることが出来る。この場合には音源位置を物理的に移動することで、レールに沿った浮揚物体の移動が可能である。しかしながら、移送時や外乱によって浮揚物体位置が揺動した際にノードに沿うように移動して落下等を生じる事がある。このうち壁面に接して垂直に伸びるノード線は、揺動した浮揚物体を壁面に吸い寄せるとして作用して接触状態に至り易い。

ここで反射壁面に音波の空間周期よりも小さな周期構造を設ける事で壁面部に低音圧・ノード層を生じさせることが出来る。これにより超音波浮揚器を構成する際に望ましくない壁面への被浮揚体の吸い寄せ作用を低減する効果に関してFEM解析結果を用いた考察を示す。



1-P-37

1-P-37 野生コウモリの多点計測へ向けた独立分散 型マイクロホンアレイシステムの開発

A Distributed Microphone Array System for Multi-Point Measurement of Wild Bats

☆明石悠佑, 山田恭史(はこだて未来大・シス情)

- ◆コウモリはアクティブセンシングにより超音波パルスを放つことから、マイクロホンアレイを用いることで非接触ながらも行動計測を行えることが先行研究で示された。
- ◆しかし、マイクロホンアレイと処理用PCの間に物理的な接続が必要であることから、計測範囲に制限が生まれ、処理用PCにも集中的な負荷がかかってしまう。
- ◆本研究では、マイクロホンアレイと解析用PCを一体化した計測デバイスを製作することで、物理的な制限を払拭し、処理負荷を分散させた長時間計測を可能とする手法を提案する。
- ◆実計測により、一台での長時間計測を可能とし、計測されたデータをもとに解析をすると、コウモリの座標推定に用いる時間差が確認され、座標推定に有意であることが判明した。



Fig. 1 一体型デバイス

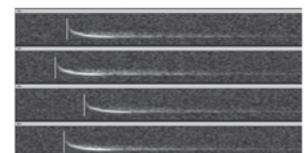


Fig. 2 計測された計測時間差

1-P-38

1-P-38 低周波パラメトリック超音波 CT に向けた伝搬時間の推定精度改善

Improving the Estimation Accuracy of Propagation Time for Low-Frequency Parametric Ultrasound CT

☆西岡珠希, △朱沛賢, 野村英之(電通大)

- ◆本研究では、集束ビームを用いたパラメトリック差音に加え、圧縮パラメトリック信号を用いた差音を使用することで、低周波パラメトリック超音波 CT(Computed Tomography)における伝搬時間推定精度の改善を目的とする。
- ◆平面音源および集束音源を用いた、圧縮パラメトリック差音による伝搬時間(Time of Flight, TOF)の結果を Fig. 1 に示す。
- ◆集束ビームを用いた圧縮パラメトリック差音では平面音源に比べて伝搬時間推定の精度が改善された。これらの結果は、低周波パラメトリック超音波 CT において、TOF 推定精度の向上が再構成画像の画質向上につながる可能性を示唆している。
- ◆現在利用している CT のアルゴリズムは直進性を仮定しているため、ビーム幅を持つ超音波では限界がある。そのためビーム幅を考慮した再構成方法の検討の必要がある。

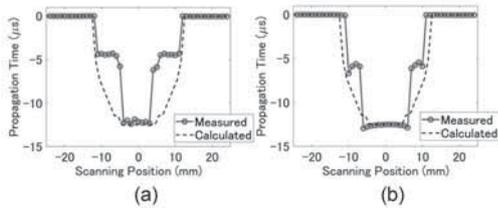


Fig.1: Experimental results of time-of-flight measurements, (a) plane (b) focus

1-P-40

1-P-40 超音波準ベッセルビームの自己回復性の検証

Verification of self-healing property of ultrasonic quasi-Bessel beam

☆齋藤順之介, △朱沛賢, 野村英之(電通大)

- ◆生活環境の改善やVR 体験を目的として、超音波ベッセルビームを用いた気流制御が提案されている。ベッセルビームは自己回復性を持ち、障害物に遮られても形状を保って回復する効果がある。
- ◆本研究では、超音波素子の配置制御により実現される、非軸対称な準ベッセルビームについて自己回復性を検証した。
- ◆理想的なベッセルビームのシミュレーションと素子配置制御によって実装した準ベッセルビームの音場測定実験を行い、音場を比較した。障害物として、メインローブの幅と同程度および2 倍の直径の球体を音軸上 0.4 m の位置に設置した。
- ◆音軸方向の音圧分布から、球体後方で音圧が回復しており、シミュレーションに近い傾向が確認された(Fig. 1 (a))。
- ◆方位方向の音圧分布から、音圧が回復した位置においてメインローブの形状も元のビームと同様に回復することがわかった(Fig. 1(b))。

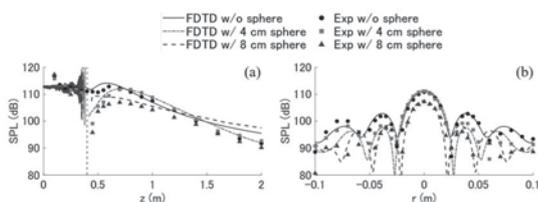


Fig.1: Sound pressure distribution of FDTD simulations and experiments, (a) Axial, (b) Horizontal at z = 0.8 m.

1-P-39

1-P-39 非接触加振における単一送受波時間反転法の再集束特性評価

Evaluation of refocusing characteristics in single transmit-receive time reversal method for non-contact vibration excitation

☆北嶋晃希, △朱沛賢, 野村英之(電通大)

- ◆超音波非破壊検査において時間反転法は有効な手法であるが、音源接触による散乱の影響が懸念される。本研究は非接触式加振による時間反転法の基礎検討として、単一受信点条件での再集束特性を評価した。
- ◆集束音源による非接触式加振および圧電素子による接触式加振を用い、空気中のアルミニウム板上で時間反転実験を行った。周波数 40 kHz の順方向伝搬時信号を 1 点で受信し、時間反転して再放射した際の振動速度をレーザードップラー振動計で測定した。
- ◆時間反転信号の空間集束特性を Fig.1 に示す。接触式では再集束位置以外に副次ピークが観測されたのに対し、非接触式では再集束点のみピークが確認された。これより、非接触式加振は副次ピーク抑制の観点で有効である可能性が示唆された。

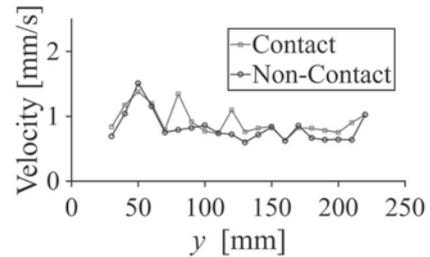


Fig. 1: One-dimensional spatial distribution of time-reversal focusing. The intended refocusing point (excitation point) is at y=50 mm.

1-P-41

1-P-41 ステンレス球殻上のレーザー誘起 Lamb 波の特性評価

Characterization of laser-induced Lamb waves on stainless-steel spherical shell

○會澤康治, △千代泰旗, △藤田大翔, △内山雄太, 西田蒼(金沢工大)

- ◆我々は球殻上のレーザー誘起 Lamb 波の調査を通して、これを利用した低侵襲での経皮薬剤送達の可能性を検討している。
- ◆これまでに行なってきた実験では球殻材料にポリプロピレン(PP)を使用したが、シミュレーションによれば金属球殻上の Lamb 波の振幅は PP 球殻よりも大きい。
- ◆本研究ではこのシミュレーション結果を PP 球殻との比較実験で確かめた。
- ◆実験には材質がステンレス(SUS304)の球殻(外径 41 mm、殻の厚さ 0.5 mm)と比較用に PP 球殻(外径 40 mm、殻の厚さ 1 mm)を用意し、球殻表面に接着したレーザー吸収材に向けてナノ秒パルスレーザー(波長 532 nm、パルス幅 10 ns)を単発照射して発生したガイド波を球殻表面から 2 mm 離れた空気結合トランスデューサで測定した。
- ◆観測されたガイド波は Lamb 波 A0 モードと衝撃波であった。
- ◆Lamb 波 A0 モードに着目すると、ステンレス球殻の方が PP 球殻よりも測定波形の最大振幅は 2 倍以上大きかった。また帯域通過フィルタ付外部プリアンプを通した測定では、100 kHz 以上でステンレス球殻の方が PP 球殻よりも最大振幅が大きく、40 kHz 以下では小さかった。

1-P-42

1-P-42 ボルネオ熱帯雨林における環境音からの生物種多様性に関わる音響指標の算出

Derivation of acoustic indices related to biodiversity from environmental sounds recorded in the tropical rainforest in Borneo

○西村 明(東京情報大)、△松林 尚志(東京農大)、△Arthur Chung (ボルネオ森林局)

熱帯雨林地帯において土地環境が異なり生物多様性も異なる保護林3地域で行った、前報2024年8月の録音と、新たな2025年3月録音の複数帯域から、前報と同様に各種音響指標値算の日内変動を算出した。そして、同じ地域内の異なる録音地点間および異なる観測日間で似た日内変動を示し、かつ異なる地域間では異なる日内変動を示す音響指標とその帯域、および日内変動の特徴、録音時期の違いを調べた。

異なる観測日間で相関が高かった音響指標の日内変動ベクトルについて、その特徴を捉えるため主成分分析を行った。その結果、200 Hz ~ 96 kHz 帯域から算出した生物音響指標(BAI)が、観測日が異なっても似た日内変化を示し、かつ同じ地域が主成分負荷マップ上で近くに布置され、地域による音環境の違いをよく表現できた。また、録音時期の違いによって異なる日内変動を示す地域と、ほぼ同様な日内変動を示す地域があった。

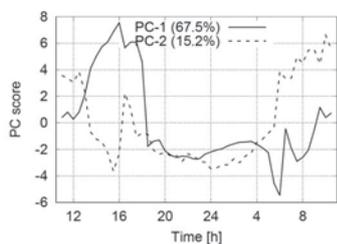


Fig.1 Factor scores of the 1st and 2nd principal components of the diurnal variation of BAI extracted from 200 Hz - 96 kHz.

1-P-44

1-P-44 FEUDT 弾性表面波素子による液中出力特性の検討

Study on the output characteristics of FEUDT transducer in liquid

◎孔徳卿, 田中智也, 佐々木実(豊田工)

- ◆浮き電極型一方向性変換器 (Floating Electrode Unidirectional Interdigital Transducer, FEUDT) という一方向性 SAW デバイスを利用し、液中に浸漬したデバイスから放射された縦波の一方向性を検討した。
- ◆短絡型と開放型浮き電極を利用し、周期長(波長) 200 μm の FEUDT を設計した。COMSOL 固有周波数解析により FEUDT の振動モードを検討し、Fig. 1 に示すように、フォトリソグラフィ技術を利用し、10 対の FEUDT を用いた弾性表面波素子を製作した。液中の絶縁性を考慮し、デバイス表面に SiO₂ 膜を成膜した。
- ◆Fig. 2 を示すように、駆動電圧を 52 V_{pp} とした場合の流速分布を示した。トレーサ粒子の最大速度 54.1 mm/s を観測することができた。液中に非対称性の放射波を確認することができた。FEUDT の両側の最大流速により、液中に放射されたパワーは 86%の一方向性を実現した。

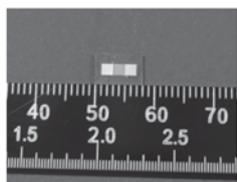


Fig. 1 Photo of FEUDT device.

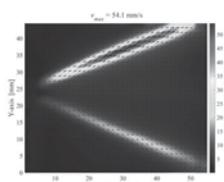


Fig. 2 Flow velocity with PIV.

1-P-43

1-P-43 多孔質セラミック材料を用いた熱駆動型広帯域超音波センサーの開発—出力音圧向上に関する検討—

Development of a Thermoacoustic Broadband Ultrasonic Sensor Using Porous Ceramic Materials: An Investigation into Output Sound Pressure Enhancement

◎布川貴康, △長谷川勘太(同志社大)

浅田隆昭(同志社大・音響ナビゲーション研究センター)

高柳真司, 飛龍志津子(同志社大)

- ◆従来の共振型送信器は高周波で平坦な周波数特性を保つのが難しく、小型・高音圧・広帯域を同時に満たせない。一方、サーモホンは熱音響変換により広帯域応答が得られ、材料設計で音圧向上が期待される。
- ◆本研究では、コウモリの広帯域超音波を模倣する新型サーモホンを開発し、音圧特性向上と超音波センシング応用の可能性を評価した。
- ◆多孔質セラミック断熱層を用いたサーモホンは、広帯域で高音圧(最大 79.6 dB SPL)の超音波放射を可能にし、2.5 m 先の反射も検知可能で、材料特性が音圧向上に寄与することが示された。

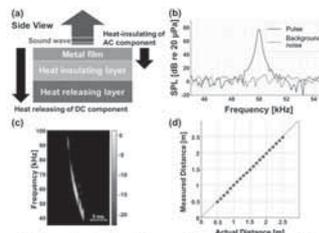


Fig.1: Structure and the principle of sound generation of thermophone (a), spectrum of a porous silica sample driven by a 50-kHz, 10-ms burst signal (b), the spectrogram of bat-mimicking FM signals emitted from the thermophone (c), and the results demonstrating the detection of echoes from obstacles at a distance of 2.5 m (d).

1-P-45

1-P-45 コウモリのバーチャルエコー空間構築のための3次元音響シミュレーション手法の開発と評価

Development and evaluation of a three-dimensional acoustic simulation method for constructing a virtual echo space of bats.

☆前川隼人, △沖見正広, △矢尾遼太, △長谷川勘太, 飛龍志津子, 土屋隆生(同志社大)

- ◆コウモリに届くエコーは微弱であるため、コウモリがエコーロケーションで取得する音響情報の全てを実空間で実測することは、従来の方法では技術的に不可能である。
- ◆そこで本研究では、FDTD 法を用いた3次元音響シミュレーションを開発し、飛行中のコウモリに到達するエコーを再現した。
- ◆その結果、2次元シミュレーションでは再現が困難であった天井や床の反射や壁のコーナーからのエコーが、実測音声に近い精度で再現されることを確認した (Fig.1)。
- ◆本手法により、飛行中のコウモリが知覚するエコー空間をバーチャルに可視化でき、行動戦略のより高度な分析に有用であると考えられる。

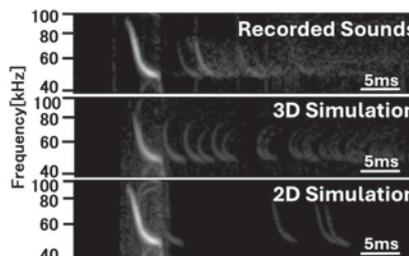


Fig.1: Spectrogram Comparison of echoes in recorded sounds and 2D and 3D simulations.

1-P-46

1-P-46 不等間隔配置による空中超音波アレイの音波集束特性の基礎検討

Fundamental Study on Acoustic Focusing Characteristics of an Airborne Ultrasonic Array with Non-uniform Element Spacing

☆吉澤拓翼(日大院・理工), 清水鏡介(愛媛大院),
伊藤洋一, 大隅歩(日大・理工)

- ◆ハルトン列に基づく不等間隔 64 素子空中超音波アレイ (開口 240 mm) を試作した。
- ◆集束距離 150 mm において, 集束点を含む x-y 平面 (± 30 mm) を走査し, 音圧が最大となる時刻帯の瞬時音圧分布を用いて音波集束特性を実験的に評価した。
- ◆Fig.1 に作成した空中超音波アレイの配置パターン図, Fig.2 集束域で音圧が最も高い時刻帯における瞬時音圧分布を示す。
- ◆結果より, 集束域の音圧分布から評価した半幅幅は約 8 mm であり, 波長と同程度の集束径が得られた。したがって, 本試作アレイは回折限界近傍まで集束幅を狭められていることを確認した。

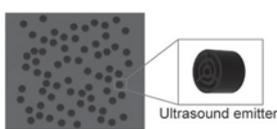


Fig.1 Schematic view of an airborne ultrasound array with non-uniform element spacing.

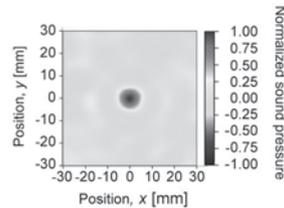


Fig.2 Sound pressure distribution.

1-P-48

1-P-48 ステップホーン構造の 40 kHz 用小型ボルト締めランジュバン型振動子の開発

Development of a compact bolt-clamped Langevin transducer for 40 kHz with a step horn structure

☆西川博基, 三浦 光, 浅見拓哉(日大・理工)

- ◆本稿では, 筆者らが提案している 40 kHz 用小型振動子の縦振動変位拡大構造について検討し, その基礎特性について報告する。
- ◆Fig. 1 は振動子の概略図である。振動子は放射面が矩形となるステップホーン構造である。本振動子の振動特性を測定した結果として, 振動子の放射面(z 軸上の 0 mm の位置)及びネジ側端面(z 軸上の 68 mm の位置)はほぼ一様に縦振動しており, 縦振動は 2.2 倍程度拡大していることが確認された。一方, 振動子側面のたわみ振動変位分布は振動子の薄い部分は大きくたわみ振動する結果となった。

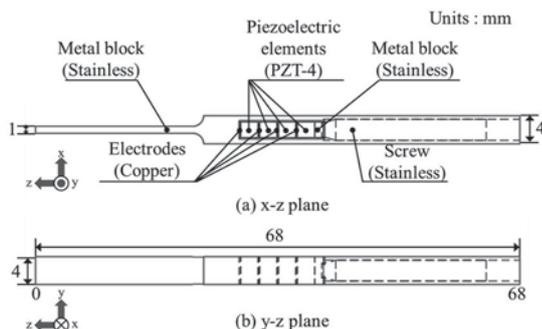


Fig. 1 Schematic of bolt-clamped Langevin type transducer

1-P-47

1-P-47 熱音響システム発電システムについての基礎検討—開管枝付きループ管型熱音響プライムバーにおける PA の設置検討—

Basic study on thermoacoustic power generation system—Study on PA placement in open-ended branch loop-type thermoacoustic primemover

☆吉田康汰朗, 坂本真一, △平山智士(滋賀県立大学大学院・工学研究科)

熱音響システムは, 未利用の熱エネルギーを有効活用できる技術として期待されている。熱音響システムは, 熱と音の相互変換である熱音響現象を利用したシステムである。この熱音響システムの中でも, 熱エネルギーから音エネルギーへと変換するものを熱音響プライムバーと呼び発電への応用が期待されている。しかし, 熱音響発電には多くの課題がある。その一つに共振温度の高さが挙げられる。現在, 排熱として多くの割合を占める熱は 300 °C 以下の熱となっている。そのため, 本研究では開管枝付きループ管型熱音響プライムバーに PA(Phase adjuster)を用いることで共振温度の低下を目指した。Fig. 1 に PA の設置位置に対する共振温度の変化を示す。Fig. 1 より設置位置 400 mm 地点で共振温度が 552 K となり 300 °C 以下の共振が可能となった。

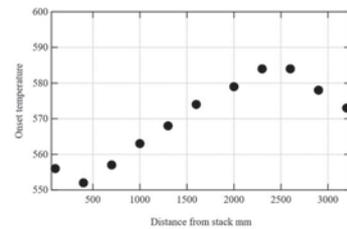


Fig. 1 Results of stability analysis.

1-P-49

1-P-49 200 kHz で駆動させた積層圧電アクチュエータから放射される音波について

Acoustic waves emitted from a multilayer piezoelectric actuator driven at 200kHz

☆渡邊太陽, 三浦 光, 浅見拓哉(日大・理工)

- ◆本報告では非接触非破壊検査の高分解能化を図ることを目的とし, 積層圧電アクチュエータを 200 kHz で駆動させ, それにより放射された音波についての検討を行った。
- ◆過去の検討にて, アクチュエータを駆動させる際, インピーダンスが非常に低いことから LC 直列回路を並列に挿入した。これによりインピーダンスを上昇させ駆動効率を上昇させた。本発表では, LC 直列回路の挿入による放射音波の変化について検討を行った。
- ◆Fig. 1 はアクチュエータから放射された音波の音圧波形である。黒色線がアクチュエータのみでの駆動, 灰色線が LC 直列回路を挿入した状態での駆動の音圧波形である。音圧波形は, LC 回路の有無により, 立ち上がり, 立ち下がり, 及び定常状態に変化がみられた。

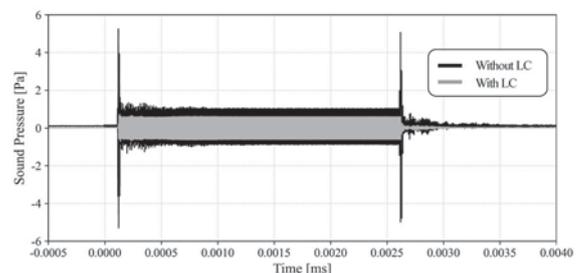


Fig.1: Acoustic waves emitted from a multilayer piezoelectric actuator

1-P-50

1-P-50 水熱合成法によるKNbO₃膜を用いたフィルム状圧電振動子の検討

Film-Type Piezoelectric Transducer Using a KNbO₃ Film Deposited via Hydrothermal Synthesis.

☆藤井利津¹, 呉沢農¹, 原田祐生², 石河睦生³, 小山大介¹

(¹同志社大, ²山梨大, ³桐蔭横浜大)

- ◆水熱合成法によって作成したKNbO₃膜の膜厚と結晶配向を評価した。
- ◆膜厚60 μmのエピタキシャル成長した薄膜の合成に成功し、これを用いた超音波トランスデューサを作製した。
- ◆トランスデューサの厚み縦振動の共振周波数は50.3 MHzであった。
- ◆高周波超音波とレーザーによる音響光学効果の実験を行った。脱気水で満たしたセル内にトランスデューサを設置し、トランスデューサ先端近傍に音軸に対して垂直にレーザーを入射した。
- ◆超音波照射の有無によるレーザー透過光分布を測定したところ変化が観測された(Fig. 1)。超音波により媒質中に屈折率変調が形成され、音響光学効果に基づく光の回折現象が生じていると考えられる。

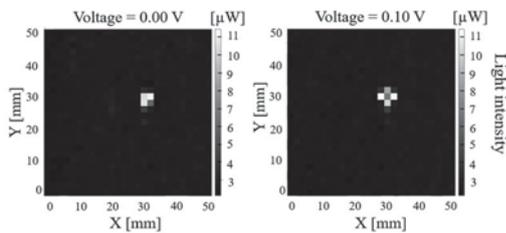


Fig.1: Light intensity distributions without and with ultrasound radiation.

1-Q-1

1-Q-1 マイクロスリット型音響メタマテリアルの輸送パラメータに基づく吸音特性予測モデルの構築

Development of an absorption prediction model for micro-slit acoustic metamaterials based on transport parameters

○菅原彬子, △岸本康生(近畿大), △城修平(大林組), 平栗靖浩(近畿大)

- ◆単純立方格子にマイクロスリットを組み込んだ音響メタマテリアル Microslit-AMM のスリット幾何形状と吸音特性の関係を整理した。
- ◆スリット幾何形状と JCAPL モデルの輸送パラメータとの関係を表す回帰モデルを構築し、数値解析および実験結果との比較により吸音特性の推定精度を検証した。
- ◆提案した回帰モデルはスリット幾何形状に対する輸送パラメータの依存性を適切にとらえており、推定された吸音率は数値解析・実験結果の双方と良好に一致した。

Table 1: Regression models

Regression model
$\alpha_{\infty, E} = 0.394 \left(\frac{1-P}{P}\right)^{0.780} \psi_s^{-0.089}$
$\alpha_{0v, E} = 0.224 \left(\frac{1-P}{P}\right)^{1.023} \psi_s^{0.010}$
$K_{0, E} = 2361 \left(\frac{1-P}{P}\right)^{-1.683} \psi_s^{-0.852}$
$\Lambda_E = 4.568 \left(\frac{1-P}{P}\right)^{-0.542} \psi_s^{-0.408}$
$\Lambda'_E = 6.202 \left(\frac{1-P}{P}\right)^{0.025} \psi_s^{-0.051}$

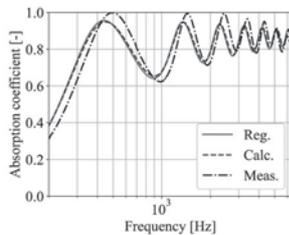


Fig.1: Sound absorption coefficient at normal incidence

1-P-51

1-P-51 空中非線形超音波を利用した金属板内複数欠陥の振幅イメージング

Amplitude imaging of multiple defects in metal plate using air borne nonlinear ultrasound

☆五味健太郎(日大・理工), 石川周男(日大院・理工), 清水鏡介(愛媛大院), 伊藤洋一, 大隅歩(日大・理工)

- ◆金属薄板内に複数の減肉欠陥が存在する場合の可視化および識別性について検討している。
- ◆本報告では、強力空中超音波の非線形性を利用し、振幅ピーク分布に基づくイメージングを2つの減肉欠陥に適用した。
- ◆その結果、高調波の次数を高くすることで1.25mm程度の間隔で隣接した欠陥を個々に分離して可視化可能であることを示した。

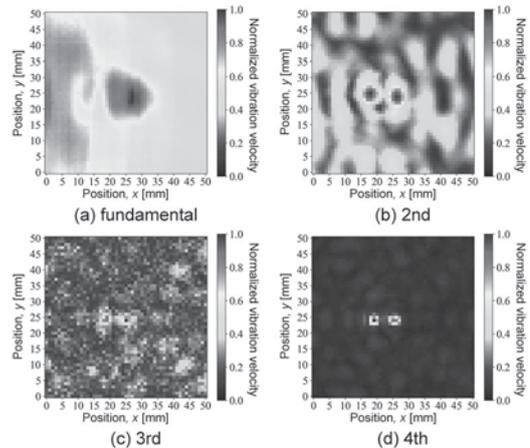


Fig.1 Amplitude peak map.

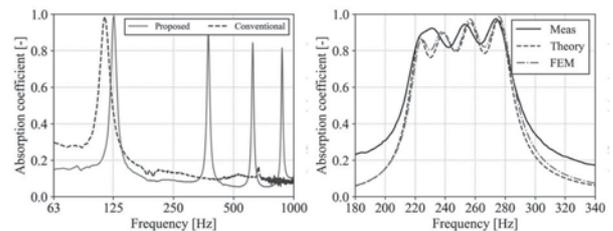
1-Q-2

1-Q-2 矩形スペースコイル型音響メタサーフェスによる1/3オクターブバンド吸音体*

One-third-octave band sound absorber by rectangular-shaped space-coiling acoustic metasurface

☆森口尚寿, 奥國健(神戸大院・工学研)

- ◆スペースコイル型音響メタサーフェスは、低周波数吸音と製造の容易さを両立する実用性の高い構造として期待されている。
- ◆しかし、従来構造は低周波数の吸音において、幾何学的な制約から所望の吸音特性実現のための設計空間が制限される懸念がある。
- ◆本研究では、コイル状空気層の幾何構造を工夫し、設計自由度を高めた矩形スペースコイル型ユニットセル構造および、それを用いた結合共振器型広帯域吸音構造を提案する。
- ◆125 Hz を対象に波長の1/94厚さで準完全吸音を実現できること、総厚36mmという薄さで中心周波数250Hzの1/3オクターブ幅において平均吸音率0.89の吸音を達成できること、を実証した。



Performance of proposed rectangular-shaped space coiling AMS: (a) unit cell structure for 125 Hz absorption, (b) coupled-resonator for 1/3 octave band absorption

1-Q-3

1-Q-3 ヘルムホルツ共鳴と板振動の連成効果を利用した樹脂成形の吸音構造の開発

Development of a blow-molded sound absorber utilizing a coupling effect of Helmholtz-resonance and plate vibration

◎藤本望夢, 眞田明(岡山県工技セ), 高田卓, △坪井裕之(みのり化成)

- ◆孔あき表面板の振動を活用することで、共鳴に加えて板振動による吸音効果を付与した吸音構造とその設計手法を提案している。
- ◆孔あき吸音板と空洞部からなる単純構造であるため、成形金型を用いた樹脂加工による低コストかつ高い生産効率での製造が可能である。
- ◆共鳴と板振動の連成効果を利用した吸音機構であり、ダイナミックダンパの最適設計法に基づき、共鳴器と弾性板の共振周波数比と弾性板の損失係数を設定することで、広帯域な吸音特性を実現した (Fig. 1)。
- ◆本研究では、一般的な樹脂成形技術であるブロー成形により共鳴と板振動の連成効果を利用した吸音構造を試作した (Fig. 2)。
- ◆垂直入射吸音率の測定結果から、樹脂成形製の吸音構造においても連成効果が生じており、広帯域な吸音特性が得られることを確認した。

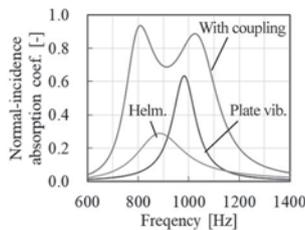


Fig. 1 : Calculated normal-incidence absorption coefficient

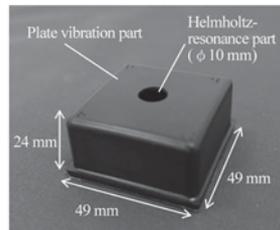


Fig. 2 : Photograph of the blow-molded sound absorber

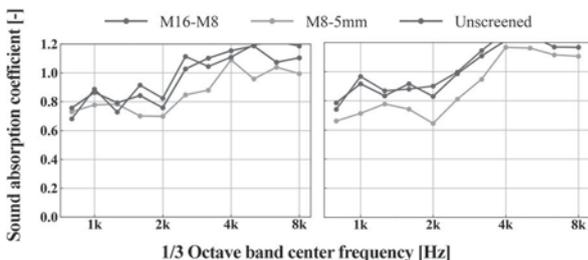
1-Q-5

1-Q-5 製紙廃材の有効利用に向けたペーパースラッジの残響室法吸音率測定

Reverberation-Room Measurement of the Sound Absorption Coefficient of Paper Sludge for the effective utilization.

☆江田博紀, 菅原彬子 (近畿大), △森岡正行 (ダイワペーパーネクスト), △岡田大樹, △北村真則 (大和板紙)

- ◆ペーパースラッジ (PS) は製紙工程の排水処理過程で生じる微細セルローズ繊維を含む有機性汚泥である。
- ◆大部分は焼却・埋立処分され、最終処分場の逼迫や処理コストの増大が課題である。このため、PSの新たな活用方法が求められている。
- ◆本研究では残響室法吸音率を測定し、粒径条件および形状保持を目的とした成形・固定方法が吸音特性に及ぼす影響を比較検討した。
- ◆その結果、粒径条件により吸音性能に差がみられたものの、全体として中高周波数帯域で良好な吸音性能を示し、接着剤および不織布による成形・固定方法による影響は限定的であった。



(a) Normal (b) Bond-NW

Fig.1: Sound absorption coefficients

1-Q-4

1-Q-4 ヘルムホルツ共鳴器を用いた音響メタサーフェスの広帯域・準完全吸音の簡易設計法

Simplified Design of Broadband Near-Perfect Absorbing Acoustic Metasurfaces Based on Helmholtz Resonators

○中西伸介(広島工大・工)

- ◆背景・目的: 低周波域用吸音構造の薄型化に向け、ネックを折り畳んだヘルムホルツ共鳴器による音響メタサーフェス (AMS) が提案されている。本報では、目標特性から AMS の構造諸元を閉形式で直接算出可能な「簡易設計法」を提示する。
- ◆簡易設計アルゴリズム: 以下の5段階で構成される。
 1. 目標吸音帯域および並列配置するユニットセル数の決定。
 2. 各セルの共鳴周波数を目標帯域内で適切に配置。
 3. 共鳴周波数および隣接セルの特性交差周波数における目標アドミタンス比の設定。
 4. 閉形式の数式を用いた各セルの幾何学的諸元の算出。
 5. 得られた諸元に基づく全体特性の検証。
- ◆結果・考察:
 - 9 個の共鳴器を並列配置した AMS の解析により、125, 250, 500 Hz の各中心帯域で準完全吸音 ($\alpha \geq 0.8$) を確認した。
 - 3D プリント製試験体の実測値は計算値と乖離が見られたが、傾向は既報と同等であり本設計法の有用性が示された。
 - 乖離要因は理論モデルと実形状の差によるモデル化誤差や造形精度に起因する。
- ◆結論: AMS の仕様を迅速に決定可能な簡易設計法の妥当性が実証された。

1-Q-6

1-Q-6 Physics-informed Neural Operator を用いた 2次元音響シミュレーション

Two-dimensional acoustic simulation using physics-informed neural operator

☆坂爪孝照, △三好孝典, △小倉正崇, 横田和哉 (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技術として、Physics-informed Neural Operator (PINO) に関する研究が近年活発化している。
- ◆PINO は、ニューラルネットワークに物理法則を組み込むことで、支配方程式を満たしながら高速な予測を実現する手法である。
- ◆本研究では、PINO を用いた 2次元音響シミュレーション手法を提案する。
- ◆本研究の最終目的は、トランペットのような複雑な形状における音響伝播の高速予測であるが、本稿ではその前段階として、直管形状での学習結果を報告する。

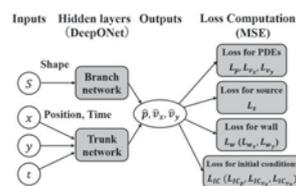


Fig.1: Physics-informed Neural Operator

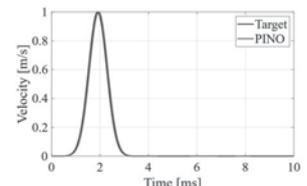


Fig.2: Learning result of Source boundary condition

1-Q-7

1-Q-7 Physics-informed Neural Operator を用いた 室内音響インパルス応答の算出

Calculation of room acoustic impulse responses using physics-informed neural operators

☆塩田朔矢, 横田和哉 (長岡技術科学大学)

- ◆機械学習を用いた室内インパルス応答(RIR)の生成に関する研究が近年活発化している。教師データの準備にはコストを要する。
- ◆Physics-informed Neural Networks (PINNs)は、ニューラルネットワークの損失関数に支配方程式に関する拘束を導入することで、教師データの必要なく数値シミュレーションが可能な手法である。
- ◆PINNs を用いて RIR を出力する Neural Operator を構築した。
- ◆提案法は教師データなしで RIR を高速・高精度に出力可能であった。

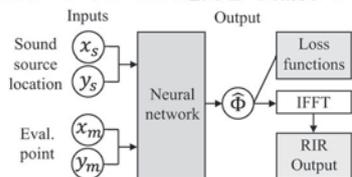


Fig.1: Proposed physics-informed neural operator for RIR calculation.

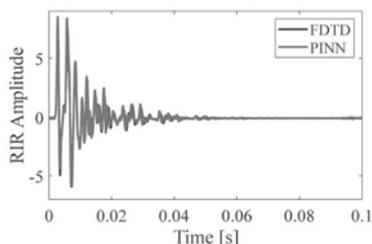


Fig.2: Calculated room impulse response.

1-Q-9

1-Q-9 ホールステージの視覚的要素が 聴感印象に及ぼす影響 —評価要因の分析—

The influence of visual elements on auditory impressions in concert halls —Analysis of evaluation factors—

☆杉本みこり, 松尾綾子, 上野佳奈子 (明治大)
井上瑞紀, 青木亜美 (日建設計)

- ◆視覚的要素による聴感印象への効果を検討するため、BoSC システムとHMDを用いた視聴覚実験を行った。本稿では評価グリッド法によって抽出された評価構造から、評価の要因を考察した。
- ◆評価構造図では、上位項目は主に『心地よさ』、『集中・没入』、『音の聴こえ』、『特別感』という4つの観点に整理された。『心地よさ』は、主に「視覚に開放感がある」とつながり、自然や外が見えるという視覚条件に関係づけられた。また、「落ちつく」を介し、壁や天井が暗いという視覚条件との関係がみられた。

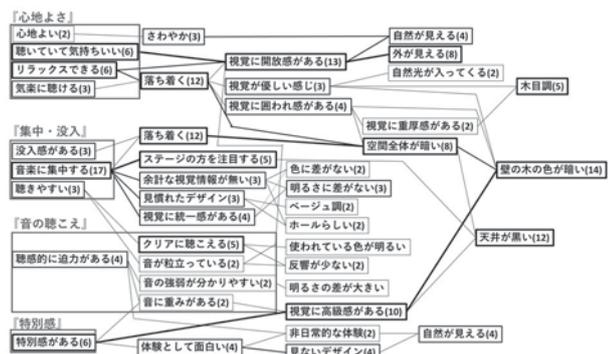


Fig.1: Evaluation structure diagram during audiovisual condition presentation

1-Q-8

1-Q-8 バンド構造解析に基づく Gyroid のバンドギャップの等方性の検証

Investigation of the isotropy of band gap in Gyroid based on Band Structure Analysis

☆矢野拓実(近畿大院), 菅原彬子, △堂寺知成, 平栗靖浩(近畿大)

- ◆Gyroid はその周期性によって特定の光が構造内を伝播できないバンドギャップ(BG)を形成する。このBGは音の遮音に応用可能と考えた。
- ◆周期構造であるGyroidに対し、周期境界条件の下で音圧の固有周波数を求めるバンド構造解析を用いた。波数ベクトルと周波数の関係を示すバンド図において固有周波数が存在しない帯域を、音波が構造内を伝播できない周波数帯であるBGとして評価した。
- ◆空隙率30%のGyroidでは約6kHz帯で全ての伝播方向に対してBGが現れ、等方性を確認した。5周期の有限厚みモデルに対する透過損失解析では、約6kHz帯で全入射方向に高い遮音性能が得られ、バンド図で示した等方的BGが実際の遮音性能に反映されることを示した。

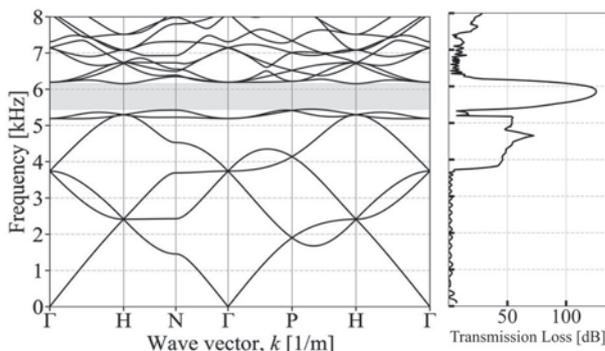


Fig.1: Comparison of Band Diagram and Transmission Loss

1-Q-10

1-Q-10 連続音場測定法を用いた 直接音成分による明瞭性評価の試み

An Attempt to Evaluate Speech Intelligibility Based on Direct Sound Components Using a Continuous Sound Field Measurement Method

◎立花杜斗 (TOA 株式会社), △小林由佳, 河原一彦 (九州大・芸工)

- ◆空間全体の音声明瞭性を評価するためには多数の測定点が必要となり、測定に多大な時間と労力を要するという課題がある。
- ◆測定者が空間内を連続的に移動しながら測定を行うことで、短時間で多点の音響情報を取得可能な連続音場測定法を提案してきた。
- ◆本手法では理想的なインパルス応答の取得が困難であり、MTFを用いた補正ができず、残響成分の影響によりSIIの評価値が過大評価される傾向がある。
- ◆連続音場測定法により得られたインパルス応答から直接音成分に着目し、直接音到達後50msまでの成分を用いた明瞭性評価を試みた。
- ◆直接音成分のみを用いた評価ではSIIが大きく低下し、過小評価となる傾向が確認された。
- ◆連続音場測定法から得られるインパルス応答を用いた評価では、直接音成分以外の成分を考慮して扱うことの重要性が示唆された。

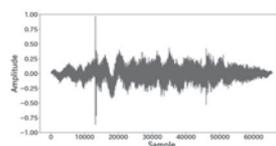


Fig.1 Impulse Response Obtained Using the Continuous Sound Field Measurement Method

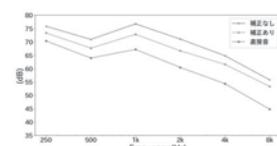


Fig.2 Speech Spectra Obtained for Each Condition

1-Q-13

1-Q-13 定常および振幅変調低周波音に対する心理・生理反応の測定と比較検討

Measurement and comparative analysis of psychological and physiological responses to steady and amplitude-modulated low-frequency sound

☆堀内誠(日大院), 松田礼(日大)

- ◆低周波音は可聴音とは異なる感覚特性をもっており、騒音問題として長年にわたり注目されている。特に、風力発電用風車から発生する低周波数成分を含み、音圧レベルが時間変動する騒音(AM音)の心理・生理的影響は不明な点が多く、健康影響のリスク・ファクタとなる可能性も指摘されている。
- ◆本報では、AM音が人体に及ぼす影響を明らかにすることを目的として、AM音暴露時の心理反応(不快感、振動感、圧迫感)と生理反応(心電図(LF/HF)、脳波、唾液アミラーゼ活性値)を測定し、定常音との比較検討を行った結果について報告する。
- ◆本研究の音刺激条件の範囲では、AM音の等価音圧レベルと定常音の音圧レベルを対応させ、かつ両者の搬送波周波数が同じ場合、同一の心理・生理指標を用いて評価できる可能性が示唆された。

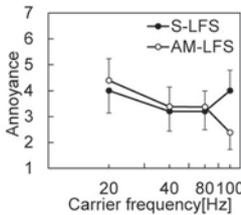


Fig.1: Relationship between carrier frequency and annoyance

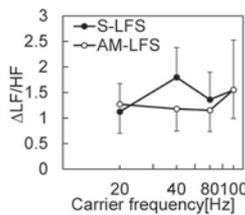


Fig.2: Relationship between carrier frequency and the rate of change in LF/HF

1-Q-16

1-Q-16 複数話者音声認識と話者ダイアライゼーションの同時自己回帰モデリングのための話者ベクトル制約

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

◎牧島 直輝, 河田 尚孝, 山根 大河, 庵 愛, 田中 智大, 鈴木 聡志, 折橋 翔太, 増村 亮 (NTT)

- ◆話者ダイアライゼーション及び音声認識は、それぞれ、誰が、いつ、及び何を話したかを推定する技術である。
- ◆従来法のSOMSREDは、複数人に対応した音声認識結果と話者ベクトルを重畳音声からend-to-endに推定する自己回帰モデルである。
- ◆SOMSREDでは、学習時に話者ベクトル空間を離散化し、話者ベクトルをトークンとして扱うが、この離散化により話者ベクトルの話者識別性が劣化する。
- ◆本稿では、この課題を解決するため、離散化しない話者ベクトル空間において、モデルの話者ベクトル出力を正解の話者ベクトルに近づけるための損失関数を導入する。

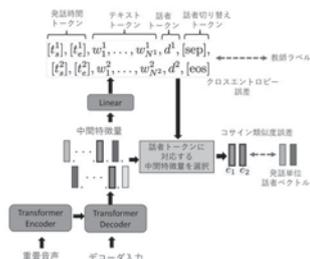


Fig.1: Overview of the proposed method.

1-Q-15

1-Q-15 テキストベースの大規模言語モデルを用いた音声事前学習モデルの評価

Evaluating Self-Supervised Speech Models Via Text-Based LLMs

◎前角高史, △後藤啓太, 篠原雄介(LINE ヤフー), △Jinchuan Tian, 渡部晋治(カーネギーメロン大)

- ◆自己教師あり学習(SSL)音声モデルは高性能が進む一方、SUPERB等の従来ベンチマークは追加学習を要し評価コストが高い。
- ◆本研究では、SSLから得た離散トークン列をテキストLLMに入力し、対数尤度に基づく学習不要・ラベル不要・パラメータ不要の評価指標(MLL)を提案する。
- ◆実験ではLLMの種類やプロンプトのパターン(P1/P2)に依らず、WERが最も低いXEUSが最も高いMLL、WERが最も高いHuBERTが最も低いMLLを示し、MLLが音声認識性能と強く相関することを確認した(Fig.1)。
- ◆また、プロンプトを変えても同様の傾向が維持されることから、提案指標はプロンプト設計の違いに対して比較的頑健であり、タスク固有の微調整なしにSSLモデル間比較の代理指標として利用できる。

Model	WER↓	MLL↑					
		Gemma3-4B		Qwen3-4B		Phi-4-mini	
		P1	P2	P1	P2	P1	P2
HuBERT	14.78	-1.780	-1.731	-1.689	-1.668	-3.162	-3.190
WavLM	3.44	-1.770	-1.712	-1.683	-1.647	-3.138	-3.139
XEUS	3.34	-1.750	-1.677	-1.668	-1.613	-3.135	-3.101

Fig.1:WER and MLL for three SSL models

1-Q-17

1-Q-17 どの顔がいつ何を話したかを推定する複数話者重畳音声における音声映像モデリング

Audio-Visual Modeling for Recognizing Which Face Spoke When and What in Multi-Talker Overlapped Speech and Video

◎牧島 直輝, 河田 尚孝, 山根 大河, 庵 愛, 田中 智大, 鈴木 聡志, 折橋 翔太, 増村 亮 (NTT)

- ◆複数話者が会話している動画の理解において、どの話者が、いつ、何を話したかの推定は重要である。
- ◆従来法では、複数人の映像に対して音声とのattention scoreを計算することで、end-to-endにどの話者が、いつ、何を話したかを推定する。
- ◆しかし、従来法は複数話者が同時に話していない非重畳音声のみを対象としており、重畳音声に対する検討はされていない
- ◆本稿では、各発話がどの映像に紐づくかを表す発話者トークンを導入し、重畳音声と各話者の映像から、どの話者が、いつ、何を話したかをend-to-endに推定するモデルを提案する。

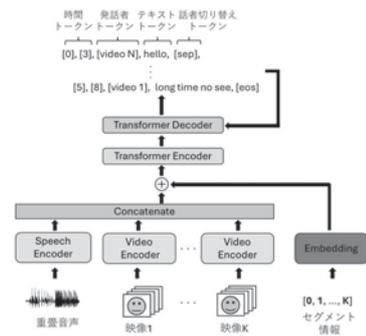


Fig.1: Overview of the proposed method.

1-Q-18

1-Q-18 Transformer による非母語話者の発話誤り検出

Transformer-based Anomaly Detection for Speech Error Detection in Non-native Speakers

☆Zhou Chenglong, △橋本 直幸(愛知工科大), 北岡 教英(豊橋技科大), 實廣 貴敏(愛知工科大)

本研究では、非母語話者の日本語発話に含まれる誤りを自動的に検出するため、異常検知に基づく発話誤り検出として、Transformerを用いた手法を提案する。日本語母語話者音声のみを用いて再構成モデルを学習し、再構成誤差から算出したマハラノビス距離を異常度指標として発話誤りを検出する。従来のAutoencoder系モデルに代えて、自己注意機構を有するTransformerを導入し、Conformerを用いたモデルも検討した。実験では、日本語母語話者音声を用いた学習データ、非母語話者音声の評価データとして用い、ROC曲線および等誤り率(EER)により評価を行った。その結果、Transformerを用いたモデルは従来手法と比較してEERが低下し、高い検出性能を示した。また、Conformerを用いたモデルでは、浅い層構成の有効性が確認された。これらの結果から、自己注意機構による大域的文脈情報の有効性が示された。

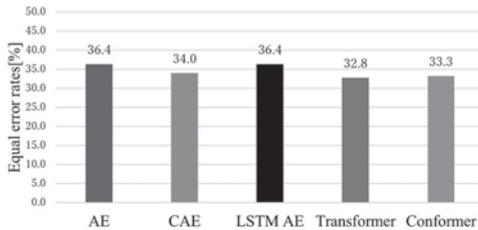


Fig. 1 Equal error rates.

1-Q-20

1-Q-20 BEST-RQ に基づく日本語音声基盤モデルの構築

Construction of the Japanese Speech Foundation Model Based on BEST-RQ.

◎高城 翼成 (豊橋技科大/SB intuitions), △Jeon Haesung,

藤田 雄介, 周藤 唯 (SB Intuitions)

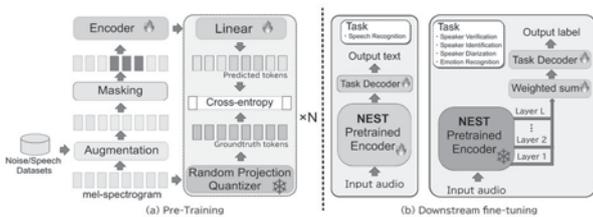


図 (a)NEST の事前学習, (b) 下流タスク用の2種類の fine-tuning

- 日本語の音声処理には日本語での事前学習が有効
→ しかし公開されている事前学習済みモデルは wav2vec2.0 or HuBERT
英語では様々な自己教師あり (SSL) 手法が提案されている。
ランダム射影量子化器を導入した BEST-RQ は効率的で高性能。
BEST-RQ + ノイズ拡張を組み合わせた NEST [1] を構築。
→従来の SSL と同等以上の日本語音声認識精度を達成

Table 1 Comparison of Japanese ASR performance across different SSL models (CER %).

Table with columns: Model, Organization, Parameter, Data (h), eval1, eval2, eval3, average, COJADS, EARS. Rows include HuBERT, NEST, and NEST SB Intuitions.

[1] H. Huang et al., "NEST: Self-supervised Fast Conformer as All-purpose Sensing to Speech Processing Tasks", ICASSP 2025.

1-Q-19

1-Q-19 RNN-Transducer を用いたマルチタスク補助情報音声認識手法

Multitask Speech Recognition with Auxiliary Information Using RNN-Transducer

○小堀 聡太, 藤江 真也 (千葉工大)

- 日本語会話音声における補助情報の同時認識
フィルラー, 言い直し, 感動詞を音声認識と同時に検出
発音形複合トークン方式 (提案手法)
補助情報を含むトークンを直接学習し, 音響的特徴と補助情報の関連性を一体的にモデル化
例) ウ+I ン+I エ+F →+F ト+F ソ+D ソ ウ ダ ネ
マルチタスク学習方式 (比較手法) (Fig. 1)
Transformer/Transducer ベースの2種類のアーキテクチャ

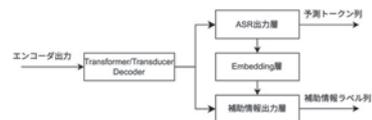


Fig. 1 MultiTaskModel overview

- 提案手法が各補助情報に対して最も高いF1値を達成 (Table. 1)

Table with columns: 補助情報, 方式, 適合率 (%), 再現率 (%), F1 値 (%), 適合率 (%), 再現率 (%), F1 値 (%). Rows include フィラー and 感動詞.

Table 1 experimental results

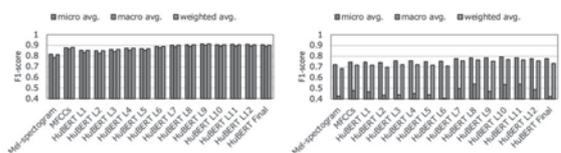
1-Q-21

1-Q-21 構音障害音声を用いた重症度分類モデルの構築と音響特徴量の比較検討

Development of a Speech-Based Model for Dysarthria Severity Classification and Comparative Evaluation of Acoustic Features

◎高畑 翔吾/福森 隆寛/△川見 員令/△長尾 青空/△門田 陽介/△辻 篤司/△宮松 直美(滋賀医大)/△小畑 大輔(神戸大病院)

- 構音障害は、脳卒中やパーキンソン病などに起因して生じる発話の明瞭性低下を特徴とする障害であり、その評価は言語聴覚士による主観的な聴覚印象評価に依存している。
重症度を客観的に評価するため、音声データに基づく自動分類技術が求められている。しかし先行研究では重症度分類の精度が十分とは言えず、日本語音声データセットも未整備という課題がある。
本研究では、独自に収集した日本語構音障害音声データを用いて、重症度分類モデルを構築した。
複数の言語聴覚士による評価に基づき、「障害無・軽度・中等度・重度」の4段階でラベル付けを実施した。
メルスペクトログラム, MFCC, HuBERT, wav2vec など複数の音響特徴表現を用いてモデル学習を行い、構音障害の有無検知および重症度分類の性能を比較検討した (Fig. 1)。



(a) Dysarthria detection (b) Severity level classification

Fig. 1: Comparison of F1-scores across features and HuBERT layers

1-Q-22

1-Q-22 パラメータ加重平均モデルが脊髄性筋萎縮症者の合成音声に与える影響の客観的評価

Evaluating the Effects of Parameter-Weighted Averaging on Speech Synthesis for Individuals with Spinal Muscular Atrophy

☆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)

- ◆先行研究では、健常者および脊髄性筋萎縮症者の音声で学習した TTS モデルのパラメータを、モジュールごとに加重平均した TTS モデル(パラメータ加重平均モデル)を構築する手法を提案し、脊髄性筋萎縮症者の話者性を維持したまま、健常者の明瞭な発話特徴を持つ音声を合成可能であることを確認した。
- ◆本稿では、効果的な話者性の維持と明瞭性の向上に向けて、各モジュールの加重平均の重みの変更と、それに伴う合成音声の発話特徴の変化との対応関係を明らかにすることを目的とする。
- ◆実験では、TTS モデルに VITS を使用し (Fig. 1)、各モジュールの加重平均の重みを個別に変更したパラメータ加重平均モデルを用いて音声を合成した。
- ◆得られた合成音声に対して、話者性・明瞭性に関する客観指標を用いた評価を実施することで、各モジュールが合成音声の発話特徴にどのような影響を与えるのか検証した。

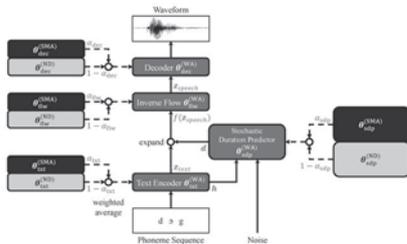


Fig. 1 Synthesis procedure using a VITS-based parameter-weighted averaging model.

1-Q-24

1-Q-24 構音障害者音声認識における多段階 fine-tuning による話者適応

Speaker Adaptation by Multi-stage fine-tuning for Dysarthric Speech Recognition

☆齋藤藤¹, 高城翼¹, 若林佑幸¹, 太田健吾², 西村良太¹, 高野哲朗³, △渡邊聡³, 北岡教英¹
(¹豊橋技科大, ²阿南高専, ³(株)ヒューマンテクノシステム)

- ◆背景: Whisper を用いた構音障害音声認識における最適な学習戦略を検証
- ◆手法: 一般コーパスを経由する「多段階 fine-tuning」と、対象データのみの「グループ適応」を比較
- ◆結果:
 - JNAS 経由は学習データと対象のドメイン差により、逆に精度が悪化
 - 構音障害音声のみを用いた「グループ適応」が、多段階 fine-tuning よりも良好な CER を達成
- ◆結論: 強力な基盤モデルにおいては、ドメインの異なるデータを挟まずグループ適応させる手法が最も有効

Table 1 Comparison of CER % in each adaptation stage

Model \ Speaker	S1	S2	S3	S4	S5	S6
Baseline (Whisper large-v3)	6.5	32.8	127.4	87.93	191.5	96.62
1. DYS Group Adaptation (w/o General Corpus)						
Dys_group	6.1	17.2	43.6	52.7	18.3	83.6
Dys_group → Speaker (Full)	6.1	19.5	36.1	48.9	22.6	86.5
Dys_group → Speaker (EncOnly)	4.9	17.7	46.8	49.1	15.4	83.2
2. Multi-stage fine-tuning via General Corpora						
JNAS	8.7	34.6	111.8	141.7	73.4	251.7
JNAS → Dys_group	7.3	18.8	44.5	61.1	16.7	89.4
JNAS → Dys_group → Speaker (Full)	6.5	23.3	50.2	65.5	20.8	88.6
JNAS → Dys_group → Speaker (EncOnly)	5.6	19.6	42.4	63.0	15.4	86.6

1-Q-23

1-Q-23 CTCとWFSTを用いたハイブリッドシステムによる舌切除者の単語及び連続数字音声認識

Word and Digit Recognition for Individuals with Tongue Resection Using a Hybrid System with CTC and WFST

☆土師梧刀(神戸大学), 高島遼一(立命館大学/JST さきがけ), 杉山千尋, 田中信和, 野原幹司, 野崎一徳(大阪大学), 滝口哲也(神戸大学)

- ◆本研究で対象とする舌切除者は、口腔腫瘍(あるいは舌がん)の治療として舌の大部分を切除したことにより、健常者とは発話スタイルが大きく異なる。
- ◆舌切除者は発話に負担があり、音声認識モデルの学習に必要な、十分な量の学習データを収録することが困難である。
- ◆本研究では重み付き有限状態トランスデューサ(Weighted Finite-State Transducers: WFST)を導入し、限られた音声データを効率的に用いることを目指す。CTC-WFST ハイブリッドシステムを構築し、舌切除者の特定話者音声認識モデルを作成する (Fig. 1)。
- ◆学習データの不足に対処するため、テキスト音声合成によるデータ拡張を活用した多段階ステップのファインチューニングを実施する。
- ◆さらに、単語及び数字認識のタスクに合致した合成音声でファインチューニングすることにより認識性能が向上することを示す。



Fig.1: Our proposed system.

1-Q-25

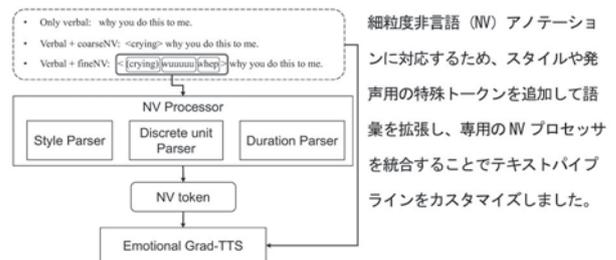
1-Q-25 細粒度非言語的表現制御による自然な感情テキスト音声合成の実現

☆周王子茜, アトマジヤ バグストリ, サクティサクリアニ
(奈良先端科学技術大学院大学)

現在の感情 TTS は、細粒度なアノテーションの不足により、感情表現に不可欠な非言語的発声 (NV) を精密に制御する能力が制限されています。

提案手法:

- 細粒度非言語表現データの構築
 - EARS コーパスを再処理。
 - 非言語音声の文字起こし (Non-verbal transcription) を作成。
- 非言語感情 TTS



結果:

主観評価の結果、提案する細粒度 NV 手法は自然さにおいて僅かな低下が見られたものの、それを補って余りある優れた表現力 (eMOS=4.2) と最高の感情認識精度 (78.8%) を達成しました。

1-Q-26

1-Q-26 Common Voice データベースを用いた複数話者 TTS の構築

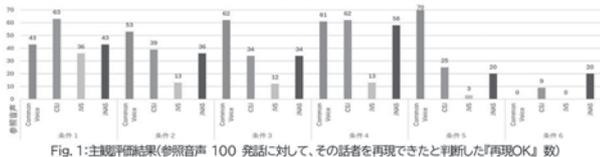
Development of Multi-Speaker TTS using Common Voice Database.

☆張 珠熒, 西村 竜一(和歌山大)

- ◆参照音声から抽出した深層話者埋め込み特徴(x-vector)を利用して、任意の話者の発話を再現する複数話者 TTS に注目
 - モデル学習時の話者不足のため、話者再現性が不十分となる課題
 - 話者数が多い Common Voice と長時間講演音声を含む CSJ を組み合わせた学習データの5条件を比較(条件1~条件5)
 - 話者特徴抽出器として日本語 x-vector モデルを導入(条件6)
 - 合成音声を聴取し、参照音声の話者再現性を主観評価

【実験結果(さまざまなコーパスの音声を参照音声として入力)】

- 条件1 学習: CSJ(先行研究)
- 条件2 学習: Common Voice
- 条件3 学習: CSJ+ファインチューニング: Common Voice
- 条件4 学習: Common Voice+ファインチューニング: CSJ
- 条件5 学習: Common Voice(話速変化: 0.9, 1.0, 1.1倍)
- 条件6 学習: Common Voice(日本語 x-vector を使用)



- ◆ Common Voice で初期学習を行い、CSJ でファインチューニングを行ったとき、未知のコーパスの参照音声に対しても高い話者再現性を確認

1-Q-28

1-Q-28 ソース・フィルタモデル型汎用ボコーダを用いた多様な話者の合成音声における品質向上

Improving the quality of various speakers' speech with a general-purpose source-filter vocoder

☆秋吉真吾(大分大院・工学研), 古家賢一(大分大・理工), 水野秀之(諏訪理工大・工)

- ◆多様な話者の音声をボコーダで合成することが求められるが、従来の特徴量だけでは品質が劣化する。
- ◆話者の身体的特徴を表す「話者ベクトル」をボコーダに追加入力することで品質向上を目指す。
- ◆本研究では、SiFi-GAN に話者ベクトルを追加入力する。
- ◆話者ベクトルの入力箇所を変更したモデルを用意し、客観評価・主観評価を行った。

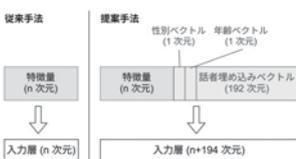


Fig.1: Combining speaker vectors and intermediate features.

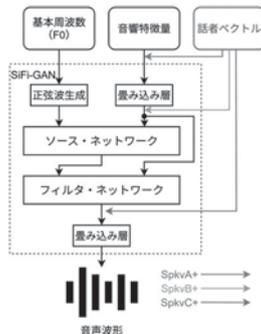


Fig.2: Schematic diagram of the introduction of speaker vectors into SiFi-GAN.

1-Q-27

1-Q-27 AnimeMOS:「アニメ声らしさ」の客観評価に向けたデータと MOS 予測モデルの設計

AnimeMOS: Data and MOS prediction models for objective evaluation of 'anime-like' voices

☆朴 浚鎔(東大院・情報理工 / Spellbrush)/ ΔLi Jerry (Spellbrush)



1-Q-31

1-Q-31 大規模モデルを用いた演じ分けに着目した落語音声合成

Speech Synthesis for Rakugo Focusing on Character Voice Differentiation Using Large-Scale Model

☆山川優綺, 上乃聖, 李昇伸(名工大)

- ◆大規模音声合成モデルをベースとし、指示文に役情報を入力することで役の特徴を反映した音声の学習および生成する手法を提案した。
- ◆各発話に役情報を付与した落語音声コーパスを構築し、役に応じた演じ分けの学習に用いた。
- ◆提案モデルによる合成音声の声質の類似度はベースラインを上回る値を示し、役の類似度評価においては同役と異役のスコア差が拡大する結果が得られた。

Table 1. The SMOS with 95% confidence intervals for speaker similarity and role similarity relative to the Ground Truth under various speaker and role conditions.

	Speaker similarity (↑)	Role similarity (↑)
Baseline		
Same spk. : same role	2.35 ± 0.35	2.21 ± 0.31
Same spk. : diff. role	2.21 ± 0.32	2.04 ± 0.30
Diff. spk. : same role	1.65 ± 0.23	1.53 ± 0.19
Diff. spk. : diff. role	1.75 ± 0.24	1.90 ± 0.26
Proposed model		
Same spk. : same role	4.44 ± 0.24	3.72 ± 0.29
Same spk. : diff. role	4.37 ± 0.21	3.37 ± 0.37
Diff. spk. : same role	2.84 ± 0.30	2.70 ± 0.30
Diff. spk. : diff. role	2.97 ± 0.29	2.47 ± 0.31

1-Q-32

1-Q-32 f_0 -BERT を用いた日本語音声合成 アクセントラベル推定の検討

A study of accent label prediction for Japanese TTS using f_0 -BERT

©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)

- ◆ テキストから f_0 を予測する手法である f_0 -BERT を拡張し、アクセントラベルを直接推定する手法を提案
- ◆ 辞書やアクセント結合規則に依存しないデータ駆動型アプローチ
- ◆ アクセント予測をメインタスクとし、トークン予測と f_0 予測を補助タスクとしたマルチタスク学習により高精度な予測を実現
- ◆ JSUT コーパスでの評価実験により、辞書ベース手法を大きく上回るアクセント正解率 93.58%、文完全一致率 22.00%を達成
- ◆ アクセントラベルを直接得ることで、TTS における制御性・解釈性の向上が期待できる

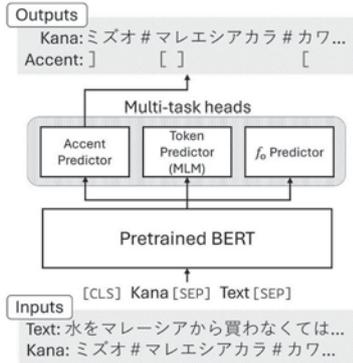


Fig: Overview of the proposed method. Text and katakana reading are input to Japanese BERT, and accent label is predicted for each mora through multi-task learning. { } = Rise, { } = Fall (accent nucleus).

講演取消

1-Q-34

1-Q-34 音声のキャラクター性設計に向けた 役割語の人称代名詞に関する印象評価

The Impression Evaluation of Personal Pronouns in "Yakuwarigo" for Character Voice Design

☆劉承受(明治大院), 林大輔(日本たばこ産業), 森勢将雅(明治大)

- ◆ 声には、話者属性とは独立したステレオタイプの印象が存在する。
- ◆ 「役割語」は、特定の属性を想起するステレオタイプの言葉遣いであり、人称代名詞や文末表現に現れやすいとされる。
- ◆ 本研究では、声におけるキャラクター性について検討するために、「役割語」による韻質要素がどのような印象を持つか調査した。
 - > 今回は、人称代名詞を 14 個用いて検討を行った。
 - > 性別、年齢、性格の 3 軸を基に 6 つの評価項目を設定し、SD 法を実施した。
 - > 加えて、「役割語」から想起される人物像を選択する実験を行った。
- ◆ 主成分分析の結果、3 軸を明確に分けられ高い解釈性を示した (Fig. 1)。
- ◆ 「役割語」から想起される人物像の、選択傾向が参加者間で類似した。

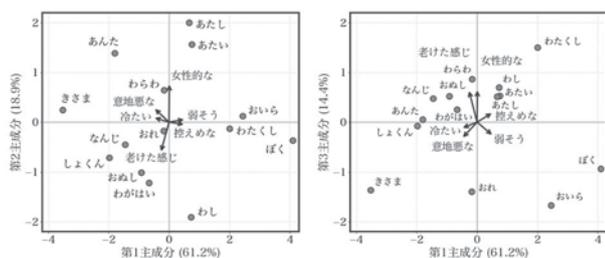


Fig. 1 : Biplots of the first and second principal components and the first and third principal components using the average scores between participants from impression evaluation of personal pronouns in "Yakuwarigo" by the SD method.

1-Q-33

1-Q-35 Improving Automatic Dubbing Speech Synchronization with In-Context Paraphrasing and Synchrony-Based Re-Ranking

☆ Jan Meyer Saragih, Faisal Mehmood, Sakriani Sakti
Nara Institute of Science and Technology

- ◆ We have improved speech synchronization on our previous automatic dubbing framework by including in-context paraphrasing after machine translation and before text-to-speech.

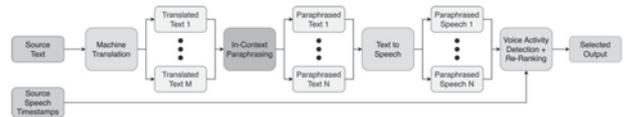


Figure 1: Proposed Framework

- ◆ In-context paraphrasing managed to improve the variety of generated translation candidates significantly while preserving the semantic of the candidates
- ◆ Increased variety manages to improve the speech synchronization while maintaining high translation quality.

Table 1: Paraphrasing Impact on Dubbing Metrics

System	Re-ranking	BLEU	COMET	SLC	SO
Baseline	-	20.74	0.655	0.813	0.570
Top-1 MT	-	48.00	0.974	0.721	0.319
Top-5 MT	Sorted by SLC	41.45	0.971	0.763	0.335
	Sorted by SO	43.15	0.976	0.744	0.360
Paraphrase	Sorted by SLC	41.96	0.973	0.871	0.402
	Sorted by SO	44.38	0.976	0.776	0.530

SLC = Speech Length Compliance, SO = Speech Overlap.

1-Q-40

1-Q-40 TV-CAR 分析に基づく ARMAX-LF 分析の検討

A study on ARMAX-LF analysis based on TV-CAR analysis

○舟木慶一(琉球大学)

音声信号の生成モデルを声帯音源波モデルで生成される声帯音源波と声道フィルタのソースフィルタモデルと仮定し、音声から LF モデル [1] と AR 係数を同時推定する ARX-LF 分析 [2] が李らにより提案されている。ARX-LF 方式は SEDREAMS 法 [3] で GCI(声門閉鎖点)を推定し、IAIF 法 [4] で推定された声帯音源から DyProg 法 [5] により推定された LF モデルパラメータを予備推定値とし、その近傍に対する ARX モデル同定により LF モデルパラメータを本推定する方式である。一方、我々は解析信号に対する TV-CAR 分析 [6] の応用として、TV-CAR 分析に基づく SEDREAMS 法 [7]、IAIF 法 [8][9]、DyProg 法[10] の提案を行い、さらに、それらに基づく ARX-LF 分析を構築した [11]。今回、ARX を ARMAX モデル同定に拡張した ARMAX-LF 分析を構築したので、報告する。

1-Q-42

1-Q-42 マルチモーダル情報を用いた Voice Activity Projection Turn-taking prediction using multimodal information

○花川主税, 藤江真也(千葉工大)

- ◆音声と非言語特徴を用いたターンテイク予測
 - 視線・頭部姿勢の角度を直接利用する方法と、各推定モデルの中間表現特徴を利用する方法の比較

- ◆複数の統合方法
 - 話者毎に音声・非言語を先に統合してから話者同士を統合(Fig. 1)
 - モーダル毎に話者同士を統合してからモーダルを統合(Fig. 2)

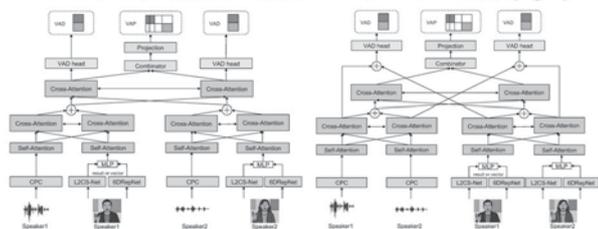


Fig. 1 Speaker Late Fusion

Fig. 2 Modal Late Fusion

- ◆評価タスクに応じて有効な統合方法が異なることが示された
 - 短い発話に対して Speaker Late Fusion が有効
 - 長い発話には Modal Late Fusion が有効
- ◆非言語情報推定モデルの中間表現特徴を用いることで、話者交代や話者保持の精度が向上
 - 特に抽象的な情報が話者交代タスクにおいて有効

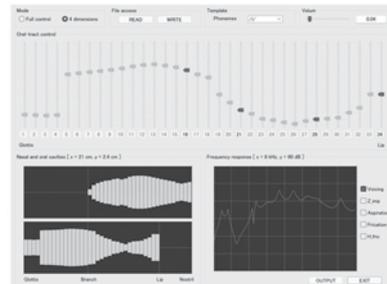
1-Q-41

声道断面積関数の手指を用いたリアルタイム制御に関する検討

A study on real-time control of the vocal tract area function using fingers

○伊藤仁, 加藤楓雪, 櫻田峻(東北工業大学)

- ◆手足の身体動作を入力として、音韻や韻律を即時的に制御する Motion-To-Speech (MTS) 型音声合成に、人間の発声機構を模擬した合成アルゴリズムを導入する。
- ◆本稿では、口腔と鼻腔で構成される声道断面積関数を、口腔内 4 か所の断面積および口腔から鼻腔への分岐を制御する変数の 5 次元入力から推定する手法について検討した。
- ◆日本語 5 母音および鼻音/m,n/の声道計測データに対する声道断面積関数の推定誤差は、最大 1.27 cm²、平均 0.34 cm²であった。また断面積関数から計算した声道伝達特性は、平均 RMS 誤差 2.94 dB で実測データと良好に一致した。
- ◆上記の 5 次元入力を、片手の指の動作として取得するためのデバイスとして、四節リンクを用いた構造を提案した。



The developed simulator for estimating and visualizing the vocal-tract

1-Q-43

1-Q-43 アニメ調キャラクター顔画像に適するクロスモーダル音声合成のための再現性・親和性改善の検討

A Study on Improving Reproducibility and Face-Voice Affinity in Cross-Modal Voice Synthesis for Anime-Style Characters.

☆菊池遥斗, 能勢隆, 小林清流, 林崎由(東北大), 橋本佳(名工大), 伊藤彰則(東北大)

- ◆本稿では、キャラクター顔画像から音声話者特徴量を推定するクロスモーダル音声合成手法 F2S-SBV2 の改良として、顔画像特徴量を音声話者特徴量に変換する F2S (Face-to-Speech) エンコーダの精度改善のため、分散補償や年齢ラベルを用いた手法について検討した。
- ◆分散補償を利用した場合、話者再現性の向上は達成できなかった一方、分散の観点では推論音声話者特徴量が元音声話者特徴量に近づき、分布縮小が改善されることを確認した。
- ◆年齢情報を推論に利用した場合、話者再現性の向上には至らなかったものの、年齢情報の表現方法や導入方法が推論結果に与える影響について今後のモデル設計に有用な示唆が得られた。

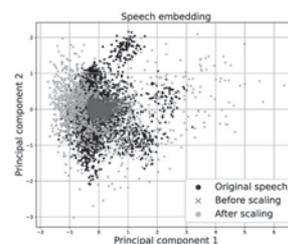


Fig. 1: Effect of variance compensation on inferred feature representations

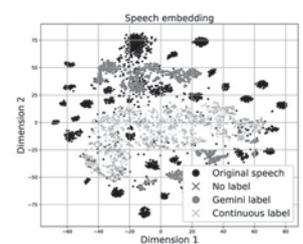


Fig. 2: Effect of age labels on inferred feature representations

1-Q-44

講演取消

1-Q-45

講演取消

1-Q-46

1-Q-46 Moshiに基づく音声対話モデルの日本語ファインチューニングにおける対話データ特性の影響
Effects of Dialogue Data Characteristics on Japanese Fine-Tuning of Moshi-Based Japanese Spoken Dialogue Models

☆阿部 雄斗(早大/NII LLMC)/佐伯 真於(早大/エキュメノボリス)/大橋 厚元(名大)/高道 慎之介(慶大)/藤江 真也(千葉工大/早大)/小林 哲則(早大)/小川 哲司(早大)/東中 竜一郎(名大/NII LLMC)

◆目的: Moshiに基づく日本語音声対話モデルを基盤として、学習に用いる対話データの特性がモデル性能に与える影響を調査。

◆貢献:

➢ LLM-jp 対話ワーキンググループの活動の一環として、商用利用可能な日本語音声対話コーパス LLM-jp-Zoom1 を整備し、日本語全二重音声対話モデル LLM-jp-Moshi を構築。

➢ 雑談対話音声による対話継続タスクで、音質客観評価および言語的客観評価・主観評価を実施し、LLM-jp-Moshi の性能と学習・評価に用いる対話データ特性の関係を分析。

◆主観評価実験結果 (Fig.1)

- LLM-jp-Moshi は既存の J-Moshi を上回る性能を達成
- 多様な対話データを用いた学習の有効性を確認

Table 1 日本語 CallHome, LLM-jp-Zoom1 の継続対話音声に対する主観評価結果

Pre-training	J-Moshi	LLM-jp-Moshi	LLM-jp-Moshi+	GT
	J-CHAT (69k hours)	J-CHAT (69k hours)	J-CHAT (69k hours)	
	Tabidachi			
Fine-training	内製データ (200 hours)	LLM-jp-Zoom1	京都観光業内対話データベース	
	日本語 CallHome, CSJ		VisualBank	
日本語 CallHome	自然性 (1-5) 2.35 意味性 (1-5) 1.54	2.82 2.34	3.12 2.56	4.03 4.10
LLM-jp-Zoom1	自然性 (1-5) 2.36 意味性 (1-5) 2.14	3.02 2.61	3.19 2.80	4.31 4.46

1-Q-47

1-Q-47 チャンク単位話者クエリを用いたターゲット話者発話区間検出
Target speaker voice activity detection using chunk-level speaker queries

○依直弘, 堀口翔太, デルクロアマーク(NTT)

- ◆TS-VAD (Fig.1) は初期ダイアリゼーション結果から得た話者埋込みで各話者の発話区間を再推定する手法で、重なり発話に強い一方、セッション全体の固定埋込みに依存するため局所的音響変動に弱い
- ◆数秒の短いチャンク単位で話者埋込みを動的推定することで、局所的な音響変動に頑健なチャンクレベルTS-VADを提案 (Fig.2)
- ◆DIHARD III において、提案手法が従来のTS-VADより低いDERを達成することを確認 (Table)

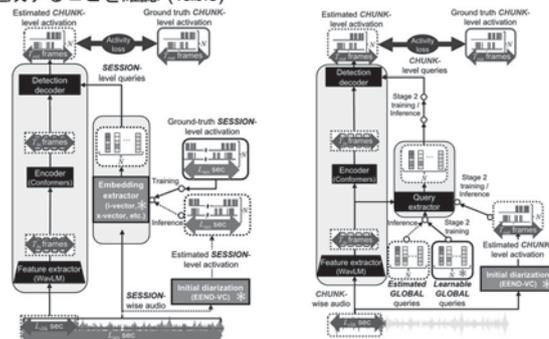


Fig.1: Conventional session-level TS-VAD Fig.2: Proposed chunk-level TS-VAD

Table: Diarization performance (%) of conventional and proposed TS-VAD

Model	Miss	False accept	Confusion	DER
Conventional TS-VAD	6.43	3.75	3.15	15.23
Proposed TS-VAD	6.30	4.68	3.18	14.16

1-Q-48

1-Q-48 演技音声に対する話者照合システムの有効性の検証

Verification of the Effectiveness of a Speaker Verification System for Acting Voices

☆山本峻太(明治大学), 上村卓也(NTTコミュニケーション科学基礎研究所), 林大輔(日本たばこ産業株式会社), 森勢将雅(明治大学)

- ◆声優の演技音声に代表されるように、人は年齢や性別などが異なるように知覚される声を発することが可能である
- ◆人間は異なるキャラクターの声に対して話者が同一であるかを認識することが難しいことが示されている
- ◆このような人間の認識特性を踏まえ、演技音声に対して、地声音声で学習された話者認識モデルの有効性を評価する
- ◆異なるキャラクターの演技音声では、同じキャラクターの演技音声と比べて、類似度が全体的に低下し (Fig.1), 演技音声では話者認識性能が低下した

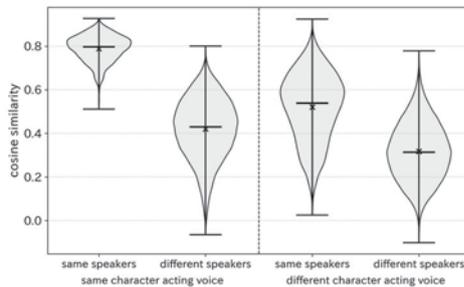


Fig.1: Violin plot showing the distribution of similarity. The horizontal line indicates the median, and the cross indicates the mean.

1-Q-52

1-Q-52 ラベルなし発話をモデル学習に含めた話者年齢層推定

Speaker age-group estimation incorporating unlabeled utterances into model training.

☆平間 光, 西村 竜一(和歌山大)

- ◆話者年齢層推定手法において半教師あり学習の導入を検討
 - 若年者・高齢者の年齢ラベル付き発話データセットが不足
 - 年齢ラベルなし発話データを識別モデルの学習に利用
- ◆提案手法(藤居ら, 2025)
 - 前段: 深層話者年齢埋め込みベクトル (Age-vector) 抽出モデル
 - 後段: 8クラス(学校区分)の識別モデル (SGAN)
- ◆識別モデル (SGAN) の学習に年齢ラベルなし発話データを追加
 - 「教師あり学習のみ」と「半教師あり学習を追加」を比較
 - 小学生・若年者(10代)発話を含む年齢ラベルなしの発話を使用
- ◆Accuracy と QWK(Quadratic Weighted Kappa) で性能向上を確認
 - 小学校「中学年」「高学年」以外のクラスでの識別性能向上

Table 1 混同行列 (条件1: 教師あり学習のみ)

	話者	年齢						
		幼稚園	小学校低学年	小学校中学年	小学校高学年	中学校	高校	大人
幼稚園	13	33	4	7	3	6	4	
小学校低学年	27	11	31	18	8	6	8	
小学校中学年	30	35	14	30	10	4	19	
小学校高学年	9	8	28	12	42	17	23	
中学校	1	3	11	27	56	36	47	
高校	5	0	0	5	20	50	40	
大人	20	8	13	25	94	153	718	
ACC		0.549	0.555	0.570	0.528	0.264	0.186	0.834

Table 2 混同行列 (条件2: 半教師あり学習を追加)

	話者	年齢						
		幼稚園	小学校低学年	小学校中学年	小学校高学年	中学校	高校	大人
幼稚園	13	35	30	13	8	4	11	
小学校低学年	22	12	32	11	6	3	5	
小学校中学年	21	27	14	33	11	7	4	
小学校高学年	3	11	24	19	26	10	20	
中学校	4	4	14	31	87	30	37	
高校	3	2	0	7	23	89	55	
大人	19	10	12	27	85	126	728	
ACC		0.645	0.574	0.556	0.472	0.354	0.331	0.844

1-Q-49

1-Q-49 Cross-Stream Attention を用いた振幅・位相情報に基づくなりすまし音声検出

Spoofer Speech Detection Based on Magnitude and Phase Information Using Cross-Stream Attention

☆PHAM HOANG LONG, 西田昌史(静岡大), 柘植覚(大同大), 黒岩眞吾(千葉大)

近年の音声合成 (TTS) や声質変換 (VC) 技術の急速な発展により、なりすまし音声システムの信頼性に深刻な脅威を与えている。このような背景から、なりすまし音声検出 (CM) の高精度化が重要な研究課題となっている。本研究では、複素STFT から得られる振幅スペクトルと位相を二つの独立したストリームとして扱い、両者を相互に参照しながら統合する Cross Stream Attention (CSA) というなりすまし音声検出モデルを提案する。さらに、位相の不安定性を抑制するため、位相正規化に用いる基準周波数および基準位相を固定せず、音声サンプルごとに最適な基準値を推定する MLP による位相正規化手法を新たに導入する。

評価実験では、ASVspoof2019 LA データセットを用いて学習を行い、同データセットの eval および In-the-Wild データセットにより性能評価を行った。Table 1 より、提案 CSA モデルが、従来の複素 CNN を上回る検出性能を示し、位相正規化を組み込むことで性能向上が確認された。また、事前学習モデル AASIST との比較では、LA 条件では AASIST が優位である一方、実環境条件に近い In-the-Wild データでは CSA が相対的に AASIST より高い精度を示した。

Table 1 EER of each spoofed speech detection model [%]

Model	LA dev	LA eval	In-the-Wild
AASIST	0.5	0.8	43.2
Phase-normalized Complex CNN	2.7	6.6	36.6
Non-phase-normalized CSA model	1.8	5.7	37.3
Phase-normalized CSA model with fixed reference frequency and reference phase	1.0	5.1	35.1
Phase-normalized CSA model with self-learning reference frequency and reference phase (Proposed method)	0.3	4.7	33.5

1-Q-53

1-Q-53 明瞭発声を目標音声とするゼロショット声質変換システムによる明瞭性向上手法の検討

A Study on Intelligibility Enhancement Based on Zero-Shot Voice Conversion by Using Clear Speech as the Target.

☆近野圭吾, 坂野秀樹, 旭健作(名城大),

- ◆本稿では、目標音声を明瞭発声とした Seed-VC により通常発声を変換し、明瞭性の向上がどの程度可能であるかを評価する。併せて、特定の子音成分の変換が明瞭性に及ぼす影響についても分析を行う。
- ◆実験の結果としてゼロショット声質変換技術「Seed-VC」を用いることで、音声の明瞭性が向上することを確認した。一方で、特定の子音成分を目標音声とした変換については、動的特徴量の増加が確認されたものの、MOS テストにおける明瞭性の向上には結びつかなかった。

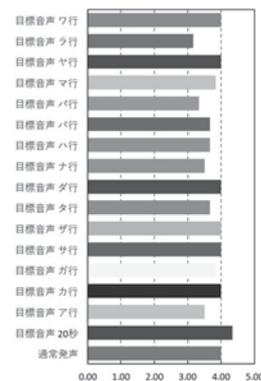


Fig.1: MOS test results for each target voice

1-Q-54

1-Q-54 ピッチ強調フィルタを用いたオープンイヤー型イヤホンにおける騒音下再生音の明瞭性向上

Improving the intelligibility of sound reproduced by open-ear earphones in noisy environments using a pitch-emphasizing filter.

渡邊悠希, 千葉大将, 鎌本優, 加古達也 (NTT)

- ◆オープンイヤー型イヤホンは、快適性が高い一方で低音域の出力が弱く、周囲の騒音が大きい環境では特に再生音が聞き取りにくくなる。
◆オープンイヤー型イヤホンでの単純な低音域の増幅は音の歪みを生じやすい。非線形歪み補正技術も、高演算量であり利用しにくい。
◆そこで本研究では、低演算量なピッチ強調フィルタを応用し、低周波調波構造を制御することで騒音下での聞き取りやすさを向上させる方法を提案する。
◆主観評価実験の結果、本制御方法 (LFHC-3(0.6)) は異なる話者の6つの音声サンプルのうち3つにおいて、音声の聞き取りやすさを有意に改善した (Fig. 1).

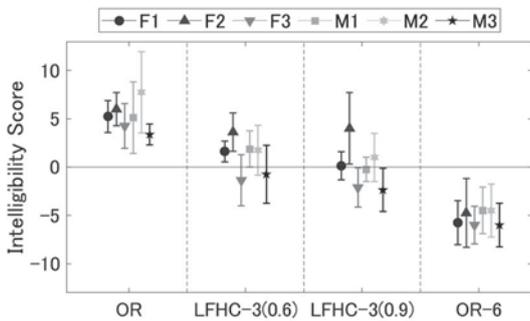


Fig. 1: Results of the subjective evaluation.

1-R-1

1-R-1 ノイズキャンセリングイヤホン使用が警告音認知に及ぼす影響: 単純加算作業および脳波測定による分析

The effect of using noise-canceling earphones on the perception of auditory warning sounds: An electroencephalography study during a simple addition task

☆宮崎大和(東海大), 程島奈緒

- ◆ノイズキャンセリングイヤホンで音楽聴取中の事件・事故が増加。
◆ノイズキャンセリングが危険を示す警告音を除去している可能性。
◆脳波測定を用い警告音の種類でその認知に影響を及ぼすか検討。
◆実験条件はイヤホンの装着状態3条件と警告音2種類の計6条件。
◆一位加算作業中にイヤホンから楽曲、スピーカーから警告音を提示。
◆Cz電極におけるβ帯線形パワーを指標として平均値を算出(Fig. 1)。
◆条件1および条件6でβ帯線形パワーがベースラインを下回った。
◆条件間で有意差はなし。個人の特性による影響が大きいと考えられる。
◆時間的推移や個人差に着目した解析が必要。

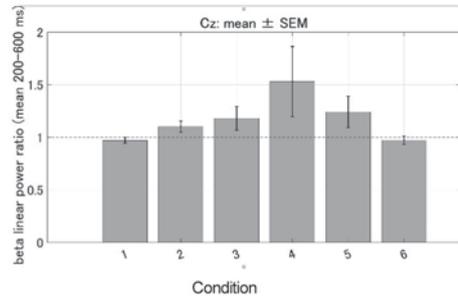


Fig. 1 Time-frequency representation at Cz

1-R-2

1-R-2 複合振幅変調音に対する聴性定常反応の計測(第2報): 2音の変調周波数差の影響

Measurements of auditory steady-state responses to stimuli amplitude-modulated by complex tones (Second Report): Effects of the modulation-frequency difference between two tones

○米玉利雅史, 大塚 翔(千葉大),

長谷芳樹, 星 貴之(ピクシーダストテクノロジーズ(株)), 中川誠司(千葉大)

- ◆聴性定常反応 (Auditory Steady-state Response: ASSR) は専ら単一周波数による振幅変調音 (単一変調音) を用いて誘発され、複合音による変調音 (複合変調音) に対する反応特性の多くは不明である
◆複合変調音に対する ASSR は、振幅包絡の時間変動を分析する“変調フィルタ”の影響を受けると考えられるが、2つの変調周波数の周波数差と ASSR の関係を体系的に調査した例は存在しない。本研究では一方の変調周波数を 40 Hz に固定し、他方の変調周波数を 11 Hz から 39 Hz の間で変化させた場合の ASSR ピーク振幅を調べた
◆複合変調音では単一変調音より ASSR の振幅が減少した。先行研究の結果とも一致しており、他の変調周波数成分によって脳内のリソースが奪われたためと考えることができる。
◆一方、40 Hz には変調周波数間の距離の有意差がなかった。各変調周波数成分が異なる変調フィルタに入ったためと考えることができるが、さらなる検証が必要である。

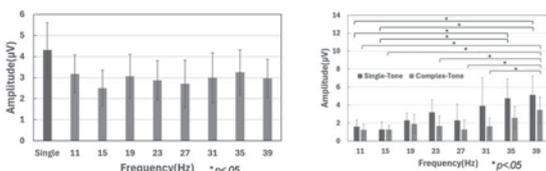


Fig. 1 ASSR peak amplitudes at 40Hz (left) and at each frequency (right) for single and complex SAM tones (mean ± SD).

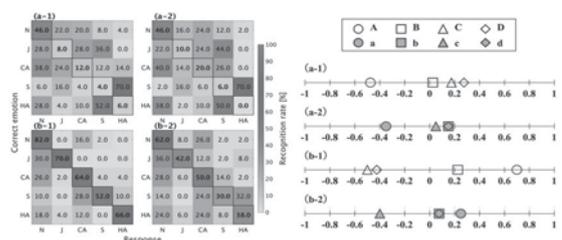
1-R-3

1-R-3 触覚刺激提示による音声の非言語情報知覚の検討

Study on the Perception of Non-linguistic Information in Speech through Tactile Stimulation

☆牧野哲平, 上江洲安史, 木谷俊介, 鶴木祐史 (JAIST)

- ◆目的: 聴覚代行の一つの試みとして、音声に触覚刺激として提示した際の非言語情報知覚の実現可能性について検討すること
◆方法: <触覚による音声知覚の手掛かりを検討するために>
①異なる身体部位(掌と頬)の触覚閾値を計測
②触覚知覚での時間変調伝達関数を計測
<触覚刺激提示時の非言語情報知覚を検討するために>
③原音声と雑音駆動音声(NVS)を刺激として触覚提示としたときの感情知覚実験ならびに緊急感知知覚実験を実施
◆結果:
・①と②から、触覚による変調知覚の特性がわかった。
・③から、原音声と NVS 提示時の結果が類似しており、触覚の変調知覚を利用して非言語情報を知覚している可能性が示唆された。



(a) Results of the emotion perception experiment (b) Results of the urgency perception experiment Fig. 1 Results of non-linguistic information perception

1-R-4

1-R-4 音声の音高と音色の知覚の個人性調査

Investigation of perceptual individuality for pitch and tone-height of speech
○森田 翔太(福山大・工)/鳥谷 輝樹(山梨大院・総合研)

- ◆日本語5母音の音高・音色(ピッチと音色的高さ)の弁別特性を調査。
- ◆方法
 - 音声分析合成システム World による音声変換
 - ピッチ: 音声の F0 を ±10 Hz の範囲で 2 Hz 刻み幅で変化
 - 音色的高さ: 音声のスペクトル傾斜を ±5 dB/Oct の範囲で 1 dB/Oct の刻み幅で変化
- ◆結果
 - 自己発話音については男性の方が F0 変化に対してピッチの高低を判断できる割合が高い
 - +方向のスペクトル傾斜変化に対して、音色的高さの相違を判断できる割合は50%以下

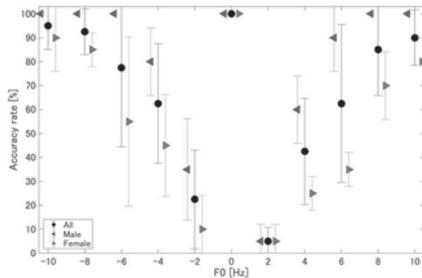


Fig.1: Accuracy rate of pitch discrimination using own F0 changed voices of 5 Japanese vowels

1-R-6

1-R-6 経頭蓋交流電気刺激による音韻修復機能の促進効果の検討

An Investigation of the Facilitating Effects of Transcranial Alternating Current Stimulation on Phonemic Restoration
☆折戸陸, 大塚翔, 中川誠司(千葉大)

- ◆音韻修復機能とは雑音などにより一部がマスキングされた音声聞いた際に、欠落した部分を脳が補完する機能である。この機能を促進することができれば、加齢等で生じる雑音下での聴取困難を緩和するリハビリテーションへの応用が期待できる。
- ◆本研究では経頭蓋交流電気刺激 (tACS) を用いて、音韻修復に関与するとされる左角回から上側頭溝にかけての領域の活動を非侵襲的に活性化することで、音韻修復の働きが促進される可能性を検証した。
- ◆θ帯域に相当する 6 Hz の tACS 後には親密度 (聞き馴染みの程度) が低い単語群で音韻修復が生じやすくなった (Fig. 1A)。一方で、γ帯域に相当する 50 Hz の tACS 後には親密度によらず、音韻修復が生じにくくなった (Fig. 1B)。
- ◆脳内の異なる周波数帯の神経律動が、音声の時間的な変調構造に同期することで、段階的に音声进行处理するという仮説がある。この仮説に基づくと、θ帯域の tACS によって音節レベルの処理が活性化され、欠落区間前後の情報を必要とする音韻修復を促進する方向に働いたと考えられる。一方で、γ帯域の tACS では、音素レベルの処理が活性化されたことで、音素間を跨ぐ処理に割かれる神経資源を減少し、音韻修復が抑制された可能性が推察される。

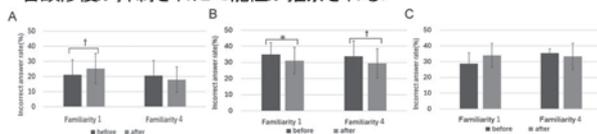


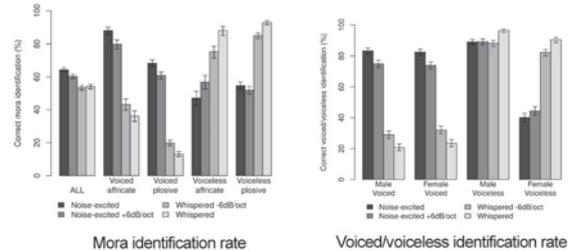
Fig. 1 Error rate of misclassifying the "replaced" speech as "superimposition" before and after 6-Hz tACS (A), 50-Hz tACS (B) and sham stimulation (C). †p<0.1, *p<0.05.

1-R-5

1-R-5 無声音の子音同定におけるスペクトル傾斜の影響

Effect of Spectral Tilt on Identification of Consonants in Unvoiced Speech
○今野英明(北海道教育大)

- ◆基本周波数を持たない音声の音韻知覚に対するスペクトル傾斜の影響を有聲・無声子音について調べた
- ささやき声 (Whispered)
- 通常音声の音源を雑音に変更して再合成した音声 (Noise-excited)
- これらにスペクトルの傾斜 ±6 dB/oct を付加した音声 (Whispered -6 dB/oct, Noise-excited +6 dB/oct)
- ◆既報の被験者を男性 8 名, 女性 8 名の計 16 名にした報告である
- ◆男性 3 名, 女性 3 名が発声した 55 単音節 (有聲・無声破裂子音, 有聲・無声摩擦子音) の音韻同定実験
- ◆有聲・無声子音の聴き取りにスペクトルの傾斜が影響
 - Whispered 全般と男声 Noise-excited の有聲子音
 - 低周波数域エネルギーの増加は有聲子音に聴取
 - Noise-excited は女性/男声で聴取が異なる



1-R-7

1-R-7 オープンイヤー型および密閉型ヘッドホンを用いたロンバード音声の比較と模擬難聴状態での知覚影響の調査

Comparison of Lombard Effect Perception Using Open-Back and Closed-Back Headphones and Investigation of Perceptual Effects under Simulated Hearing Loss Conditions
☆寺村実優, 坂野秀樹, 旭健作(名城大院)

- ◆本研究では、オープンイヤー型および密閉型ヘッドホンで収録したロンバード音声の違いを比較し、それらが模擬難聴状態で音声知覚に与える影響を明らかにする。
- ◆比較分析では基本周波数とパワー分析を行い、聴取雑音の音圧レベルが大きくなるほどロンバード音声の基本周波数、パワーが上昇する傾向もあったが、個人差が見られた。
- ◆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。
- ◆オープンイヤー型ヘッドホンにおいて、聴取雑音の音圧レベルが上がるとつれ、単語正答率も上昇傾向が見られた。

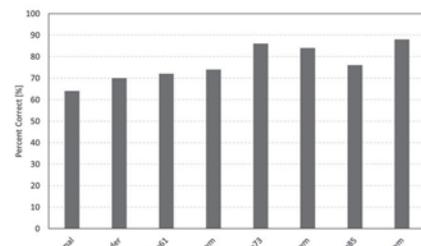


Fig. 1 Correct answer rate for word

1-R-8

1-R-8 音声了解・品質の客観評価指標を用いたGANモデルによる難聴者ための音声強調処理

Speech enhancement processing for the hearing impaired using a GAN model with objective metrics for speech intelligibility and quality

☆孫 浩正, 羅 文煜, 山本 絢子, 時政 和征(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 GESI-GAN より改良した音声強調モデル「GESIV-GAN」を提案する。雑音付加音声と雑音を入力とし、了解度 (GESI) および音質 (VISQOL) 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、SSDRC 処理した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。

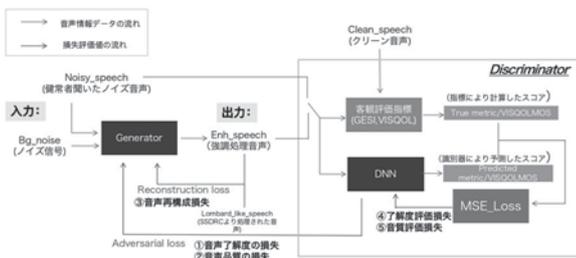


Fig.1: Overall architecture of the proposed model GESIV-GAN

1-R-9

1-R-9 個人の聴覚特性を反映した DeepGESI による難聴者の音声了解度予測

Speech Intelligibility Prediction for Hearing-Impaired Listeners Using DeepGESI Incorporating Individual Auditory Characteristics

☆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、個人ごとの聴覚特性を反映した非侵入型音声了解度予測モデルを提案する。
- ◆提案モデルは、従来の DeepGESI に個人ごとのオーディオグラムを導入することで、音声に対する予測精度の向上を実現する。
- ◆本手法は非侵入型 (Non-intrusive) の計算方式を採用しており、基準音声を必要としない。
- ◆RoPE により音響系列のための動的な位置エンコーディング手法 RoPEAS を開発し提案モデルに導入することで、可変長音声に対応可能とした。
- ◆実験結果より、提案手法は従来法と比較して予測誤差を大きく低減できることを確認した。

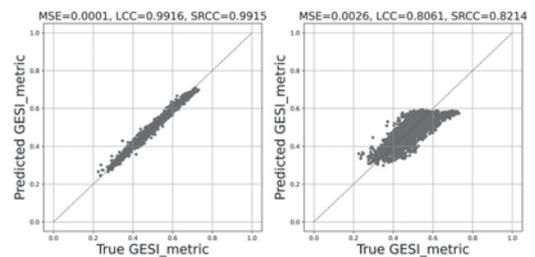


Fig.1: Scatter plots comparing the prediction accuracy of the proposed method (left) and the baseline method (right) for the GESI metric.

1-R-10

1-R-10 改良した DeepPIQ による模擬難聴音声の了解度・音質予測

An Improved DeepPIQ Model for Predicting Speech Intelligibility and Quality of Simulated Hearing-Impaired Speech

○時政和征, 羅文煜, 北芝優成, 入野俊夫, 陳金輝 (和歌山大)

- ◆本研究では、非侵入型の音声了解度・音質の客観評価指標 DeepPIQ の「共有部拡張による軽量化」と「オーディオグラム情報の導入」を試みた。
- ◆まず、ネットワークの共有範囲を従来の畳み込み層から全結合層まで拡張することで、モデルのパラメータ数を約 7.5%削減する軽量なアーキテクチャを構築した。
- ◆次に、個々の難聴者に適応した予測を実現するため、入力にオーディオグラム情報を追加する手法を導入した。実験により、聴力情報の統合が特定の対象に対する了解度予測精度の飛躍的な向上に有効であることを確認した。

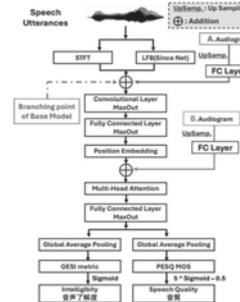


Fig.1 : DeepPIQのベースモデル

Fig.2 : 共有部拡張 DeepPIQ

1-R-11

1-R-11 音質評価のための非侵入型客観評価指標 DeepPESQ

Non-intrusive Objective Speech Quality Assessment using DNN

☆北芝優成, 時政和征, 入野俊夫, 陳金輝(和歌山大)

- ◆音質評価の客観評価指標には、従来 PESQ が広く使用されてきた。しかし、PESQ は評価にあたって「基準音声」と「劣化音声」の両方を必要とする「侵入型」の指標であり、実環境での応用には制約がある。
- ◆このような背景から、本研究では、DNN を用いて従来の PESQ スコアを教師データとして学習させることで、基準音声を必要とせず、リアルタイム処理が可能な客観的音質評価モデル DeepPESQ を開発した。

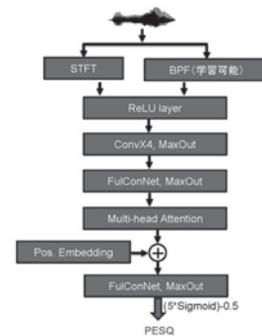


Fig.1: DNN architecture of DeepPESQ.

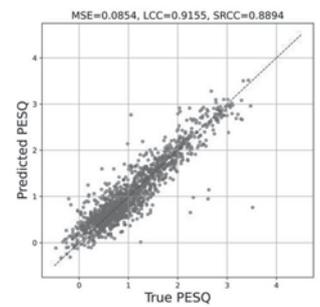


Fig.2: Prediction results of DeepPESQ on test dataset.

1-R-12

1-R-12 Integrating Peripheral and Central Processing into Speech Intelligibility Prediction for Hearing-Impaired Listeners

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

- ◆ **Aim:** To investigate how modeling peripheral and central processing can improve speech intelligibility prediction across different hearing loss conditions in noisy environments.
- ◆ **Problem:** Existing prediction methods often rely on peripheral processing and overlook the role of central processing in integrating spectral, temporal, and binaural cues.
- ◆ **Solution:** Peripheral processing is modeled to simulate auditory degradations caused by hearing loss. Central processing is modeled in parallel using spectro-temporal modulation analysis and a convolutional neural network (CNN) with binaural interaction to capture learned central features.
- ◆ **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 9.7% and reduced the RMSE by 11.5%.

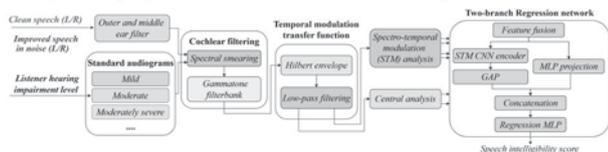


Fig.1: Simplified overview of proposed method

1-R-14

1-R-14 リアルタイム声質変換が 発話行動に与える影響の解析

Analysis of the effects of real-time voice conversion on speech behavior

☆ 水上遼人(明治大), 林大輔(日本たばこ産業), 森勢将雅(明治大)

- ◆ リアルタイム声質変換を用いた発話において、発話者の発話行動にどのような違いが生じるのかを検証するために、発話実験を実施した。
 > フィードバック音声の条件は、キャラクター音声3条件と地声条件の計4条件とし、各条件に対して、練習発話、本番発話、主観評価を実施した。
- ◆ 収録した音声ファイルから基本周波数を抽出した結果、地声条件に比べて、キャラクター音声を用いた発話において、基本周波数が変化する可能性が示唆された (Fig. 1)。
- ◆ 主観評価結果と関連付けると、フィードバック音声が高い声と感じた場合には発話の基本周波数が高くなり、低い声と感じた場合には基本周波数が低くなる可能性が示唆された。

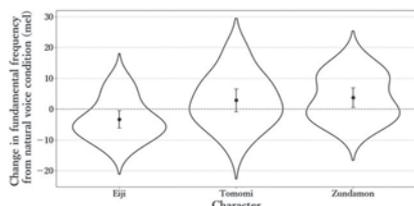


Fig.1: Violin plot showing changes in fundamental frequency across character conditions based on natural voice condition.

Error bars indicate 95% confidence intervals.

1-R-13

1-R-13 話者情報が異なって知覚される 演技音声における発話者の同一性判断: VTuberの動画視聴者を対象とした実験による検証

Speaker verification by human participants with acting voices perceived as different characters: An experimental study on VTuber fans

◎ 林大輔(JT), △ 博衣こより(ホロライブ), 山本峻太, 森勢将雅(明治大)

- ◆ 本研究では、VTuberである第二著者の動画内において、その視聴者に対して実験を行い、年齢や性別などの話者情報が異なって知覚される演技音声であっても、継続聴取により発話者の同一性を認知できるのかについて調べた。
- ◆ 実験の結果、特定の発話者の音声の継続聴取によりその人の声に対する感度が高まり、話者情報が異なって知覚される演技音声であってもある程度は発話者の同一性が認知できることが示唆された。
- ◆ また、継続聴取している発話者の声が提示されることに対する期待が高まることで、他の声も含めて「継続聴取している発話者の声である」という判断をしやすくなることが示唆された。

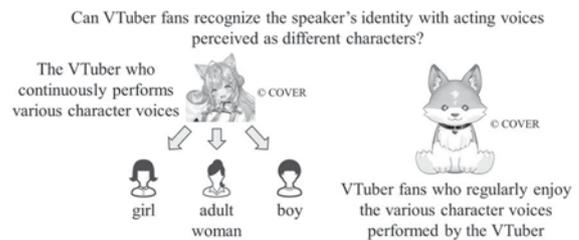


Fig.1 The schematic description of the present experiment

1-R-15

1-R-15 リズムタッピングと発話の相互干渉 発話の周期性に着目して

Mutual Interference between Rhythmic Tapping and Speech Production: Focusing on Speech Rhythm Periodicity

☆ 荒木俊輝(筑波大院), Oli Jan, 橋克輔(産総研), 寺澤洋子(筑波大)

- ◆ 背景と目的: 指によるリズムタッピングと発話の同時遂行時、特に間隔に長短のあるリズムでは相互干渉が生じる。本研究では、タッピングリズムが発話の時間構造に与える影響を検討した。
- ◆ 方法: 16名の実験参加者に対し、リズムタッピング中に文章を発話するという課題を実施した。発話音声に対して Montreal Forced Aligner (MFA) を用いた強制アライメントを行い、母音区間を抽出した。得られた母音系列に基づき、タッピング中の発話リズムを調査するための指標として normalized Pairwise Variability Index (nPVI) を算出した。
- ◆ 結果: タッピング中の発話について、等間隔リズムでは nPVI が低下、非等間隔リズムでは増加する傾向が認められた (Fig. 1) ことから、タッピングのリズムが発話の時間構造に影響を及ぼすことが示された。この結果は、発話とリズム制御の関連を理解する上で重要な知見を提供する。

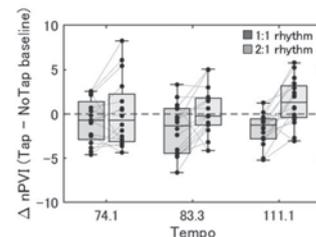


Fig.1 : ΔnPVI relative to the no-tapping baseline for 1:1 and 2:1 rhythm conditions across tempo. Paired dots indicate individual participants.

1-R-16

1-R-16 視聴覚情報の空間的一致性が異なる環境下での発話意識の主観評価および発話音声の音響分析

Subjective Evaluation and Acoustic Analysis of Speech Under Different Conditions of Audiovisual Spatial Congruence
©加藤日花里, △今野智明, 堀内俊治(KDDI 総合研)

- ◆空間の印象と残響が発話に与える影響を明らかにすることを目的とし、大小異なる空間において、両耳室内インパルス応答 (BRIR) を用いて聴覚フィードバックを変化させた際の発話意識の主観評価および音響分析を行った。
- ◆BRIR を用いることで空間の印象を変えられる可能性が示唆された。
- ◆空間大においては、発話レベルと子音と母音の強度比 (CVR) について、いくつかの条件間での有意差が確認された。
- ◆上記は、特に大きな空間においてオーバーラップマスキングの程度が大きいことに起因する発話意識の音声への表れであると考えられる。

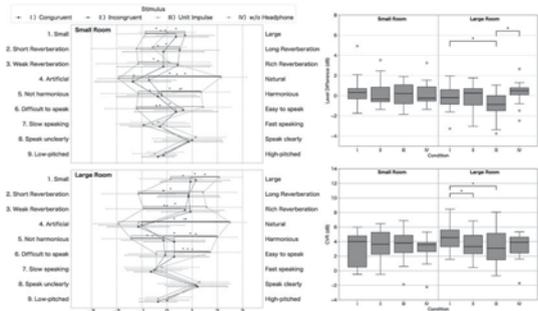


Fig.1: (left) The evaluation scores, (upper right) speaking level difference, (lower right) CVR, *: p < 0.05

1-R-18

1-R-18

雛の採餌時に母鶏が発する鳴き声の合成

Synthesis of maternal hen vocalizations produced during chick foraging
☆ 田中智己, 矢田部浩平, 新村毅 (農工大)



採餌時の母鶏の鳴き声 (Food Call)
→ 雛の恐怖心を低減させる

- 母鶏不在下での母鶏とのコミュニケーション再現

- 従来: 録音したFood Callの再生

Food Callの多様な音響特徴の違いを捉えられない

- 提案: 音響特徴に基づくラベリングを条件付けとするFood Callの生成

1-R-17

1-R-17

鶏の母子間コミュニケーションにおける母鶏の鳴き声の分析

Analysis of maternal calls in mother-offspring communication of domestic chickens

☆ 中谷優太, 矢田部浩平, 新村毅 (農工大)



1-R-19

1-R-19 牛の鳴き声による状態識別および個体識別におけるマルチタスク学習の効果

Effect of Multi-task Learning on State and Individual Identification Using Cattle Vocalizations

☆西尾優亜, △坪倉和哉(愛県大・情) △神谷咲月, △小島秀介, △兼子明美(愛知県農業総合試験場) 入部百合絵(愛県大・情)

- ◆牛の鳴き声を用いて状態識別と個体識別を同時に行うマルチタスク学習手法を提案する。
- ◆牛の発情に伴う鳴き声の変化傾向は個体ごとに異なるため、個体差を考慮した発情検知手法が求められる。そこで本研究では、個体情報と状態情報(発情/非発情)を同時に学習するマルチタスク学習を導入し、発情検知および個体識別の精度向上を目指す。MLP および CNN による識別モデルを構築し、入力特徴量による違い、CNN における事前学習重みおよび入力チャネル数の影響を分析した。さらに、状態と個体を独立に学習するシングルタスク学習との比較からマルチタスク学習の有効性を検証した。
- ◆ Fig1 の結果から、マルチタスク学習では個体識別精度 0.915、発情検知精度 0.861 を達成し、シングルタスク学習と比較し、個体識別6%、発情検知5%の精度向上が得られた。また Fig2 より、13 頭中 10 頭で発情検知精度が向上し、個体単位で確認しても、その有効性が確認された。

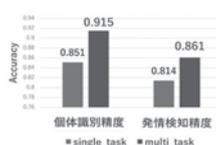


Fig. 1 Single vs. Multi-task Learning Performance

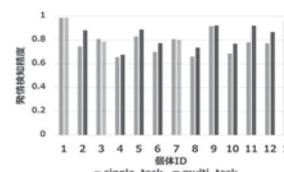


Fig. 2 Estrus Accuracy per Individual (Single task vs. Multi task)

1-R-20

1-R-20 音源方向推定と映像情報を用いた 鶏の発声個体の推定

Estimation of vocalizing chickens
by using sound source localization and video information

☆ 仁平善久, 照沼卓磨, 瀬口瑛子, 矢田部浩平, 新村毅 (農工大)

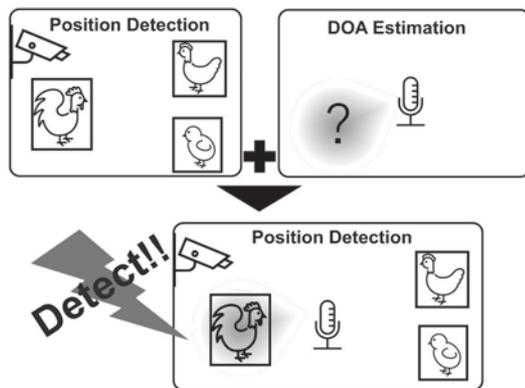
複数の鶏が存在する環境で個々の鳴き声解析を行いたい

従来 音声のみ ... どの個体が鳴いたかわからない

提案 映像情報 ... 物体検出により位置を推定
+
音声 ... ビームフォーミングにより方向推定

映像情報と音声を組み合わせることで、発声個体を特定

▶ 鶏のコミュニケーションの解析への活用を目指す



1-R-22

1-R-22 指さし動作が聴覚空間注意に与える影響

The Effect of Pointing Movements on Auditory Spatial Attention

☆ 李戦 (東北大通研/院情科研), 坂本修一 (東北大通研/院情科研)

- ◆ 本研究では、指さし運動が聴覚空間注意に与える影響を検討した。
- ◆ 被験者は、左右スピーカからの出力音に対する音圧変化(増加, 減少)判断とキューで指示された方向への指さしを行った。
- ◆ 音圧変化判断対象のスピーカの方向と指さしの方向の異同 (Same, Opposite) と、指さしの有無 (Move, Static) を組み合わせ、MovS, StatS, MovO, StatO の4条件を設定した。
- ◆ 実験の結果、指さしが聴覚空間注意を促進する可能性が示された。

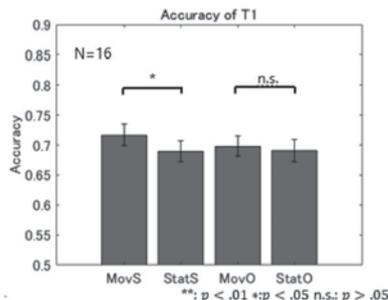


Fig.1: Accuracy for 4 conditions.
(MovS: Move-Same, StatS: Static-Same, MovO: Move-Opposite, StatO: Static-Opposite)

1-R-21

1-R-21 人間のエコーロケーションにおける 背景音の影響

Effect of Masking Sound in Human Echolocation

☆ 勝倉朱音, 亀川徹, 丸井淳史 (東京藝大)

- ◆ 様々な背景音が存在する日常空間におけるエコーロケーションを想定し、背景音が人間のエコーロケーションに与える影響を検証した。
- ◆ 障害物の存在有無を識別するタスクと、障害物の距離を識別するタスクを実施した。背景音の有無、背景音の提示音圧レベル、背景音の到来方向、放射音の種類の違いが与える影響を検証した。
- ◆ 背景音の有無、提示音圧レベル、放射音の種類の違いによる知覚精度の変化は認められなかった。背景音が存在する環境でも、人間はある程度エコーロケーションを行うことができる可能性が示唆された。
- ◆ 背景音の到来方向に関しては、背景音が前方から到来する場合と比較して、後方から到来する場合にエコーロケーションの精度が低下する可能性が示唆された。

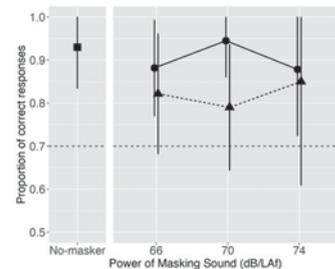


Fig.1: Participants' performance in detecting the object by echolocation under white noise masking sound. Emit sound is whitenoise. (Direction of Masking Sound: ●Front ▲Rear)

1-R-23

1-R-23 VR 環境下における異なる聴覚 ディスプレイによる音像定位精度の比較

Comparison of sound localization performance between different spatial audio displays in a VR environment.

☆ 深谷拓也, 山高正烈 (愛知工科大院)

- ◆ 本研究では、VR 環境下における視覚刺激の提示が、異なる聴覚ディスプレイ (VAD) で提示された音空間知覚に与える影響を検討した。
- ◆ 視覚刺激はUnity VRにて構築し、視覚刺激なし、正面0度、右30度の3条件とした。聴覚刺激は、(1)Panning (2)Ambisonics デコード (Ambi) (3)モノラル音源へのHRTFの畳み込み (Mono+HRTF) (4)Ambisonics デコードへのHRTFの畳み込み (Ambi+HRTF) の4種類を使用し、0~90度までの15度刻みの7方向の何れかに、ピンクノイズを提示した。
- ◆ 実験結果

- 視覚刺激無し条件の場合、提示角度0度以外は、音刺激の提示角度によって音像定位精度がばらつくことが確認された。(Fig.1)
- 視覚刺激が0度の場合、音像定位が全体的に視覚刺激の提示方向にシフトしており、特にAmbi条件でその傾向が顕著に表れた。(Fig.2)
- 視覚刺激30度の場合、視覚刺激0度条件よりも視覚による影響が顕著に表れた。また、正面近傍においてAmbi条件の音像定位精度が低下した。(Fig.3)

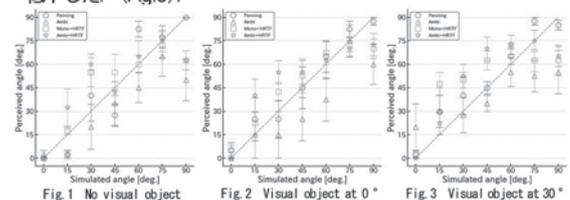


Fig.1 No visual object

Fig.2 Visual object at 0°

Fig.3 Visual object at 30°

1-R-24

1-R-24 正面 0.5m にある音源の放射方向知覚実験における正面音声の特異性の影響について

A study of peculiarities of sounds coming from the front for perception of radiation direction of sound sources located 0.5m in front.

☆大島若菜, △米澤杏瞳 (阪産大・デザイン工).

中山雅人 (立命大・情報理工), △李旭, 高橋徹 (阪産大・情報デザイン)

【背景】発話者の頭部方向推定 (HOE) は、会話開始時の認知や高臨場感音響再生において重要である。特に、正面 0 度の刺激は識別されやすいと報告されている。

【目的】正面 0.5m に位置する発話者に対して、人が頭部方向の角度差をどの程度識別できるかを調査し、0 度刺激の特異性の有無を検証する。

【手段】-30 から 30 度の 7 種類の頭部方向音声を用い、二肢強制選択手法により識別実験を行った。さらに、提示角度を ±2 度シフトした 3 種類のデータセット (E, M, P) を用い、0 度刺激の影響を比較検証した。

【結果】0 度刺激の特異的な識別優位性は確認されず、Fig.1 と Fig.2 でも特徴的な傾向は見られなかった。Fig.3 に示す通り、0 度条件に限定した角度差平均においても、3 データセット間に差は認められなかった。

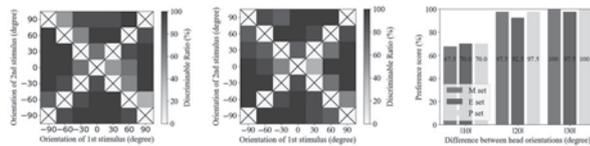


Fig.1: Preference score map between head orientations for neutral dataset (E set).

Fig.2: Preference score map between head orientations for negative shifted dataset (M set).

Fig.3: Mean preference score for each absolute orientation difference.

1-R-29

1-R-29 動的声帯モデルを用いた声門内圧測定実験

Measuring intraglottal pressure of physical model of oscillating vocal folds

☆久米田員男, △上田凌生, △塩谷悠人, 徳田功(立命館大)

- ◆アクリル製 M5 声帯モデルを回転動作させることで、発声時における声門形状の動的変化を再現し、声門内圧を測定した。
- ◆計測の結果、従来の静的モデルでは観測できなかった、圧力分布のヒステリシス特性を確認した。
- ◆圧力分布から算出された声帯作用力にもヒステリシスループが認められ、ループ面積は振動周期が短いほど大きくなることが分かった。
- ◆ヒステリシスの作り出す声帯作用力の差は、声帯が呼気流からエネルギーを得て振動を維持するうえで重要な役割を果たしている可能性がある。

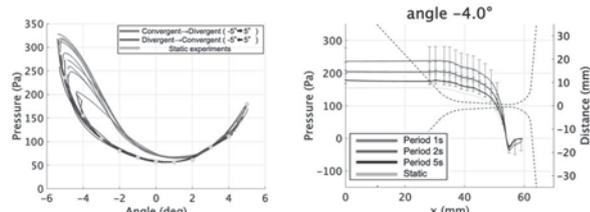


Fig.1: Dependence of subglottal pressure on vocal fold angle. (Period 1s)

Fig.2: Pressure distribution inside glottis. (Convergent shape)

1-R-28

1-R-28 声帯溝症について溝の位置が発声に及ぼす影響

Effect of sulcus location on phonation of sulcus vocalis model

☆竹内理人(立命館大), △葉山奈南(立命館大), 徳田功(立命館大)

- ◆ヒトの声帯を MRI で撮影して得られたデータをベースに、声帯溝症を模擬した物理モデルを作製した。
- ◆溝のない声帯の物理モデルと、4 種類の溝の深さと 3 種類の溝の位置を組み合わせた溝症モデル 12 種類を用意した。
- ◆吹鳴実験を行い、各モデルについて声門下圧・音声・高速度映像を計測し、溝の深さと位置の違いが声帯振動に与える影響を比較した。
- ◆溝ありモデルは溝なしモデルに比べて、オンセット圧が高い傾向を示した。また、溝の位置によって溝の深さが発声に及ぼす影響の傾向が異なった。

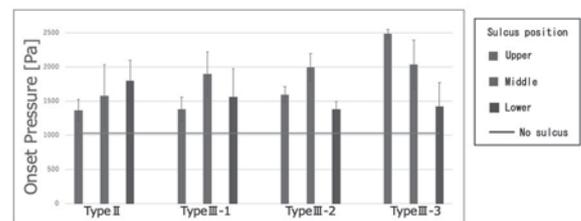


Fig.1: Experimental results of phonation onset pressures.

1-R-30

1-R-30 声帯硬さが声帯-仮声帯振動に及ぼす影響

Effect of vocal fold stiffness on vocal-ventricular fold co-oscillations

☆水本 直宏(立命館大), △谷 良信(立命館大), 徳田 功(立命館大)

- ◆ヒトの声帯と仮声帯の一体型モデルをシリコンで作成し、声帯の硬さが、声帯振動および声帯-仮声帯振動に及ぼす影響を調査した。
- ◆吹鳴実験を通して、声帯部分の硬さが異なる一体型モデルを比較し、振動特性の変化を調べた。
- ◆声帯が硬いほど声帯振動の基本周波数が上昇した。
- ◆声帯が硬くなると声帯振動と仮声帯振動の周波数比が 2:1 から 3:1 に変化した。

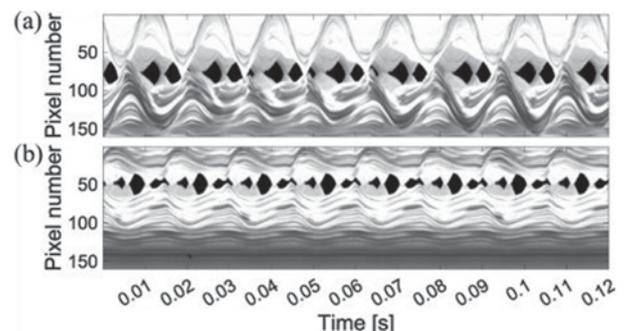


Fig.1: Fig.3 Kymagrams of (a) soft, (b) hard models.

1-R-31

1-R-31 マカク属における声帯および声帯膜のヤング率測定

Measurement of Young's moduli of vocal folds, and vocal membranes in macaque monkeys

☆杉江昂輝, 山本康旭, △山本憲隆(立命館大)
△西村剛(大阪大), 徳田功(立命館大)

- ◆アカゲザル4個体およびニホンザル2個体の喉頭から摘出された声帯、声帯膜に対して一軸引張試験を実施した。
- ◆応力-ひずみ線図の、発声開始ひずみを考慮した範囲 ($0.05 \leq \epsilon \leq 0.15$) で線形近似し、ヤング率 E を算出した。
- ◆個体間のばらつきが大きい、多くの個体で声帯と声帯膜のヤング率が近い値を示した。
- ◆声帯膜を用いた発声でみられる基本周波数の上昇に、弾性特性の差以外の要因が関与することが示唆された。

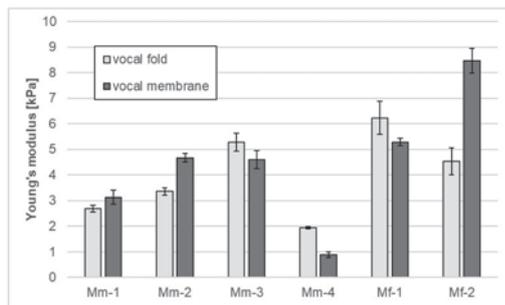


Fig. 1 Young's moduli of vocal folds and vocal membranes measured from four rhesus macaques and two Japanese macaques. (Mm: rhesus macaques (*Macaca mulatta*), Mf: Japanese macaques (*Macaca fuscata*)

1-R-33

1-R-33 キンカチョウのマイクロCT画像を用いた鳴管物理モデルの構築

Construction of a physical model of syrinx using micro-CT images of zebra finch

☆藤江真己人(立命館大), △Coen P.H. Elemans (南デンマーク大学), 徳田功(立命館大)

- ◆キンカチョウ (zebra finch) 鳴管のCT スキャンデータに基づき、外側ラビア (LL) および内側ラビア (ML) から構成される物理モデルを構築した。
- ◆鳴管直前の気管支圧 (p_b) と鳴管を囲む気嚢圧 (p_a) を2台の圧力コントローラーによって独立に制御し、物理モデルを吹鳴させた。
- ◆物理モデル実験では、気嚢圧により LL が内転し、気管支圧が気嚢圧よりも高くなったときに、ML と LL が振動する様子が観測された。
- ◆摘出鳴管吹鳴実験においても、気管支圧が気嚢圧より高いときに、音声が生産された。
- ◆鳥類の発声において、気嚢圧と気管支圧の相互関係が、発声の有無を決定していることが分かった。

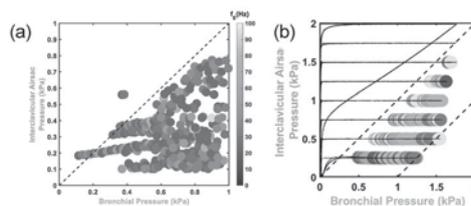


Fig.1: Fundamental frequencies produced in the pressure control space of (a) physical model, (b) excised syrinx of zebra finch

1-R-32

1-R-32 膨張型の喉頭嚢が物理モデルの発声に与える影響について

The effect of an expandable air sac on a physical model of the vocal folds

☆松原大輔, 廣野凌, △大岸幸誠, 徳田功(立命館大), △西村剛(大阪大)

- ◆サル類に見られる、開口時に喉頭嚢の膨張を維持するメカニズムの妥当性を検証するため、チンパンジー等の喉頭形状を模した物理モデルを使用し、3つの条件下: (a)声道開口部の閉鎖、(b)喉頭嚢開口部の閉鎖、(c)負圧による伸長、で実験を行った。
- ◆開口部を閉鎖する条件(a)および(b)では、発声に極めて高いオンセット圧が必要となり、現実的ではないことが示された。
- ◆一方、周囲の組織による伸長を模した条件(c)では、1.7 kPa という妥当なオンセット圧が得られた。これにより、喉頭嚢の膨張維持には筋肉等による能動的な伸長機構が不可欠である可能性が示唆された。
- ◆喉頭嚢の付加およびその容積の変化が音圧レベル (SPL) に影響を与えることも確認され、喉頭嚢が発声の音響特性を制御する重要な役割を担っていることが示された。

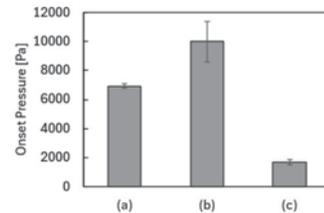


Fig.1: Onset pressure when the air sac is expanded under 3 conditions: (a) Constriction was created at the open end of the vocal tract. (b) Constriction was created at the opening of the air sac. (c) Air sac was expanded in a container with negative pressure.

1-R-34

1-R-34 ハシトガラス発声器官の三次元形状データに基づく声道モデルの生成

Generation of a vocal tract model based on three-dimensional shape data of the vocal tract of *Corvus macrorhynchos*

☆木原優太, 高橋義典(工学院大・情報)

- ◆ハシトガラスの「ka」コールのフォルマント分析 (Fig. 1)
 - 「ka」コールの波形は前半と後半に分割され、前半と後半でフォルマント周波数が異なる (Table 1)。
- ◆ハシトガラスの声道模型の制作
 - ハシトガラスの死体から気管部分と口腔部分をシリコンで型取りして、3D スキャナで 3D モデルを作成した。
 - 気管部分と口腔部分の断面積の実測値に基づいて、3D プリンタで声道模型を作成した。
- ◆声道模型によるフォルマント周波数の推定 (Fig. 2)
 - 声道模型のインパルス応答を測定し、共鳴周波数からフォルマント周波数を推定した (Table 1)。

Table 1. Estimated formant frequency

	1 st Formant	2 nd Formant
"ka" call first half	1142 Hz	2670 Hz
"ka" call latter half	1063 Hz	2259 Hz
Vocal tract model	1162 Hz	2687 Hz

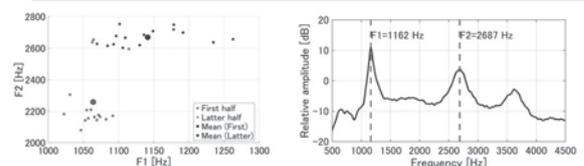


Fig. 1: Scatter plot of the 1st and 2nd formant frequency

Fig. 2: Results of the resonant characteristics of the vocal tract

1-R-35

1-R-35 ネコの声の言語表現について —英語話者へのアンケート調査から—

Representations for cat sounds:
A questionnaire survey of English speakers

○横山安紀子(日本大学)

- ◆本研究では、ネコがノドを鳴らす時の声の表現について、英語話者を対象としてアンケート調査を行った。これは、動物の鳴き声が人間言語の「ことば」として表現されるまでの過程と要因を解明することを目的とする研究の一部として行ったものである。
- ◆調査対象者は20代、30代、40代、50代、60代の米国在住の成人英語話者である。性別、年代別に、ネコの飼育経験の有・無による群に分け、各群25人ずつ、合計500人からオンラインによる回答を集計し、その結果を比較した。
- ◆結果としては、英語辞書項目である“purr”および /p+/r/ を含む語が飼育経験の有無にかかわらず、高い回答率を示した。男女差、年代差について有意差は認められなかった。
- ◆日本語話者を対象とした先行研究では、/g+/r/ を含む語の回答率が顕著に高かった。それと比較すると、英語話者は、動物の声に近い音を「ことば」として表現する傾向にあり、そこには聴覚情報に基づいた語の生成過程があると考えられた。
- ◆動物の声を知覚し表現する場合、異なる言語話者による形式のちがいが見られるが、それぞれの言語において、様々な要因がその過程には関与していることが考えられ、「ことば」の生成については音声情報だけでなく、多角的な視点から、話者が受容する情報を検証する必要があることが示唆された。

1-R-37

1-R-37 Text-to-Speech モデルによるアライメントを用いた音素列からの調音運動データの生成

Generation of Articulatory Movement Data from Phoneme Sequences
Using Alignments with Text-to-Speech Model

☆柳翔太(東京理科大)、澤田隼(日本工業大)、大村英史、桂田浩一(東京理科大)

調音運動を観測する方法の1つにreal-time MRI (rtMRI)を用いる方法がある。rtMRIは測定範囲が広く、声道全体を捉えることができる。しかし、rtMRIを用いた調音運動の観測は特殊な機材を用いるため収録は容易ではない。そのため、他のモダリティ、特に言語情報からrtMRI動画像を推定する研究が行われている。言語情報から調音運動を推定するには、離散的な言語情報を時系列データである調音運動 (rtMRI 動画像) に対応付ける必要がある。既存の手法では、rtMRI 動画像と同時に収録された音声に対して強制アライメントを行うことで、音素継続長を取得している。この情報を用いて音素とrtMRIフレームを対応させている。そこで本研究では、Text-to-Speech (TTS)のモデルであるAlignTTSから得られるアライメント情報を用いて音素継続長を推定し、その継続長を用いて音素列からrtMRIデータの推定を行う。提案モデルの構成をFig.1に示す。

実験では、強制アライメントを用いる既存の手法と比較して同程度の品質の動画像の推定が可能であることを確認した。また、音素継続長の推定に音声を利用しない調音運動の推定を可能にした。



Fig.1: Structure of rtMRI video model from phoneme sequence.

1-R-36

1-R-36 動画生成 AI による rtMRI 動画のアップスケールの試み

A study on real-time MRI movie upscaling using image processing artificial intelligence

☆藤木 祐輔, 北村 達也(甲南大), 前川 喜久雄(国語研)

- ◆リアルタイム MRI (rtMRI) による調音運動観測における時間分解能は27 fpsにとどまり、子音などの速い調音運動の観測には不十分。
- ◆深層学習をベースにした画像処理 AI を活用することによって、時間分解能を向上させることができれば、MRI 装置の性能の制約を超えることができる。そこで、2つの手法でrtMRI動画の時間補間を試みた。
- ◆Topaz Studio (Topaz Labs 社) バージョン 1.0.4 の frame interpolation 機能では、舌運動を正しく補間することはできなかった。
- ◆FILM (Reda et al., 2022) による補間では、2フレームの舌運動の中間段階が得られ、調音運動のrtMRI動画の時間分解能向上に利用できる可能性を示した。

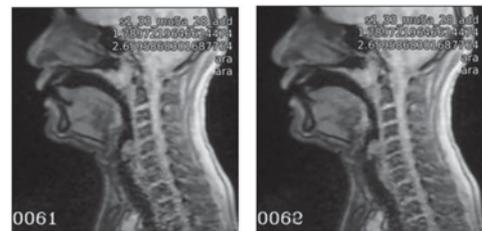


Fig.1:(left) Original frame and (right) frame interpolated by FILM (Reda et al., 2022).

1-R-38

1-R-38 TTS 合成音声による謝罪の許容度向上に関する検討:

人間話者との基本周波数入れ替え

Evaluation of sincerity for speech of apology:
Does human speaker F0 improve TTS synthesized speech sincerity?

☆澤田拓郎, 菅野聖真, 松井淑恵(豊橋技科大)

- ◆人間らしい謝罪を TTS 合成音声に行わせるために、人間話者による謝罪音声のピッチパターンを TTS 合成音声に移植することを試みた。
- ◆収録した人間話者の音声と、生成した TTS 合成音声をそれぞれ WORLD で分析し、音声モーフィングを実施した。基本周波数 F0 のモーフィング割合のみを変化させることで人間話者の抑揚を TTS 合成音声に移植した。
- ◆今後の聴取実験で、人間話者の自然な抑揚を TTS 合成音声に持たせることで合成音声による謝罪の許容度が向上するか検証する。

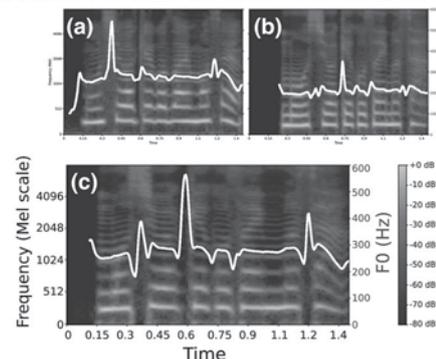


Fig.1:Spectrograms of (a) human speech, (b) TTS synthesized speech, (c) TTS speech with human F0. The white line indicates the F0.

1-R-39

1-R-39 IP通信におけるパケット損失が日本語音声の聞き取り精度に及ぼす影響の検討
 The Effects of Packet Loss in IP Communication on the Intelligibility of Japanese Speech

☆Huang Yixuan, 森永 実季, 竹内 朗, 坪谷 総志, 細谷 勇斗 (Studio Infinity)

- ◆本研究は、IP通信におけるパケット損失が日本語音声の知覚に及ぼす影響について、書き起こし精度に加え、聞き取り過程に着目して検討した。従来用いられてきた単回聴取後の誤り率では、劣化音声に対する聞き手の負荷や改善行動を十分に捉えられない可能性がある。そこで本研究では、複数回の聴取を許容する課題形式を採用した。
- ◆実験参加者は、各音声刺激を最大6回まで自由に聞き直ししながら逐次書き起こしを行った。刺激は、パケット損失率0%、10%、20%、40%の4条件で提示し、初回聴取時の文字誤り率 (CER_{first})、最終書き起こし時の文字誤り率 (CER_{last})、聞き直し回数 (Final Listen Count: FLC)、および回答時間を分析指標とした。
- ◆その結果、CER_{last} は CER_{first} に比べて PLR の違いをより反映し、特に低いパケット損失条件において感度の高い指標であることが示された。また、聞き直しによる精度向上は主に3~4回目までに生じ、それ以降では有意な改善は見られなかった。さらに、CERのみを用いたモデルよりも、CER_{last} に FLC を加えたモデルの方が PLR の違いをより適切に特徴づけることが明らかになった。これらの結果は、初回の聞き取り結果のみを用いるよりも、参加者が自身の判断に基づいて聞き直し回数を決定した後の最終的な聞き取り精度と、その聞き直し過程を併せて評価することが、音声劣化条件を特徴づける上で有効である可能性を示している。

1-R-41

1-R-41 音声による説得において話者の表現と聴取者の特性が果たす役割: Hall (1980) の実験の再検討

The role of speaker expression and listener characteristics in persuasion through speech: A reexamination of Hall's experiment (1980)

☆三石竜勢, 澤田拓郎, 菅野聖真, 松井淑恵 (豊橋技科大院)

- ◆Hall (1980)では、音声による説得の成功は、優れた発話者と優れた聴取者の組み合わせで起こること、また、説得中の発話者の感情的なトーンに対して、聴取者間で感情的な反応が異なる可能性があることが報告されている。一方で、音声の加工の条件が不明確で再現が難しいことから、Hallの実験結果の解釈には検討の余地が残されている。
- ◆Hallらの実験の再現を目指して、説得音声の収集とその評価を行った。
- ◆実験の結果、説得音声の評価は全体的に低かったが、評価が異なる説得音声を集めることができた。今後、話者の説得力と聴取者の感情弁別能力の組み合わせが説得の成功に与える影響を測定する。

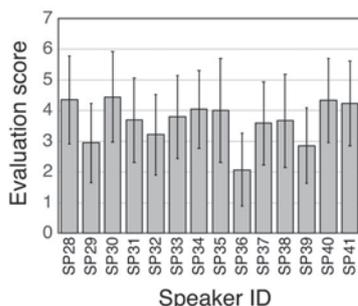


Fig. 1: Mean and standard deviation of persuasiveness ratings for each speaker.

1-R-40

1-R-40 宛先のない発話の音響的特徴と発話者のノンバーバル行動

Speech features and nonverbal behaviors of unaddressed utterances

☆佐藤 匠, 森 大毅 (宇都宮大)

- ◆本研究では、人間が発話の宛先の有無を判断している手がかりを調査するために、発話の基本周波数(F0)と発話者の視線を分析する。
- ◆宛先の有無による平均F0の分布の違いを示した(Fig. 1)。
- ◆宛先のない発話(独り言)に対し談話における機能を定義し、機能の違いが平均F0に与える効果を線形モデルにより分析した。その結果、感情を表出する独り言は平均F0が低く、自分を勢いづける独り言は平均F0が高い傾向があった。
- ◆宛先のない発話と視線の関連について分析を行った結果、独り言では話し始めに視線を逸らす傾向がより強いことが示された。

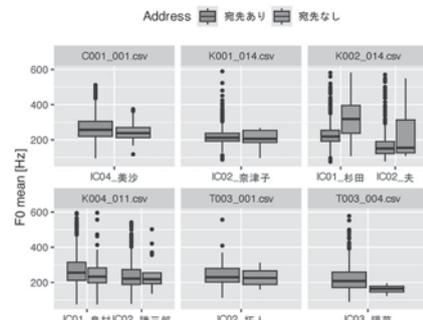


Fig.1: Distribution of F0 Mean

1-R-42

1-R-42 発話内容の違いと声質の変化が対人評価へ与える影響: オンライン聴取実験による調査

The impact of differences in speech content and changes in voice quality on interpersonal evaluation: An investigation using an online listening experiment

☆菅野聖真, 松井淑恵 (豊橋技科大院)

- ◆リアルタイム会話で声質変換(VC)を用いると、仕事仲間としての魅力である道具の魅力度が低下した。この道具の魅力度低下に、発話内容の差異とVCによる声質の変化が与えた影響を測定するオンライン聴取実験を実施した。
- ◆声質は同一で、発話内容が異なる音声同士の比較を行った実験では、対人魅力度の評価に有意な差は見られなかった。
- ◆ほぼ同一の発話内容で、声質をVCで変化させた音声とその元音声の比較を行った実験では、道具の魅力度と、好意や関心の対象としての評価である情緒的魅力度の両方に有意な差が見られた。
- ◆先行研究で報告したVCの利用による道具の魅力度の低下は、発話内容が条件間で異なることよりも、VCによる声質の変化が原因である可能性が高いことが示唆された。

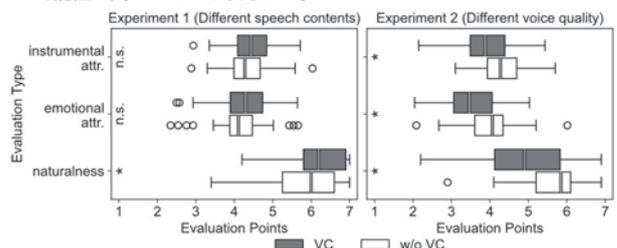


Fig. 1: The evaluation result of interpersonal attractiveness and speech naturalness. (white vertical line: median, open circle: outlier, asterisk: p < 0.05)

1-R-43

1-R-43 感情強度を段階的に制御した 音声刺激を用いた認知症高齢者の 音声認知に関する予備的検討

A preliminary study of emotion recognition in older adults with dementia using speech stimuli with graded emotional intensity

○齊藤千晶(認知症介護研, 名古屋市立大院), 中村篤(名古屋市立大院)

- ◆我が国では認知症者数が増加しており、疾患の早期から感情内容の理解に困難を生じることが指摘されている。一方で、感情の強さの違いによる影響は十分検討されていない。
- ◆本研究では感情の強さを精緻に制御した音声刺激を作成し、認知症者の感情認知の特徴を明らかにすることを目的として、認知症高齢者2名を対象に、サーストンの一対比較法を用いた予備的検討を行った。
- ◆検査方法の feasibility については、検査時の対象者の様子から課題理解の持続性、検査の遂行度合い、負担感の観点から評価を行った。その結果、認知症高齢者においても本検査の実施は一定程度可能であることが示唆された。一方で、認知機能障害の程度によっては課題遂行が困難となる場合もあり、検査適用には事前に認知機能の状態を考慮する必要性が示された。
- ◆感情音声の認知においては、総平均正答率は怒りの音声よりも喜びの音声で高い傾向が認められた。怒りの音声では、強度の高い刺激であっても必ずしも怒りと判断されず、喜びの方が段階に応じて選択される傾向がみられた。今後は対象者数を増やし、感情の種類や強度、認知機能との関連について多角的な検討が必要である。

1-R-45

1-R-45 日本人中学生の英単語音読音声における L・R 置換の自動評価

Automatic evaluation of L/R substitutions in isolated English words read by Japanese junior high school students.

☆池田華奈 加藤恒夫(同志社大) △菅原真理子(同志社大) △山田玲子(ATR)

筆者らのプロジェクトは、日本人の英語音声に児童期の言語環境が与える影響と経年変化を明らかにするため、様々な言語的背景をもつ中学生の英語音声を継続的に収集している。小学生期に、日本国内で日本語による教育を受けた生徒 (JP)、日本国内で日本語・英語のバイリンガル教育を受けた生徒 (BP)、日本国外で英語による教育を受けた生徒 (EP) の3群に分類した。本研究では、収集済みの英語音声データを用いて、/l/ と /r/ の置換率を調べた。中学生51名を対象に、中学1年時と中学3年時の2回、/l/・/r/ を含む計17単語の英語音読音声を収録した。また、英語母語話者 (NS) と比較した。強制アライナ (MFA) を用い、辞書に /l/ と /r/ を置換した対立仮説を追加登録することで、強制アライメント時に置換の有無を自動判定した。

結果として、全単語を総合すると、中学1年時、3年時ともに、JP群の置換率が最も高く、次いでBP群、EP群の順となり、NS群は皆無であった。統計的にもJP群は他群より有意に高かった。また、全体的な経年変化はいずれの群でも有意ではなかった。音素環境別に見ると、JP群の単独頭子音の /l/ の置換率は中学3年時でも改善が見られなかった一方、/r/ の置換率は中学3年時に減少した。

これにより、小学校時の言語環境が /l/ と /r/ の発音に与える影響は大きいことが示唆された。音素環境別では、JP群にとって単独頭子音の /r/ の発音は /l/ よりも難しいことが示唆された。

1-R-44

講演取消

1-R-46

1-R-46 日本語音声基盤モデルは英語をどう聞くか？ — 音韻知覚の分析 —

How do Japanese-based speech models hear English?
— An analysis of phoneme perception —

☆白井透, Bagus Tris Atmaja, Sakriani Sakti(NAIST)

- ◆自己教師あり学習モデル (SSL モデル) はラベルなし音声から表現を学習し、大規模音声データによる事前学習を通じて、音声認識などのタスクで高い性能を示している。
- ◆SSL モデルが獲得する音声表現がどの程度学習言語に依存しているのかは十分に明らかになっていない。
- ◆本発表では、英語音声および日本語音声で学習された SSL モデルが英語音声をどのように知覚しているのかを明らかにするために、音素レベルの知覚を分析する。
- ◆ABX テストを用いた二つの実験を行う。
 - 英語音素の弁別性能の評価
 - 日本語話者特有の母音挿入を伴う子音連続音の弁別性能の評価

モデル	事前学習		ABX 正答率 (%)	
	言語	音声データ量 (h)	英語音素弁別	子音連続弁別
HuBERT	英語	960	89.6	81.2
wav2vec 2.0	英語	960	87.8	78.5
kushinada	日本語	62,215	88.2	79.8
izanami	日本語	5,315	85.0	80.2

Fig.1: Pre-training configuration and ABX results

1-R-47

1-R-47 百人一首の決まり字における音響的特徴と時間的決定境界の基礎検討

A Time-Frequency Analysis of Acoustic Features and Temporal Decision Boundaries of Kimariji in Hyakunin Isshu Recitation Speech

©近藤碧, 浦川千歌, 大島風雅, 中山仁史(広島市大院)

- ◆競技かるたの「決まり字」は勝敗に直結するが、確定直前の微小な音響変化は客観化しにくい。
- ◆本研究では、決まり字付近の音響的变化を手掛かりに候補札を早期に絞り込む枠組みを検討する。
- ◆「あ」から始まる16首の読上げ音声を収録し、時間周波数解析により決まり字近傍の瞬間的な変化を確認した。
- ◆CTC デコーダ出力から候補文字列ごとの累積スコアを更新し、候補間正規化によりグローバル事後確率の時間推移を算出する。
- ◆グローバル事後確率と局所変化指標を統合し、候補が1つに収束する時刻を時間的決定境界として推定・可視化する。

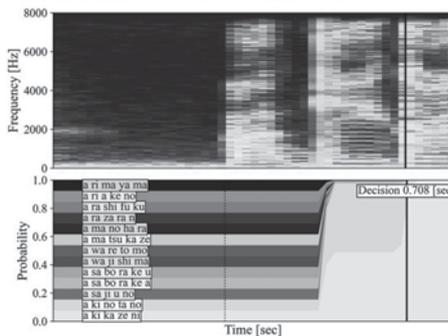


Fig. 1: Example of the temporal decision boundary for a Kimariji

1-R-49

1-R-49 教師なし音節発見モデルを用いた子どもの年齢別発音バリエーションの分析

Analyzing Age-related Phonetic Variation in Children's Speech Using Unsupervised Syllable Discovery Model

©堀井こはる, 俵直弘, 小川厚徳, 荒木章子 (NTT)

- ◆子どもの成長に伴う発音バリエーションの分析は、言語発達の理解を促進し、高精度な子ども向け音声認識器の開発にも応用可能である。
- ◆本研究では、最新の教師なし音節発見モデル **Sylber** を用いて、5歳から15歳の子どもの発音バリエーションの変化を、年齢間の相対比較に基づき、データ駆動的に明らかにした。
- ◆教師なし音節発見は SSL モデルから得られる音響特徴量を教師なしでセグメンテーションし、セグメントの平均特徴量をクラスタリングすることで、音声から音節クラスタを発見する手法である。
- ◆成人音声で学習された Sylber による音節境界検出は子ども音声に対しても一定の精度が確保されており、本手法の有効性が示された。
- ◆年齢別の音節クラスタ数は、5歳から7歳にかけて増加し、8歳以降は減少に転じる山型の推移を示した (Fig.1)。この傾向は、7歳頃までの音節習得と、それ以降の発音の安定化・収束という言語発達の過程をデータ駆動的に捉えたものといえる。

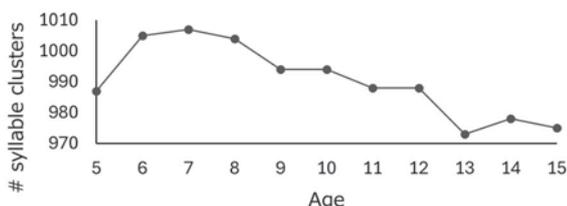


Fig. 1: Number of syllable clusters for each age.

1-R-48

1-R-48 The roles of early exposure in the development of perceptual ability and phonolexical representations in L2 English

OJun Nomura, John Rothgerber (Kwansei Gakuin University)

- ◆Goals: The present study aimed at examining early learners'
 - discriminatory ability for "difficult" sound contrasts and
 - accuracy of the phonolexical representations of words involving these contrasts.
- ◆Method
 - Three types of contrasts were tested: (i) B-V, (ii) Tense-Lax (/l/-/ɪ/), and (iii) Cluster (e.g. /glæd/-/ gu' læd/).
 - 5 early learners, 25 late learners (divided into Upper and Lower groups), and 13 native speakers of English participated.
 - Discriminatory ability was tested using an oddity task (OT).
 - Accuracy of phonolexical representations was tested using a lexical decision task (LDT).
- ◆Results
 - OT (discrimination): Early learners' performance was close to native speakers' (Fig. 1).
 - LDT (phonolexical representations): Early learners' performance depended on the type of contrast; it was similar to late learners on the Cluster condition (Fig. 2).

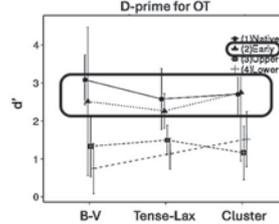


Fig. 1 Performance on the OT.

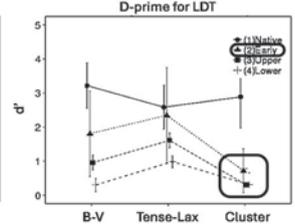


Fig. 2 Performance on the LDT.

1-R-50

1-R-50 対話における加齢障がいの課題事例

Examples of Age-related Disability Challenges in Dialogue

○市川薫, 堀内靖雄(千葉大)

日本は超高齢者社会(65歳以上人口21%以上)を超え、重高齢者社会(後高齢者人口>前期高齢者人口)を突き進み、加齢による障がい対応が大きな課題になってきている。特に「対話」に必要な視聴覚機能の低下は認知症などの原因として注目される。

人体の視聴覚器官は「対話のこぼれ」が必要とする「実時間性」などの「要求条件」を満たすため「対数化特性」や「認知的実在処理機能」などの巧妙なシステムを構成している。しかし、加齢による要素機能の低下からの感覚過敏症状や認知機能の低下から実時間対話に支障をきたしている。また日常生活を取り巻く衝撃音(例えば食器間の衝突や扉などの開閉)のストレスは当事者にしか理解できない障がいである。

対する支援機器として例えば補聴器があり、最近ではAIの導入などが試みられているが、聴覚器官の機能低下とそれに対する対応策、効果を対応付けて定量的に比較・理解可能なようには必ずしも表現されていない。価格も30万円から150万円近くになり、各社の機器を試用・比較できるように補聴器店で用意することも難しいであろう。多くの当事者は加齢障がいから生活の質(QOL)を大きく低下させている。ユーザの視点からの国際的な規格が必要と考える。医療から工学までの幅広い領域に対する課題である。

1-R-51

1-R-51 ゲームを取り入れた協調的な訓練を可能にするチューブ発声支援システム

Tube phonation support system enabling collaborative training through game-based activities

☆村井武人(甲南大院), 北村達也(甲南大), 川村直子(姫獨大)

- ◆本研究は、チューブ発声を用いた音声リハビリの研究である。
- ◆チューブ発声では患者が安定して発声訓練を行える方策が求められている。そこで、2人で同時操作可能なゲームを取り入れた発声訓練支援システムを開発し、評価実験を行った。
- ◆実験では、健康な高齢者にシステムを利用させ評価を行った。また、訓練前後の音声を録音し、音声に対する即時的効果を調査した。
- ◆実験結果より、システムはチューブ発声の継続に対する意欲に繋がる効果が確認できた。また、複数人で同時に操作を行うことによる相手の存在が、チューブ発声に対するモチベーションの維持に寄与する可能性が確認できた。
- ◆録音した音声に対して声質分析を行い、Jitter, Shimmer, HNRの変化を調査した。しかし、参加者の音声には明確な声質の変化は見られず、即時的効果は確認できなかった。



Fig.1: Multiplayer serious game.



Fig.2: Scene from the experiment.

1-R-54

1-R-54 構音障害音声のリアルタイム変換デバイスの原音入力特性の改善

Improvement of Source Sound Input Features of Real-Time Conversion Devices for Articulatory Disordered Voice

○藪謙一郎, 田中敏明, 伊福部達(東大)

- ◆筆者らは、指の位置や動きで声を楽器のように音韻を操作できる音声生成器を提案しており、この方式で、発声機能が残存する構音障害者が話者自身の声を活かしてリアルタイムに音韻を付加できるシステムを開発してきた。
- ◆本発表では、先行研究で得られた課題抽出に基づき、原音の入力音について改良を行い、特に、マイクの装着状態の改善や、入力音の周波数特性を考慮したアナログ特性の改善を行って、安定した変換音声出力を可能としたので報告する。

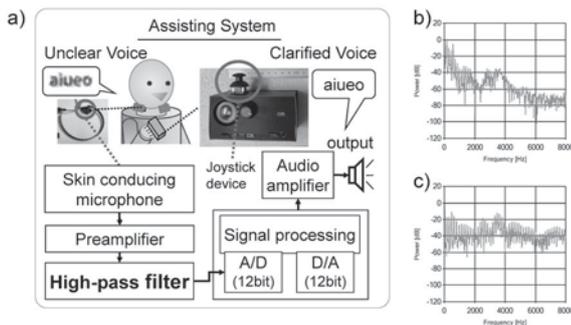


Fig.1: (a) Structure of improved speech assistance system device, (b) input voice before improvement, and (c) input voice after improvement

1-R-52

1-R-52 高齢者の日常会話におけるトピックの特徴について

Topic-shift Characteristics in Elderly Daily Conversations

☆今中伸顕(大阪工大), 脇田由実(大阪工大) △松田千登勢(摂南大)

- ◆高齢者が飽きずに習慣的に使用できる会話支援システムを実現するため、高齢者の日常会話の特徴を、年齢や生活環境(在宅か施設か)の違いを考慮しながら分析した。
- ◆単位時間当たりの発話数および会話内の異なり名詞比率は、在宅高齢者に比べ施設高齢者が少なく、施設高齢者は少ない発話の中で同一名詞を繰り返す傾向があり、新たなトピックを展開しにくい状況であった。またトピック分析の結果、前期高齢者はトピックが決まると深く話すが、後期高齢者および施設高齢者は比較的浅い傾向があることもわかった。ただこれらの特徴は、施設高齢者では個人差が大きく、会話に依存していた。
- ◆これらの知見は、高齢者の属性に応じてトピック内容や提示タイミングを調整する会話支援システム設計の有効な指針の一つとなる。今後はデータ数を増やし、知見の信頼性を確認していく。

