

3-1-1

3-1-1 実測信号とシミュレーション信号の残差最小化を目的とした深層学習に基づく光レーザーマイクロホン受信信号の擬似生成

Simulation of optical laser microphone signals based on deep learning for minimizing the residual between measured and simulated signals

☆中野裕貴(立命館大院), 張宇涛, 耿毓庭, 中山雅人, 西浦敬信(立命館大)

- ◆光レーザーマイクロホンは、音により振動している物体に対してレーザー光を照射し、音の計測を行うシステムである。
- ◆光レーザーマイクロホンを用いた音声計測では、振動物体の過渡応答等に起因する非線形歪みや雑音混入により受信信号が劣化する。深層学習に基づく音声処理の適用には大量の学習データが必要となるが、網羅的な実データ収集は困難である。
- ◆本稿では、学習データの拡張を目的とし、光レーザーマイクロホンの擬似受信信号生成手法を提案する。提案手法では、劣化要因をインパルス応答による線形歪み、深層学習モデルによる非線形歪み、および自己回帰モデルによる定常雑音に分解してモデル化する (Fig. 1)。評価実験では、生成した擬似受信信号と実測信号の類似性を評価した。

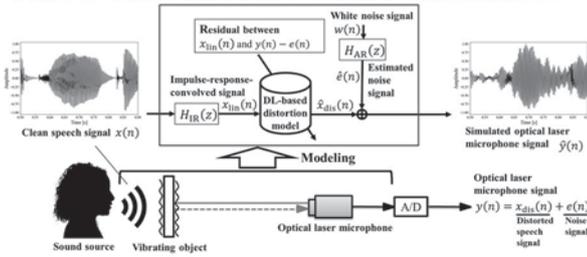


Fig. 1: Overview of the proposed method

3-1-3

3-1-3 帰還系相殺誤差が Filtered-x LMS 法の動作を不安定化させる原因

Factors on the feedback path cancellation error destabilizing the performance of filtered-x LMS algorithm

○藤井健作(コダウェイ研), 菅木禎史(千葉工大)

フィードフォワード型能動騒音制御システムに Filtered-x least mean square (LMS) 法の適用を想定する本検討では、距離減衰によって帰還系の利得は 1 未満になるとし、これによって制御が発散することはないとする。但し、このような場合でも高い騒音低減効果を実現するためには帰還系は一定程度相殺されていること、2 次系についても一定程度の誤差で推定されていることが求められる。

問題は、帰還系相殺誤差が無視できない場合に 2 次系が修正 2 次系に置換えられ、その結果として騒音制御の発散の抑制には 2 次系推定誤差よりも帰還系相殺誤差を小さくすることが重要となる。更に、実用時に想定される音響系の変化に対応するために導入される 2 次系と帰還系の常時推定機能には推定の完了を判定する処理が必要になる。これは、音響系変化の後に行われる騒音制御フィルタの係数更新は、その判定を待ってからしか行えないことを意味する。

Filtered-x LMS 法を安定して実行できるようにするためには、これ以外にも様々な機能の追加が必要になる。これらは全て、同法が有効となるために必要な条件「騒音制御フィルタの係数が緩やかに更新されること」を満たすための機能である。これに対して本検討では、これらの機能が不要なく、制御を不安定にする要素もない連立方程式法の採用が現実的な選択と言えることを示す。

3-1-2

3-1-2 デュアル遠方参照マイクロホンを用いたマルチチャネルフィードフォワードアクティブノイズコントロールシステムの実騒音環境評価

Performance evaluation of multichannel feedforward active noise control system with dual distant reference microphones in real environments

☆水谷真絃(立命館大院), 岩居健太(阪産大), 中山雅人, 西浦敬信(立命館大), 添田喜治(産総研)

- ◆フィードフォワードアクティブノイズコントロール(FFANC)システムでは、因果律の制約により誤差マイクロホン近傍に参照マイクロホンを設置することが困難である。
- ◆そこで、光レーザーマイクロホンを用いた FFANC が提案されているが、誤差マイクロホン近傍での残響成分増加により、参照信号と一次騒音のコヒーレンスが低下し、騒音低減性能が低下する。
- ◆本稿では、参照マイクロホンとして気伝導マイクロホンを追加することで、因果律の制約緩和と高コヒーレンス参照信号の取得を同時に実現するデュアル遠方参照マイクロホンを用いたマルチチャネル FFANC システム (Fig.1) を提案し、実騒音環境下での実実験により評価する。
- ◆工場騒音を用いた実実験の結果、複数実環境下で提案システムが最も高い騒音低減性能を持つことが示された (Fig.2)。

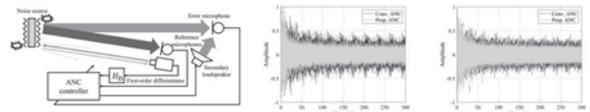


Fig. 1: Structure of the proposed ANC system

(a) Soundproof room (T60 = 100 ms) (b) Office environment (T60 = 300 ms) Fig. 2: Waveforms of error signals

3-1-4

適応フィルタにおける適応学習率アルゴリズムの検討

Investigation of adaptive learning-rate algorithms for adaptive filtering

◎内田蓮, 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)

背景: 適応フィルタ

- 適応フィルタ出力と所望信号の誤差を最小化
- NLMS: ステップサイズの設定が容易であり、広く用いられている
↓ 一般化
- V-APSM: パラメータ空間の幾何構造 (計量行列) を適応的に更新

提案 ①: 周波数領域 Adam (SpecAdam)

- 適応学習率アルゴリズム: 勾配降下法の収束性能を向上
- Adam: 次元間の相関を無視した勾配の補正
↓ 周波数領域で導出
- SpecAdam: 次元間の相関を効率的に考慮可能

提案 ②: 計量行列を適応学習率アルゴリズムで設計

- 計量行列として, Adam, SpecAdam の前処理行列を適用
- インパルス応答推定実験の結果 (Fig. 1) から, 有効性を確認
▶ 計量行列の設計には, 次元間の相関を考慮することが重要

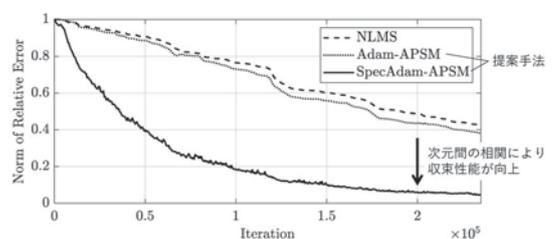


Fig. 1: Change of the relative error in impulse response over iterations.

3-1-5

3-1-5 マトリョーシカ埋め込みを用いたオーディオエフェクトチェーン推定 Effect Chain Estimation Using Matryoshka Embedding

☆和田仰(東大/産総研), 中村友彦(産総研), 猿渡洋(東大)

- ▶ **オーディオエフェクトチェーン推定**
 - オーディオエフェクトの種類と適用順を同時に推定するタスク
 - 非線形・時変なエフェクトが含まれる場合も多く、音響信号からの推定技術は発展途上
- ▶ **着眼: エフェクトチェーンの階層構造**
 - より少ないエフェクト数のチェーンが**入れ子状に包含される**
- ▶ **方策: マトリョーシカ埋め込みを用いたマルチタスク学習を導入**
 - **入れ子状に含まれるチェーンの予測を補助タスクとして導入**
 - マトリョーシカ埋め込み: 入れ子状に取り出された低次元の埋め込み
 - **異なる次元数の埋め込みでエフェクトチェーンの階層の深さを表現可能**
- ▶ **結果: マトリョーシカ埋め込みの導入により推定精度が向上**

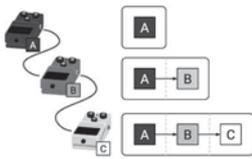


Fig. 1: Audio Effect Chain and Its Hierarchical Structure

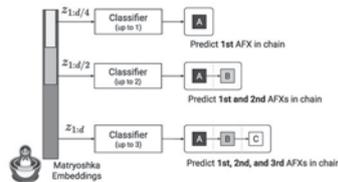


Fig. 2: Effect Chain Estimation using Matryoshka Embeddings

3-1-7

3-1-7 離散扁長回転楕円体列を窓関数とするストックウェル変換を用いた超分解能時間周波数解析手法の検討

Study of a Super-Resolution Time-Frequency Analysis Method Using the S-Transform with Discrete Prolate Spheroidal Sequences as Window Functions

☆菅沼大輝(広島市立大), 藤ノ木健介(神奈川大・工), 石光俊介(広島市立大)

- ◆ 超分解能時間周波数解析手法である **Superlet transform** において低周波領域でガウス窓の打ち切りによる「にじみ」が発生しやすい
- ◆ **離散扁長回転楕円体列 (DPSS)** を窓関数とするストックウェル変換を超分解能化することで、この影響を抑制する (Fig.1)

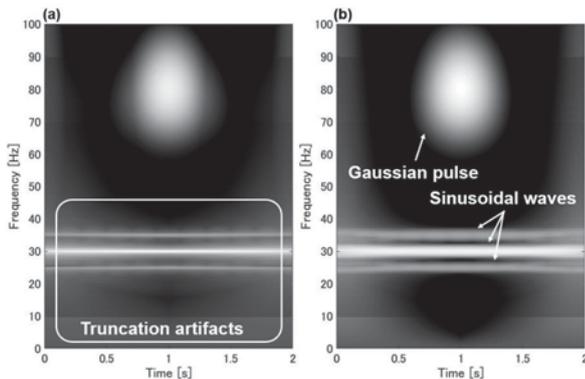


Fig.1: Comparison of time-frequency distributions of a test signal using the super-resolution Stockwell transform. (a) Gaussian window. (b) DPSS window.

3-1-6

3-1-6 SPレコード写真の輝度値情報から抽出した音溝底部の形状からの音声再生

Audio Reproduction of 78 rpm Records from the Shape of the Groove Bottom Extracted from the Luminance Information of the Pictures

☆三輪尋希, 及川靖広(早大理工), 来住伸子(津田塾大)

- ◆ 歴史的に貴重な資料であるSPレコードの音声を、レコードの音溝をデジタルカメラで撮影した画像から非接触により再生する。
- ◆ 撮影した画像に対し、座標変換による溝の直線化と高解像度化を行う。
- ◆ 音溝の底は針が接触しにくく、形状が保存されている可能性が高い。
- ◆ 溝底が溝の斜面よりも暗く映ることに注目し、溝を横切るように「暗度」を分析し、そのピークを検出することで溝底の座標データを得る。
- ◆ 画像の傾き補正を経て、ヒストグラムを使って溝一本分のデータを抽出し、それらを順番に並べることで音声データを復元する。
- ◆ 針で再生した音声と本研究で作成した音声のスペクトログラムを Fig.1 に示す。おおよそ 1000Hz 以下の音声については概ね復元することができたが、作成した音声には多くのノイズが含まれている。

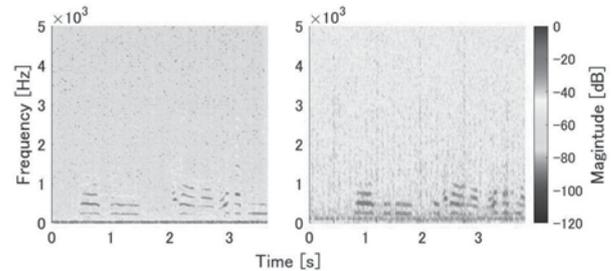


Fig.1: Spectrograms of Sound from Stylus (left) and of the Produced Sound (right)

3-1-8

3-1-8 Note Position(s) on Guitar Neck Detection

ギターネック上の音符位置検出

☆ Wai Lun MAK, Yasuhiro OIKAWA (School of Fundamental Science and Engineering, Intermedia Department)

- Prior research often claims that detecting finger positions on the guitar neck requires both camera and audio input.
- Our work introduces the world's first lightweight system that tracks fretboard fingering in real-time using audio alone.
- The system handles everything from single notes to complex chords.
- Users spend about 30 minutes recording single-note samples for training; the machine-learning stage takes roughly 20 minutes.
- The model is trained using CQT, MFCC, and four additional features.
- Our results also confirm that single-note FFT signals can be combined to accurately synthesize chord signals.
- Figures 1-4 below show the final output of the system when being used live.



Fig. 1: detecting a standard G chord



Fig. 2: detecting a barred A chord



Fig. 3: detecting a standard Dm chord



Fig. 4: detecting an A-shaped C chord

3-1-9

3-1-9 時間伸長圧縮処理を用いた音声情報ハイディング法におけるブラインド検出の検討

Study on blind detection in speech watermarking method using time-stretching and compression processing

☆川崎麦(TMCIT), 鶴木祐史(JAIST), 磯山拓都(TMCIT)

- ◆問題点: 従来の時間伸長圧縮処理を用いた音声情報ハイディング法は透かし検出にホスト信号も必要であるため、実用性が低い。
- ◆目的: 時間伸長処理された音声信号と時間圧縮処理された音声信号を区別し、透かし情報のブラインド検出を提案することを目的とする。
- ◆方法: 高域に着目した音響特徴量を抽出し、ランダムフォレストで時間伸長/圧縮を多数決判定することで透かし情報を検出する。
- ◆結果: 図1(a)より、知覚不可能性は従来法(TSC)と提案法(Normal)は同程度に高いことがわかった。図1(b)より頑健性は低かったものの、Normal及びOggにおいては、4bps、8bpsにおいてERBが10%を下回ることがわかった。

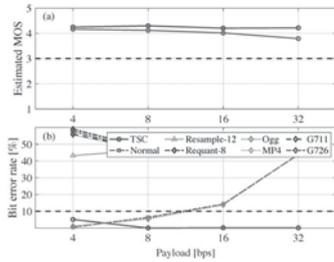


Fig.1 Results of sound quality evaluation and BER: (a) MOS values of sound quality estimated by NISQA and (b) BER.

3-2-1

3-2-1 境界が移動する場合の反射波・透過波について Reflected and transmitted waves in the case of moving boundary

○土屋 隆生 (同志社大・理工) Takao Tsuchiya (Doshisha Univ.)

1次元音場について移動境界からの反射および透過を純音響的に取り扱った。移動Mach数が比較的小さく流体力学的な取り扱いをしない仮定の下 ($M < 0.1$) で、移動境界からの反射波および透過波を理論的に導出した。その結果、移動境界からの反射・透過係数は境界の移動速度に依存することが分かった。つぎに、FDTD法に移動境界を実装した。FDTD法において境界を密度分布として表現すれば、FDTD法の簡便さを損わずに移動境界を表現できることを示した。また、1次元の数値実験により、移動境界のマッハ数が0.1までであれば、十分な精度で移動境界からの反射波・透過波を計算できることが示された。

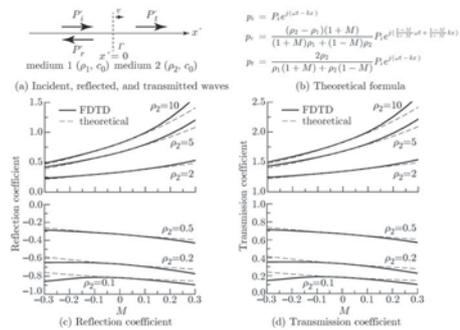


Fig.: Reflection and Transmission coefficient for moving boundary.

3-2-2

3-2-2 日本海固有水の音速構造の差が音波伝搬に及ぼす影響

Effect of difference in sound speed structure of the Japan Sea proper water on sound propagation

○鶴ヶ谷芳昭 (三陽精工), 菊池年晃 (防衛大)

- ◆日本海の音速構造の四季変化に対する音波伝搬を検討するため、深度304.8mより深い深度の音速構造を冬季の音速構造で代表した。日本海固有水において、水温は一定であり、冬季の密度の増した海水が沈降することから冬季の値で置換えた。しかし、一方で、音速計算式から海水温が一定となると、音速は深度(圧力)で線形に増大する。冬季読取値と圧力置換値との間には深度609.6mで-0.61m/s、1828.8mで-2.1m/s、3048mで+2m/sである。この差は音波伝搬に影響する。

- ◆Fig. 1にSDが220mにおける冬季の音速構造の読取値と圧力置換値の音波伝搬の比較を示す。周波数は1200Hzである。音速構造の違いから、海面付近の上向き伝搬及び1250m以深の伝搬パターンに大きな差異を生じていないが、300mから1500mでは差異を生じている。また、海底付近の転回点の距離を見ると読取値の方が音源方向へシフトしている。

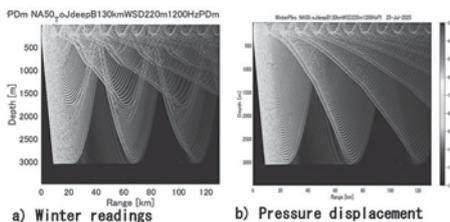


Fig. 1 Comparison of sound propagation patterns from winter readings and pressure displacement

トしている。春季において、読取値は圧力置換値+10m/sの値に近接している。

講演取消

3-2-3

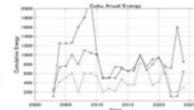
3-2-4

3-2-4 震源構造の累積エネルギーによる定量化と可視化

Quantification and visualization of deep hypocenter based on cumulative energy s

○菊池年晃(防衛大)

中部海域には地震を頻発する震源域が幾つか存在する。その震源震動を地震のエネルギーに基づいたパラメータ、累積エネルギーを用いて定量化解析を行った。この定量化による震源震動の年次変化は震源域の全体像を可視化している。この可視化により、異なる震源域や時期の震動構造を比較し得ることができる。即ち、この海域の震源域は二分化構造で、深部震源域はエネルギーの発震機能を有するが浅部震源域はエネルギーの滞留、蓄積の機能を有する。



中部海域の震源、東海沖(青)、紀伊半島沖(赤)及び遠州灘(緑)の震源域の年間累積エネルギーの年次変化を示す。これらの3曲線は類似した年次変化をしており同一プレート上に位置していることが分かる。また2004年から2025年にわたる変化は周期性を包含している。その発端は、2004年9月に、紀伊半島南東沖震源域の浅部震源域でM7.4を始めとする大きな地震が頻発した。深部震源域はエネルギーの発震機能を有するが浅部震源域はエネルギーの滞留、蓄積の機能を有する。従って、2004年の大きな地震の頻発により、それ以前の蓄積エネルギーは消失したので、その後の、震源域で発生した累積エネルギーは重要な特性を持つ。その出発は2005年に発生したM5.4の単発パルスで、その後約1年間隔で4年継続した。これらのパルスの発震深度はプレートの最先端であり、プレートの新たな動きと考えられる。従って、図に示した震動変動は周期的な性質をもつので、今後の予測に役立つことができる。

3-2-6

3-2-6 表面プラズモン共鳴型超音波センサにおけるセンシング領域の検討

A SPR ultrasonic sensor with variable sensing area

☆松寺歩美, 市橋隼人, 松川真美(同志社大)

- ◆近年、表面プラズモン共鳴 (SPR) は様々な分野への応用が期待されており、超音波検出もその一つである(1-2)。
- ◆SPR 超音波センサではセンシング形状が楕円形かつ大きな領域となる。
- ◆レーザ光径の制御により、超音波の計測領域を可変にできるシステムを検討した。
- ◆光学系にピンホールを挿入して、計測領域を変えると振幅が変化した。この変化を理解するには、領域内の音場分布やピンホール設置位置の正確さなどの詳細な検討が必要である。

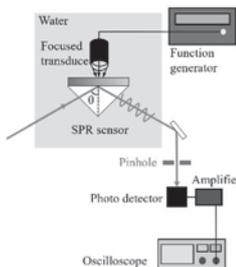


Fig. 1 Experimental system.

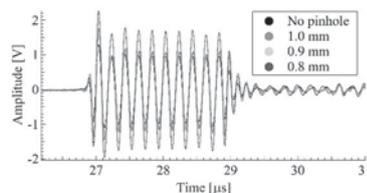


Fig. 2 Observed ultrasonic waveforms.

[1] K.Dezao et al, Applied Acoustics, 240,110967, (2025)
 [2] S.Nakatsuji et al, IEEE Trans. Ultrason. Ferroelectr. Freq. Control, 70(6), 562-568, (2023).

3-2-5

3-2-5 日本南岸沖海底地震計による低周波鯨類鳴音の追跡

Tracking of low-frequency whale calls using ocean-bottom seismometers off the southern coast of Japan

○岩瀬良一(JAMSTEC)

- ◆日本の南岸の海域に設置された海底ケーブル型地震・津波観測網 N-net 沖合システム及び DONET を構成する地震計の波形データを使用したこれまでの解析[1,2]により、鯨類鳴音の検出事例がなかった当該海域において、低周波鯨類鳴音と推定される信号を検出し、音源が2025年3月8日から13日にかけて東に移動する様子を捉えた。
- ◆今回これに先立つ3月5日から7日までの3日間の波形データを解析し、3月6日の17時から18時 JST にかけて N-net の南西端の観測点において鯨類鳴音の特徴を有する信号を新たに検出、次いで検出観測点の北東方向への拡大を確認した。
- ◆その結果、当該音源が3月6日から同13日にかけて九州の南東沖から紀伊半島南東沖まで移動したことが明らかとなった。
- ◆これに加え、同年9月17日から18日にかけて同様の信号を検出した。音源は3月と同様、九州南東沖から四国東部沖まで移動したが、紀伊半島沖までには達しなかったと推定される。
- ◆3月及び9月の鳴音検出という結果は日本海溝沿いの S-net に見られる鯨類鳴音の季節移動[3]を示唆するが、いずれも東向きという移動状況は局地的な影響も考えられ、解析結果を踏まえ今後詳細に検討する。

参考文献

[1] 岩瀬, 音講論(秋), 89-90, 2025. [2] R. Iwase, Proc. USE2025, 1P6-5, 2025. [3] 望月ほか, JpGU2025, SCG55-P26, 2025.

3-2-7

3-2-7 A trial construction of miniaturized reverberation chamber for evaluating acoustic materials at ultrasonic frequency

☆Zuhal Maulana FIRDAUS, 和田 有司, 中村 健太郎 (Science Tokyo)

- ◆ Airborne ultrasound uses for short-range sensing in vehicles, robotics, non-contact manipulation, and human-machine interaction in confined spaces are increasing, and unwanted reflection and interference become harder to avoid, necessitating acoustical treatment of the environment, which requires a reliable material absorption characteristics measurement system.
- ◆ Absorption characteristic of 15 materials were measured in the range of 20 - 50 kHz using a miniaturized reverberation chamber with an internal volume of 0.0292 m³ designed with no parallel walls and a conical diffuser for sufficient sound field diffusivity.
- ◆ The chamber is excited for a short period, and the following reverberation is recorded. The decay time was computed using Schroeder integration, and the time difference between empty chamber and the chamber with the sample material was used to calculate the absorption coefficient of the material for each frequency. Measurement was repeated 4 times under three different microphone positions for statistical and spatial robustness.
- ◆ The resulting absorption coefficients for each material were averaged across 5 frequency bands as shown in Fig. 1.

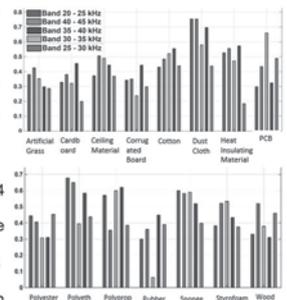


Fig. 1: Material absorption performance per frequency band. Y-axis is the average absorption coefficient.

3-2-8

毛髪がVHF音の伝搬に与える影響(その2) -パルスエコー法を用いた 毛髪の反射特性の検討*-

Effects of hair on VHF sound propagation-Part 2,

Measurement of hair reflection characteristics using the pulse-echo method

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

- ◆本研究では, VHF音(16 kHz~32 kHz)の伝搬シミュレーションに用いる毛髪モデル構築の為の基礎的知見を得るため, VHF領域における人工毛髪の反射特性を, 管内伝搬のパルスエコー法で測定した。
- ◆人工毛髪は, 長さの異なるショート, ボブ, ロングの3条件で, 測定位置は頭頂部から130 mmと200 mmとした。
- ◆人工毛髪はVHF領域の反射特性に影響を与えることが確認されたが, 測定する毛髪の位置や条件により反射率に大きなバラツキがあった。
- ◆毛量に伴う反射波の変化から, VHF音は毛束内を伝搬する間に減衰した可能性が示唆された。

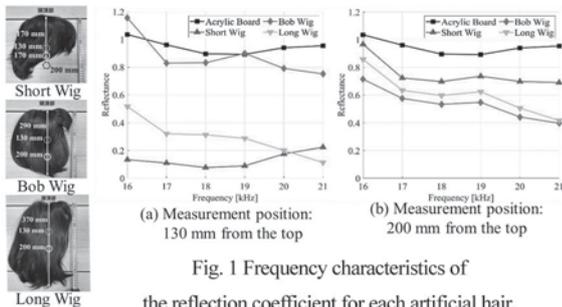


Fig. 1 Frequency characteristics of the reflection coefficient for each artificial hair.

3-3-1

3-3-1 ChatGPTとPythonで学ぶ 室内音響シミュレーション

Learning Acoustic Simulation Using ChatGPT and Python

○青木 直史 (北大)

- 近年, ChatGPTをはじめとする生成AIツールが急速に普及し, 教育現場での活用が広がっている。とくに, プログラミング教育では, AIとの対話を通じて具体的なプログラムの生成や修正を補助できるようになり, 学習者が自律的に学ぶための新たな手段として注目されている。
- 本研究では, ChatGPTを活用しつつPythonを用いて室内音響シミュレーションを実践した筆者の取り組みを, ひとつのケーススタディとして報告する。



3-2-9

3-2-9 VHF音の聴こえに関する基礎的検討 -耳栓着用時の聴覚閾値計測-

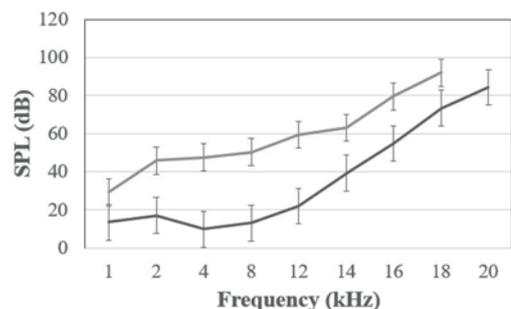
Basic study on VHF sound hearing

-Hearing threshold measurements with earplugs-

☆久保谷旭陽, 松下宗輝, 近藤奏海(神奈川工科大),

中村健太郎(東京科学大), 廣江正明(小林理研), 上田麻理(神奈川工科大)

- ◆VHF音(16 kHz~32 kHz)の聴こえのメカニズムは現在も議論が続いている。骨伝導でVHF音を感知しているという説や, 聴覚器以外でVHF音を知覚しているという説もある。
- ◆本研究では, VHF音が気導を經由して知覚されているか否かを明らかにすることを目的に, 耳栓の装着・非装着時におけるVHF領域の聴覚閾値を計測した。
- ◆測定の結果, 耳栓装着時は非装着時と比べ, VHF領域であっても, 聴覚閾値の上昇を確認した(Fig1)。



◆ — Earplug in right side only — (a+c)Silicone+Polymer clay

Fig. 1 Comparison of hearing thresholds with/without earplugs.

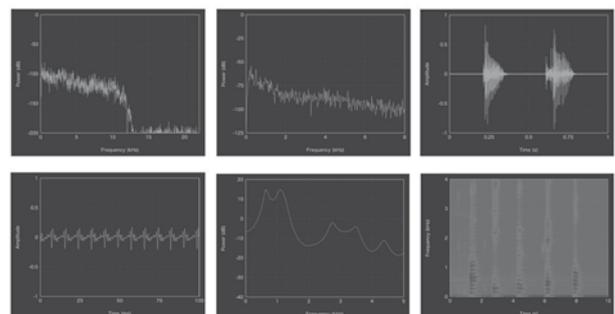
3-3-2

3-3-2 使いやすい音響分析アプリの開発

Development of easy-to-use acoustic analysis apps

○岡崎聡 (富山県立大)

- ◆音響分析アプリ群を開発した。
 - スペクトルアナライザ (-22 kHz, -8 kHz)、波形ビューワ (1 s, 100 ms)、フォルマント・ヴィジュアルイザ、スペクトログラム。
 - Webアプリケーションとして, スマートフォン, タブレット, PCで動作する。
- ◆音響教育への有用性
 - 誰でも入手可能で, 簡単に実行できる。
 - Simple, fun, and deep.
 - 教育に用いる例を列挙した。
 - 小・中・高の音響教育や, 音響を志す大学生, 言語聴覚士などの教育に役立てて欲しい。



3-3-3

3-3-3 人や生成AIとの対話能力を高める大学生向け学習教材の開発と評価

Design and Evaluation of Learning Materials for University Students to Enhance Dialogue Skills with Humans and Generative AI.

○上田麻理, 金森克洋, 田中哲雄(神奈川工科大)

本稿では、大学生の考える力、伝える力の育成のために数学、音響波形、イラスト、音の4領域にわたる言語化教材を提案した。本教材は、Web上に正解がないため学生が自ら思考せざるを得ず、不正防止と能力向上の両面で有効であることが確認された。



3-3-4

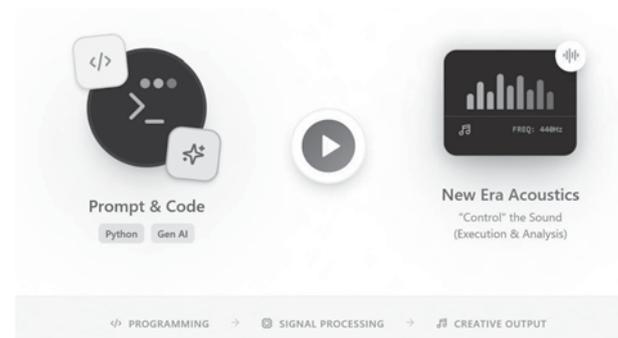
3-3-4 PythonとAIで音を操れ！プロンプトエンジニアリングで学ぶ新時代音響学授業の試行-その1

Learning Acoustics via Prompt Engineering Part 1:
A Pilot Study of a Python and AI-Integrated Course

○上田麻理, 金森克洋(神奈川工科大),

桐生昭吾(都市大), 田中哲雄(神奈川工科大)

神奈川工科大学情報学部2年生の2025年度後期の演習科目において、コードを書かない音響学Pythonプログラミング演習を試みた。さらに学生が作成したプロンプトの質と学生の音響学の基礎に対する理解度の関連を調べたので、その1とその2の2本立てで報告する。



3-3-5

3-3-5 「PythonとAIで音を操れ！プロンプトエンジニアリングで学ぶ新時代音響学授業の試行-その2

Learning Acoustics via Prompt Engineering Part 2: A Pilot Study of a Python and AI-Integrated Course

○金森克洋(神奈川工科大), 桐生昭吾(都市大), 田中哲雄(神奈川工科大), 上田麻理(神奈川工科大)

- ◆筆者らは神奈川工科大学情報学部情報メディア学科2年生を対象とする新規な音響学の演習授業を34人の学生に対し実施した。
- ◆「純音(正弦波)のグラフ再現」の演習回につき学生からのプロンプトをヒト(教員・TA)と生成AIの両方で既定の6種の評価項目で評価した。また演習最終回には音響学の理解度アンケートも実施した。
- ◆プロンプト評価ではヒトによる評価とAI評価に有意な正の相関が得られ、AIによるプロンプト評価の可能性が示された。
- ◆プロンプト品質と理解度アンケート結果には有意な相関がなく、プロンプト作成能力と従来型の授業成績とは無関係と考えられる。
- ◆プロンプト評価項目2(Item2)「音響・情報系パラメータ定義」では理解度アンケート結果と弱い正の相関が見られ、正弦波の再現という課題では音響学の知識が有効だったと考えられる。

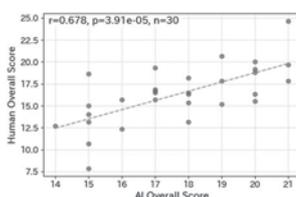


Fig.1 Relationship between AI Overall Evaluation and Human Overall Evaluation Score of prompts.

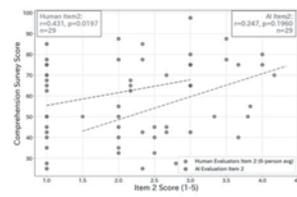


Fig.2 Relationship between Item2 Score (by human and AI) and Comprehension Survey Score.

3-3-6

3-3-6 音響に関する研究を応用した学生実験の紹介

Introduction to student Experiments applying acoustic research

☆丸目勝斗, △中村仁飛, △豊田昌弘, 小塚晃透(愛工大)

- ◆愛知工業大学の学部生の授業で行われている学生実験について紹介する
- ◆音場測定と超音波マニピュレーションを紹介する
- ◆音場測定はArduinoを用いて制御する電動ステージを用いて出力した信号を、Arduinoのアナログ入力機能を用いて電圧を取得することで音場を測定した
- ◆過去には手動で測定を行っていたが、自動化することによって高速で測定することができた。また、レイリーの式を用いたシミュレーションを行ったところ、正確に測定できていることが確認できた
- ◆超音波マニピュレーションは超音波の進行波と反射波や、逆向きの進行波によってできる定在波などで物体を非接触で浮揚させる技術である。これをAD9833というDDSのモジュールを用いて信号を発生することで超音波を出力する
- ◆正弦波を出力でき、発泡スチロール球を非接触で浮揚させることができた。また、4mmで浮揚することができた
- ◆これらの機器は安価にそろえることができ、精度も高い
- ◆学生実験や、始めて実験を行う場合などの場合にお勧めしたい。

3-3-7

3-3-7 言語聴覚士養成校の音響学・聴覚心理学 授業時間別の工夫

Teaching Strategies for Acoustics and Auditory Psychology by Class
Duration in Speech-Language Pathology Training Programs

○竹内京子(順天堂大)

- ◆ことばのリハビリを行う言語聴覚士養成校では、音響学が必修科目、学生が最も苦手な科目、その改善が求められている。
- ◆音響学、聴覚心理学の授業は各学校で時間数が違い、扱うことができる内容が違う。本発表は、発表者が今までに実施した、音響学・聴覚心理学の授業のカリキュラムのパターンを概観し、それらの長所、短所を述べる。
- ◆様々なカリキュラムを体験し、個人的に一番やりやすかったのが音響学(聴覚心理学を含む)30コマ以上の時間を単独担当または、連携を密にして2人で担当した場合であると考える。
- ◆重なるの多い2科目を連携なしで、複数で担当することは、短期間に国試という成果を求められる養成校では、時間的効率を考えると、避けた方がいいだろう。
- ◆今後の課題として、多くの養成校で取り入れられている「音響学(聴覚心理学を含む)」30コマは、よい選択肢であるが、それ以上に時間を増やすことによって、学生と教員両者の心の余裕とともに、成果としての学生の理解度も上がるのではないだろうか。
- ◆学生が、苦手な音響学は、実は、難しく理解することができないのではなく、学ぶ内容に対して、授業時間が短すぎたのではないだろうか。時間を増やすことによって、より効率的に学ぶことができると考える。

3-5-2

3-5-2 FasterVoiceGrad: 敵対的拡散変換蒸留を用いた ワンステップ拡散モデル声質変換の高速化

FasterVoiceGrad: Adversarial Diffusion Conversion Distillation
for Faster One-step Diffusion-Based Voice Conversion

○金子卓弘, 亀岡弘和, 田中宏, 近藤祐斗(NTT)

- ◆VoiceGradなどの拡散モデルベースの声質変換モデルは、高品質かつ高話者類似度の音声を生成でき、近年注目を集めている。
- ◆しかし、推論には逆拡散処理を繰り返す必要があり、変換速度が遅いことが課題であった。
- ◆この課題を克服するため、私たちの先行研究であるFastVoiceGradでは、敵対的拡散蒸留を用いて、マルチステップ声質変換モデルからワンステップで変換可能なモデルへ蒸留する手法を提案した。
- ◆しかし、計算コストの削減効果は逆拡散過程の主処理に限られており、発話内容の保持に不可欠なコンテンツエンコーダの計算コストは依然として高く、課題が残されていた。
- ◆そこで本研究では、逆拡散過程の主処理とコンテンツエンコーダを同時に蒸留するFasterVoiceGradを提案する。
- ◆特に、再構成処理ではなく変換処理において蒸留を行う敵対的拡散変換蒸留を用いることで、学習可能なコンテンツエンコーダを通した再構成による逆拡散過程の主処理の学習阻害を防ぐ。
- ◆評価実験では、FasterVoiceGradはFastVoiceGradと同等の変換性能を維持しつつ、GPU上では約6.6倍、CPU上では約1.8倍の高速化を達成することを示す。

Audio samples:

<https://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/fastervoicegrad/>

3-5-1

3-5-1 PRVAE-VC2: 音声表現の知識蒸留による ノンパラレル音声変換

PRVAE-VC2: Non-Parallel Voice Conversion
by Distillation of Speech Representations

○田中宏, 亀岡弘和, 金子卓弘, 近藤祐斗(NTT)

本研究では、自己教師あり音声表現学習を用いた非並列多対多声質変換(VC)のための知識蒸留手法を提案する。従来手法である摂動耐性変分オートエンコーダ(PRVAE-VC)は、非ストリーミングおよび低遅延ストリーミングVCにおいて高い性能を示してきたが、変換音声と目標話者音声との間には依然として差異が存在していた。

本稿では、事前学習済みHuBERTから得られる文脈的に豊かな音声表現を導入した改良手法PRVAE-VC2を提案する。さらに、HuBERTが計算量の大きいモデルである点を考慮し、知識蒸留によりストリーミング可能な学生モデルを構築する。実験の結果、提案手法は話者類似度、音質、および音声認識誤り率の全てにおいて従来法を上回る性能を示した。



3-5-3

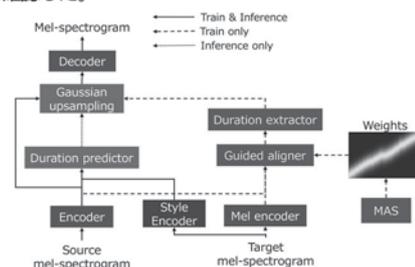
3-5-3 系列変換型複数話者声質変換方式の比較

Comparison of sequence-to-sequence
multi-speaker voice conversion models.

◎山下陽生(1, 2), 岡本拓磨(2), 高島遼一(3), 大谷大和(2),
滝口哲也(1), 戸田智基(4, 2), 河井恒(2)

(1 神戸大学, 2 情報通信研究機構, 3 立命館大学, 4 名古屋大学)

- ◆本稿では、Attention alignment機構を備えた系列長変換型複数話者声質変換モデルMS WSAS-VC (Multi Speaker WSAS-VC)を提案する。
- ◆提案モデルは、Weighted Guided Attentionに基づくalignment機構を維持したまま、Global Style Token (GST)を導入することで、話者性と発話スタイルを分離して制御可能とする。
- ◆これにより、話者ごとに異なる声質および発話長を考慮した声質変換を実現する。
- ◆実験では、合成音声品質、話者類似度、および音声の崩れに関する客観評価を行い、既知話者変換において高精度な声質変換が可能であることを確認した。



Network architecture of MS WSAS-VC.

3-5-4

3-5-4 Any-to-Many 声質変換における入力話者・性別に頑健な発話内容獲得のためのデータ拡張の分析

An Analysis of Data Augmentation for Speaker- and Gender-Robust Content Representation in Any-to-Many Voice Conversion.

☆山下凜, 能勢隆, 小林清流, 伊藤彰則 (東北大)

- ◆本研究では話者変化に頑健であるとされる音声 SSL モデルである ContentVec の学習手法に関する分析を行い、声質変換における現行のデータ拡張手法の問題点を明らかにした。
- ◆具体的には ContentVec について、以下の分析を行った。
 - ContentVec の学習時のピッチ・フォルマントシフトによるデータ拡張が実際の話者分布をどの程度再現できているか
 - ContentVec を用いた声質変換が、ピッチ・フォルマント周波数の変化に対してどの程度頑健に動作するか
- ◆結果としては、現状のデータ拡張手法は実際の話者変化に即しておらず、声質変換の観点からは改善の余地があることが分かった。
- ◆声質変換において、入力音声のパラメータ変化に対しては頑健に動作したため、データ拡張手法の改善により話者変化への頑健性を向上させられる可能性が示された。

Table 1 Speaker similarity with target speaker before and after conversion

	別話者の自然音声	データ拡張音声
変換前	0.234	0.135
変換後	0.845	0.886

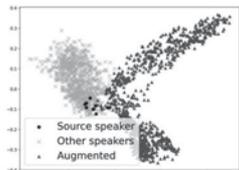


Fig. 1 Comparison of speaker distribution

3-5-6

3-5-6 ガウス混合モデルによる局所線形変換を用いた自己教師あり音声表現に基づく声質変換

Voice Conversion Based on Self-Supervised Speech Representations Using Local Linear Transformation with Gaussian Mixture Models.

☆田名部智也, 西島大史, 齋藤大輔, 峯松信明 (東大)

GMM is Back!! – When Classic Meets SSL

【背景】深層学習全盛の声質変換研究において、自己教師あり学習(SSL)音声表現と古典的な変換モデルを組み合わせた場合の性能や変換の性質に関する議論が不足。

【着想】SSL 特徴量に単一の大域的線形変換を適用する既存の LinearVC を、GMM の各混合成分に基づく複数の変換の重ね合わせへと拡張。

【提案】Fig. 1 のように WavLM から抽出されたソース・ターゲット音声の SSL 特徴量を対応づけてガウス混合モデル (GMM) を学習。

【結果】少量データ条件を含め、深層学習ベース手法に匹敵する話者類似度を示し、特定の条件下で LinearVC を上回る性能を確認。GMM の解釈性が変換の性質や SSL 特徴量空間の構造に対する理解を今後さらに深めるための有効な枠組みを与える。

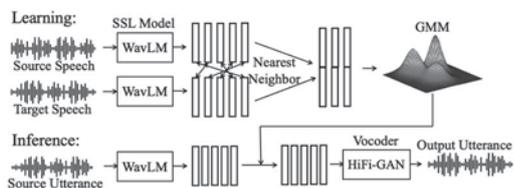


Fig. 1: Overview of the proposed SSL-GMMVC framework

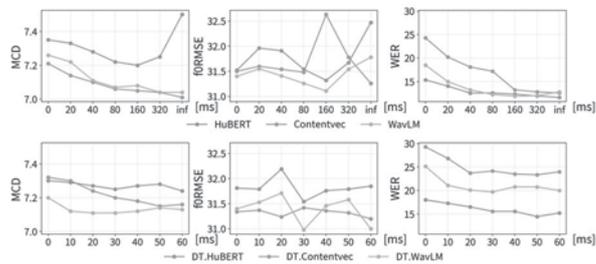
3-5-5

3-5-5 知識蒸留による因果的な音声潜在特徴抽出と音声変換への適用

Causal Speech Latent Feature Extraction via Knowledge Distillation and Its Application to Voice Conversion.

☆古田京平(名古屋大学), HUANG Wen-Chin(名古屋大学), 安田裕介(名古屋大学, 国立情報学研究所), 戸田智基(名古屋大学)

- ◆SSL 特徴を用いた音声変換は高性能だが、多くは未来情報に依存するため、低遅延リアルタイム実行では性能低下が生じる。
- ◆本研究では、HuBERT/ContentVec/WavLM の SSL 特徴を教師として、CNN ベースの因果的特徴抽出器を知識蒸留し、先読み量 (0~120 ms) を変えながら音声変換品質を比較した。
- ◆先読み量に対する性能の改善・頭打ちの挙動は教師モデル依存であり、教師 SSL が暗黙に利用する未来情報の実効長が異なることを示した。
- ◆蒸留モデルは MCD 等では教師に近い場合がある一方で WER が悪化しやすく、特に HuBERT の蒸留モデルは他より劣化が小さいなど、蒸留ギャップが教師モデルで異なることを示唆した。



3-5-7

3-5-7 Multilingual Retrieval-Augmented Generation Enhanced LLM-based Speech Recognition

☆Pengcheng Wang, Sheng Li, Takahiro Shinozaki (Institute of Science Tokyo)

- ◆In recent years, automatic speech recognition (ASR) systems based on Large Language Models (LLMs) have achieved remarkable recognition performance. However, they still have shortcomings in handling rare words and domain-specific expressions. To address this issue, retrieval-enhanced generation (RAG) has been introduced. However, most existing methods apply retrieval after decoding and depend on potentially incorrect transcriptions, resulting in error propagations.
- ◆In this work, we propose an adaptive pre-RAG framework that retrieves information directly from the speech representation prior to LLM decoding. By combining speech-based retrieval with a lightweight adaptive router, the proposed method selectively activates retrieval only when it is needed. This design improves robustness and prevent unnecessary computation.

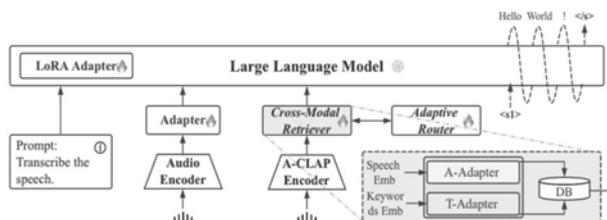


Fig. 1. Overview of our proposed RAG based ASR framework.

3-5-8

3-5-8 Whisper における Mixture-of-experts 型 Feed-forward ネットワークを用いた 多言語音声認識

Multilingual speech recognition with Mixture-of-Experts via FFN upcycling in Whisper model

©品山健太郎, 三村正人(NTT 研究所)

- ◆大規模事前学習済みモデル Whisper を用いた多言語音声認識において、言語数およびデータ量増加に伴うモデルキャパシティ不足と言語間の干渉による性能低下を防ぐ手法の構築を目指す。
- ◆Transformer の Feed-Forward Network を複製して Mixture-of-Experts (MoE) 化する手法を提案する。
- ◆提案手法では、学習時・推論時ともに 1 つの Expert のみを選択する Switch 型 MoE を採用し、アクティブパラメータ数をシンプルな fine-tuning 手法と同一に保った。
- ◆10 言語(西洋 5 言語+非西洋 5 言語)を用いた実験により、FFN の MoE 拡張によってシンプルな fine-tuning 手法より認識性能が向上し、Expert 数を増やすことでさらなる改善が得られることを確認した。

Table 1: Common Voice (西洋 5 言語) における性能比較

	WER(%)					ave.
	en	de	fr	it	es	
Mono FT	16.4	10	15.4	11.8	9.6	12.64
Multi FT	17.2	11	16.5	12.9	10.3	13.58
MoE (N=4)	17.2	11	16.4	12.8	10.3	13.54
MoE (N=8)	17.2	10.9	16.4	12.7	10.2	13.48

Table 2: Common Voice (その他 5 言語) における性能比較

	CER(%)					ave.
	bn	ru	sw	tr	th	
Mono FT	34	5.4	10.3	8.2	11	13.78
Multi FT	34.9	6.4	12	10.2	12.3	15.16
MoE (N=4)	34.6	6.3	12.1	10.2	11.9	15.02
MoE (N=8)	34.6	6.2	12	10.1	12	14.98

3-5-10

3-5-10 Textual In-Context Learning for Robust Speech Recognition

○ Benita Angela Titalim, Mehmood Faisal, Sakriani Sakti (NAIST)

- ◆ASR performance degrades severely under real-world noise, especially competing speech.
- ◆This work explores language-level prompting with LLM-based ASR to improve robustness without retraining acoustic models.
- ◆Experiments compare input-focused (p1), output-focused (p2), and noise-aware prompts (p3-p5) under competing speech.
- ◆Output-focused prompting reduces WER by ~34% compared to input-focused prompting, while noise-aware prompts yield up to 17.4% WER and 18.7% CER reductions at -5 dB SNR.
- ◆Compared to acoustic-level, language-level approach offers a lightweight alternative that improves speech recognition robustness, though severe interference remains challenging.

	Input/Output Aware Prompting		Noise-Aware Prompting		Noise+Output Aware Prompting
	p1	p2	p5	p6	p7
SNR 20dB	21.103	4.775	7.909	4.204	6.682
SNR 10dB	76.637	26.631	24.456	19.074	20.180
SNR 0dB	182.116	159.191	126.534	126.896	120.690
SNR -5dB	202.163	188.689	159.981	162.775	155.911

Table 1: Word error rate (WER, %) results of speech recognition under competing speech for different prompting strategies.

3-5-9

3-5-9 不確実性に基づく AED 型音声認識モデルの ストリーミング適応学習

Uncertainty-aware streaming adaptation for AED speech recognition models

©佐藤裕明, 佐久間旭, 菅野竜雅, △熊野正, △河合吉彦 (NHK 技研), △渡部晋治 (CMU)

- ◆Whisper に適用可能な AED 型モデルのストリーミング化手法を提案
- ◆Whisper をはじめとする AED 型モデルは、逐一の音声入力に対して音声情報が十分に得られる前に誤ったトークンを出力したり、不要なトークンを生成し続けたりする課題が存在
- ◆提案法ではトークン予測に対して「不確実性」を算出し、不確実性が高い状態ではトークン出力を保留し、低い状態になったらトークンを確定することで高精度なストリーミング動作を実現
- ◆日本語話し言葉コーパス (CSJ) と番組音声で提案法の有効性確認

Table 1: Offline and streaming performance of proposed method applied to Whisper large-v3. Each cell shows CER / 50th / 90th percentile latency (ms) (input window size: 320ms, uncertainty threshold: 0.05).

CSJ data			
Mode	eval1	eval2	eval3
Offline	3.7 / - / -	2.7 / - / -	2.9 / - / -
Streaming	3.8 / 1075 / 4417	2.8 / 1526 / 5433	3.2 / 922 / 4429

Broadcast data		
Mode	News	Corner
Offline	3.0 / - / -	4.9 / - / -
Streaming	3.5 / 2140 / 8100	5.9 / 1460 / 5540

3-5-11

3-5-11 複数話者音声認識コーパスはダイアライゼーションモデル学習に転用できるか？

Can multi-speaker ASR corpora be repurposed for training diarization models?

©堀口翔太, 依直弘, 芦原孝典, 安藤厚志, デルクロアマーク(NTT)

- ◆ダイアライゼーションモデルの学習に、複数のコーパスを結合して利用する研究が増えているが、その発話区間のラベリング基準はコーパスによって大きく異なる。具体的には、ダイアライゼーション向けのコーパスでは発話区間が厳密につけられている一方で、複数話者音声認識向けのコーパスでは発話中の無音区間も発話区間に含めていたり、発話単位音声認識の評価において始終端が切れないように区間にマージンを設けていたりする。
- ◆これを受けて本稿では以下の問いに答えるべく検証を行う。
 - 問1：発話区間の厳密性の違いがダイアライゼーションの評価にどのような影響を及ぼすのか？
 - 問2：元々の用途が異なるコーパスを結合して学習に用いた場合、モデルはどのように振舞うのか？
 - 問3：結局ダイアライゼーションモデルが推定すべきなのはどのような基準の発話区間なのか？

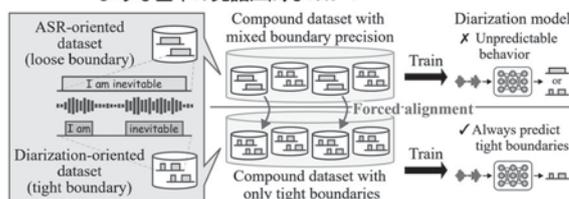


Fig.1: Overall framework of our experiments conducted in this paper.

3-6-1

3-6-1 北海道日本ハムファイターズのスポーツサイエンスを活用した選手育成の取り組み

Hokkaido Nippon-Ham Fighters' Players Development Initiatives Utilizing Sports Science.

○山口 雄大(北海道日本ハムファイターズ)

多くの若手選手が活躍する北海道日本ハムファイターズが取り組む、選手育成の 理念 と 活動事例 について紹介する。



3-6-2

3-6-2 高校野球における応援歌の影響に関する調査

A Study on the Influence of Cheering Songs in Japanese High School Baseball

☆岡村美和, 寺澤洋子(筑波大)

◆背景

試合中に音楽を用いて応援されている高校野球について、応援歌の影響を明らかにしたい。

◆手法

野球経験者に質問紙調査

プレー歴6~16年の74名から回答を得た。

◆結果

多くの選手は応援歌を好意的に受け止めている。

選手は、感情、プレー、試合の結果、球場やチームの雰囲気の良い影響を与えていると感じている。

一方で、緊張や声が聞こえず連携が取れないなどの悪い影響を感じたことのある選手もいる。

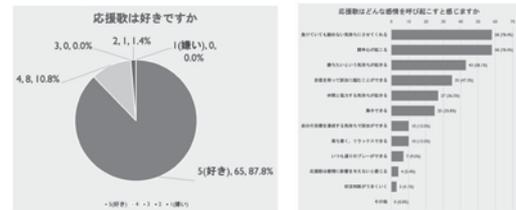


Fig. 1 応援歌に対する好意度

Fig. 2 応援歌が喚起する感情

3-6-3

3-6-3 スポーツ観戦者が叫ぶ「ヤバい」の正体 その2 —ムードメーターを用いて感情ボキャブラリーを可視化する試み—

Decoding "Yabai" II: Visualizing Emotional Vocabulary in Sports Spectatorship with the Mood Meter

○久永将太(神戸親和大学), 浅野恵子(順天堂大学)

- ◆スポーツ観戦者が叫ぶ「ヤバい」を対象に、14の観戦場面を設定し、大学生98名にムードメーターを用いて背後の感情構造を検討した。
- ◆「ヤバい」と同時に想起される感情ボキャブラリーには、場面ごとに明確な偏りがみられ、各状況に固有の情動パターンが確認された。
- ◆ポジティブ場面では高揚の種類が、ネガティブ場面では沈み方の違いが語彙として反映され、「ヤバい」は各場面の感情の質(快/不快・高/低エネルギー)に応じて語彙が体系的に変化することが明らかとなった。

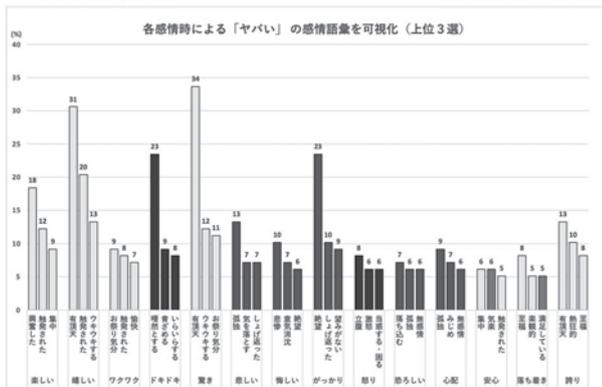


Fig. Emotional Vocabulary Associated with "Yabai" Across 14 Contexts (Top Three Terms per Emotion)

3-6-4

3-6-4 野球競技における聴覚活用のための基礎的検討 —バットの打球音の素材差の印象評価—

Fundamental Study on the Use of Auditory Information in Baseball: Assessment of Hitting Sounds Based on Differences in Bat Materials

○内藤光平, △山田俊(神奈川工科大),

三浦貴大(産総研), 上田麻理(神奈川工科大)

野球の打球音における評価実験において、スタジアムの実録音は実情に即している反面、背景騒音の混入や打撃条件の制御が困難であるという課題がある。一方、SG規格に準拠した収録方法は、制御された環境下で高いSN比と再現性を得られる利点があるが、スタジアムの実音をどの程度代表しているかについては十分な検証がなされていない。そこで、本研究では、SG規格音源とスタジアムでの実録音(フライ、ゴロ、ライナー)の同等性を明らかにすることを目的に周波数特性の相関分析や打球音の継続時間の算出と比較を行った。今回はその結果を報告する。

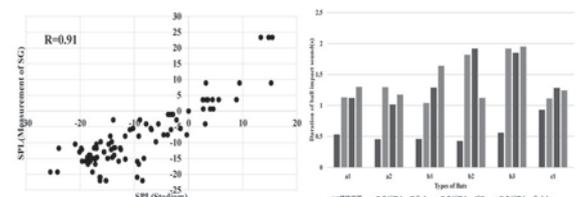


Fig.1 Correlation of frequency response between stadium and Measurement of SG environments (r=0.91).

Fig 2 Duration of impact sounds in indoor and outdoor environments.

3-6-5

3-6-5 フルコンタクトスポーツが聴力に与える影響に関する基礎的検討

A Basic Study on the Effects of Full-Contact Sports on Hearing Ability

○内藤光平, △山本奨悟(神奈川工科大), 三浦貴大(産総研), 東野哲也(国際医療福祉大), 上田麻理(神奈川工科大)

スポーツと難聴に関する研究は古くから行われており、剣道において頭部に衝撃を与える「面打ち」などを要因とする「剣道難聴」が報告されている。しかし、より接触強度の高いフルコンタクトスポーツを対象とした研究は少ない。そこでラグビーや柔道といったフルコンタクトスポーツ競技者を対象とした聴力測定を行ったためその結果を報告する。

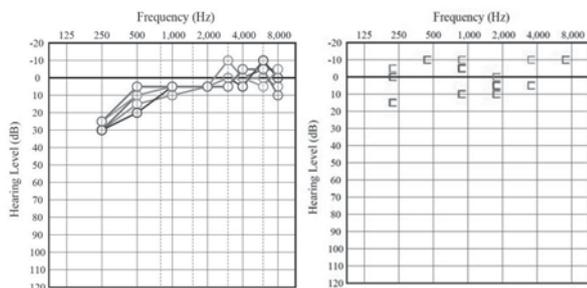


Fig.1 Air-conduction audiogram of rugby players(n = 5, right ear).
Fig.2 Bone-conduction audiogram of rugby players (n = 4, right ear).

3-6-7

3-6-7 FPS ゲームにおける聴覚情報活用に関する検討その6

-騒音下における聴覚トレーニングの実践-

A study on the use of auditory information in FPS games Part 6: Practice of Auditory Training in Noisy Environments

☆越水瑛怜(神奈川工大), 三浦貴大(産総研) △塩川茂樹, 上田麻理(神奈川工大)

- ◆本研究では、FPSゲーム「VALORANT」を対象とした聴覚トレーニングシステムの構築及び騒音下での足音判別を行う聴覚トレーニングを実施した。Fig. 1に聴覚トレーニングの回答画面を示す。
- ◆その結果、トレーニングにより騒音下の足音の判別精度が向上することが定量的に示された。



Fig.1: Answer screen for auditory training.

3-6-6

3-6-6 フラメンコ舞踊時の発生音に関する基礎的検討その2

熟練に伴う発生音の違い

Fundamental study of sounds generated during flamenco dance, part 2. Differences in sound production according to skill level.

☆木村美咲, 内藤光平(神奈川工科大), △干場かおり(スタジオ Fuente), 三浦貴大(産総研), 宮崎剛, 田中哲雄, 上田麻理(神奈川工科大)

- ◆これまでフラメンコ(Flamenco)舞踊に関する研究は、生体力学的視点に偏っており、音・聴覚に関する系統的な研究はほとんどない。
- ◆前報(その1)では、音の活用とサパテアード(足音)の音響分析について報告した。
- ◆本報(その2)では、サパテアードの発生音に着目し、音響データのみから熟練度を判別できるかという点について、比較・分析を行った結果を報告する。
- ◆熟練度を判断するアンケートにおいて全体的に中間的な評価であったが、初級と指導者の音に関しては有意に識別されていた。
- ◆自由記述において未経験者と経験者で音自体の評価と動きを読み取ろうとする傾向が見られた。

Table 1 熟練度判断の混同行列.

		評価結果			
		初級	中級	上級	指導者
提 示 音 源	初級	140 (34%)	145 (36%)	96 (24%)	27 (7%)
	中級	120 (29%)	151 (37%)	111 (27%)	26 (6%)
	上級	46 (17%)	111 (41%)	84 (31%)	31 (11%)
	指導者	26 (19%)	43 (32%)	45 (33%)	22 (16%)

太字下線は残差の検定で有意である箇所を意味する。

3-8-1

3-8-1 方言らしさを伝える手がかりを探る: 逆相 f_0 ・正弦波状 f_0 ・ささやき声

Exploring clues to convey the essence of dialects: Inverse phase f_0 , sinusoidal f_0 , and whispering voice.

○内田照久(大学入試センター 研究開発部)

- 目的 標準語音声の f_0 パターンの高低を反転した逆相 f_0 音声や正弦波状 f_0 の音声からは、「未知の方言印象」が想起される。今回、通常発話でない声種でも方言印象が想起されるかを検討
- 実験
1. 通常発話: 正相 f_0 , 逆相 f_0 , 正弦波状 f_0 の音声条件
 2. 雑音系の音声: 雑音駆動音声, 疑似ささやき声の条件
 3. 多層バンドノイズ音声: 3種の f_0 音声から生成した音声
- 結果 「標準語らしさ」、「地方の方言らしさ」などを10件法で評定
- 結果 首都圏の国公立大学1年生310名が評価実験に参加、逆相 f_0 や正弦波状 f_0 の音声は、言語音声として破綻する訳ではなく、地方の方言として認知された。しかし、雑音系音声、バンドノイズ音声では方言印象は想起されなかった。調波構造を保持するバンドノイズ音声に関しては、さらなる改良が必要。

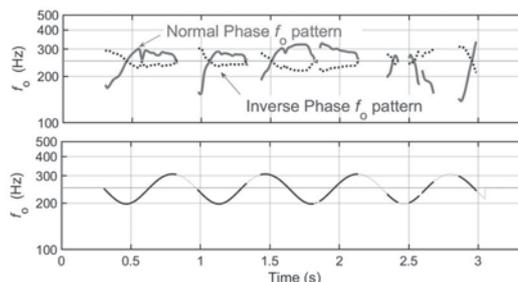


Fig. 1 Examples of original intonation and sinusoidal f_0 patterns.

3-8-2

3-8-2 自他の朗読音声か聞き手の没入感に与える影響の分析

The Influence of One's Own and Other's Voice on Immersion in Audiobook Listening

◎柳田耀, 長野瑞生, 藤田健一, 井島勇祐 (NTT)

- ◆先行研究では、音声の“自分らしさ”（聞き手と話者の類似性）が、聞き手の没入感を高める可能性が示唆されている [Kao+, 2021; Dragojevic+, 2024].
- ◆“自分らしさ”は、“同一化”（Identification：登場人物の視点から自分のものとして捉えること）を介して、“没入感”（Transportation：物語世界に入り込んだように感じる）に影響するとされている [Kao+, 2021]. しかし、聴取音声の種類（自分/他人）による構造の違いは十分に検討されていない.
- ◆本研究では、自分および他人の合成音声を用いた大規模主観評価実験を実施し、“自分らしさ”が没入感に与える影響を分析した.
- ◆その結果、他人の音声では、“自分らしさ”が同一化を高め、同一化が没入感を高める媒介構造が確認された一方、自分の音声では、“自分らしさ”から同一化への効果が限定的であった.

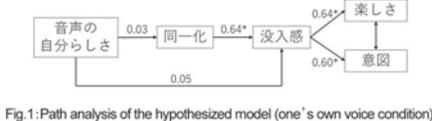


Fig.1: Path analysis of the hypothesized model (one's own voice condition)



Fig.2: Path analysis of the hypothesized model (other's voice condition)

3-8-4

3-8-4 音声の音響分析 色分類カードゲームの遊び方

How to play Acoustic Analysis Color-Classification Card Game

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

- ◆ことばのリハビリを行う言語聴覚士の養成校では、必修科目として音響学がある。しかしながら、音響学の授業は、学生の苦手科目の筆頭、その改善が求められている。
- ◆音響学の授業で、臨床で使う音声の音響分析を学ぶため、サウンドスペクトログラムの読み方を学ぶが、音声学の知識が十分でないと理解できない。また、発音記号の機械的暗記が、音声の調音様式の理解につながっていないことも多い。
- ◆「声のまようを読んでみよう！入門編」は実際のサウンドスペクトログラムを読むゲームだが、難易度がやや高いのが欠点であった。これらを解決するため、カードゲーム「声の色もよう探し」を作成した。その作成過程と想定される遊び方を紹介する。
- ◆具体的に、3種類の遊び方：1) 似た者カップル探し！ 2) どこで発音？ どうやって発音？ 3) 間違えやすい単語作りを紹介する。
- ◆今後の課題は、実は、音声の音響分析にはサウンドスペクトログラムの読み方以上に、前提となる音声学の知識を定着させることが大変であることを日々の授業で感じている。
- ◆今後は、より分かりやすい音声の分類方法についても考えていきたい。

3-8-3

3-8-3 発話における母音と子音の偏りが the speech-to-song illusion に与える影響

Effect of Vowel and Consonant Bias in Speech on the Speech-to-Song Illusion

☆籠谷春紀, 寺澤洋子(筑波大), △貞方マキ子(アムステルダム大)

- ◆the speech-to-song illusion(STS 錯覚)：発話音声を変換して音楽として聴かせることで音楽に錯覚する現象である。
- ◆音韻的偏り（母音偏り・子音偏り）を持つ刺激（例「カバダカババ叩く」）における錯覚の強度変化を調査した。
- ◆従来の教示方法「歌らしく聴こえたか？」から「Rapらしく聴こえたか？」に変更した場合の母音・子音の偏りとの相互作用を調査した。
- ◆結果1：音韻的偏りによって、有意に錯覚強度が上がった。
- ◆結果2：教示方法は音韻的偏りと相互に作用しなかった。
- ◆結果3：「Rapらしさ」を尋ねた場合、有意に錯覚強度が上がった。

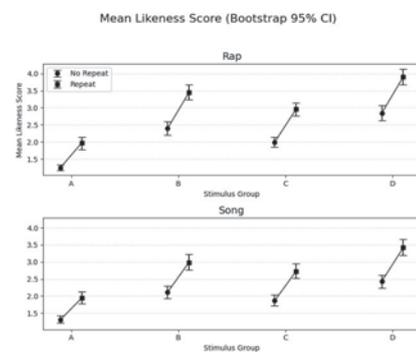


Fig.1: 反復聴取がSTS錯覚に与える影響

3-8-5

3-8-5 「新たな F1~F4 周波数値標準化による母音体系の形状と空間面積・母音間距離の自動算出シート」の特長と活用法

Features and Applications of an Automated Calculation Sheet for Vowel System Shape, Vowel Space Area, and Inter-Vowel Distances Based on a New F1~F4 Frequency Normalization Method

○齋藤孝滋(フェリス女学院大), 梁津祐(国士館大),
山下暖乃花, 原田麻未(フェリス女学院大)

- ◆この「新たな F1~F4 周波数値標準化による母音体系の形状と空間面積・母音間距離の自動算出シート」は、各母音素の該当音声各 6 回発音について、F1・F2・F3・F4 周波数値を入力するだけで、日本語標準語発話を基準とする(声道長等によるズレ補正)標準化を自動で行い「F1-F2」図、(F1~F3 周波数値を 2 次元化した)「HL-FB」図、(F1~F4 周波数値を 2 次元化した)「HLE-FBE」図を作成し、母音の平均値、母音間距離、母音体系については形状・重心・空間面積等を自動で算出し結果を提示するシートである。
- ◆本シートで提示される「HL-FB」図は、「HLE-FBE」図は、IPA 及びレントゲン・MRI との対応度が高いため、通言語的に実用レベルでの活用と、標準語教育・各言語教育・各言語研究領域への貢献が期待される。

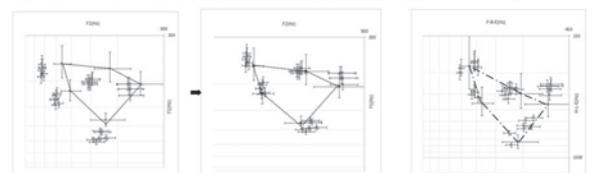


Fig.1: F1-F2 Vowel Space Plot Measured frequencies (raw values) Fig.2: F1-F2 Vowel Space Plot Normalized frequencies Fig.3: HLE-FBE Vowel Space Plot

3-8-6

3-8-6 多様な音素環境における母音の声道形状の変動分析

Analysis of variations in vocal tract shapes of Japanese vowels across various phonetic environments

☆軍司祐治, △浅野快輝, 朝倉麗仁,
竹本浩典(千葉工大), 前川喜久雄(国語研)

- ◆本研究では、3名の話者がATR503文の「サブセット」を発話した際の調音運動を記録したrtMRI動画から調音器官の輪郭を抽出して、話者ごとに各母音の声道形状の変動を主成分分析した。抽出された母音フレーム数は、/a/:651, /i/:425, /u/:331, /e/:277, /o/:448である。
- ◆その結果、/a/, /i/, /u/, /o/の声道形状の変動は、狭め部位では小さく、それ以外の部位で大きかった。これは、調音結合に伴う声道形状の変動が狭め部位を大きく変えない範囲で生じている可能性を示唆する。
- ◆一方、/e/では声道の狭め部位が明確に特定できず、声道形状の変動が大きな部位や方向が話者間で一致しなかった。これは、/e/が他の母音に比べて調音の指定があいまいな母音である可能性を示した。

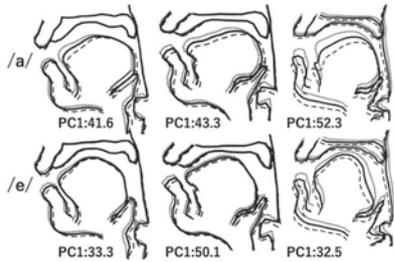


Fig. 1: /a/, /e/の調音運動のPC1と寄与率[%]
(数値:寄与率, 点線:-1.5SD, 実線:平均, 破線:+1.5SD)

3-8-8

3-8-8 ニューラルオーディオコーデック特徴量を用いた音声から話者特有の表情予測モデルの構築及び分析

Analysis of speaker-specific facial expression prediction model from speech using neural audio codec features

☆朴 浚鎔¹, 陳 晋升¹, 土井 啓成², 朴 炳宣²,
白旗 悠真², 橋 健太郎², 棟 楊¹, 齋藤 佑樹¹, 猿渡 洋¹
1) 東大院・情報理工, 2) LINE ヤフー

- ◆音声から話者特有の顔表情を BlendShape として予測するモデルを構築し、入力音声表現の情報構造が予測性能に与える影響を分析する。
- ◆自己教師あり学習 (SSL) 表現と Neural Audio Codec (NAC) 由来の連続音響特徴を比較し、話者固有の表情に対する有効性を検証する。
- ◆DITに基づく条件付き Flow Matching モデルを用い、音声特徴および話者埋め込みを条件として BlendShape 系列を生成する (Fig.1)。
- ◆全体的な精度においては SSL 表現が安定した性能を示す一方、特定の NAC 表現と上記の PCA による圧縮表現は、同等以上の表情再現精度や話者識別精度を示すことが確認された。

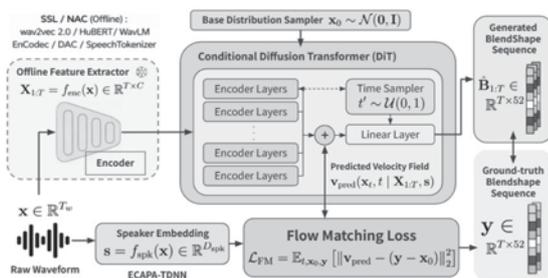


Fig.1: Architecture of BlendShape prediction model

3-8-7

3-8-7 上顎洞開放によって生じる伝達関数のピークとディップの生成要因の検討

Investigation of mechanism for generating peaks and dips in the transfer function induced by maxillary sinus opening

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

- ◆前報では、上顎洞を鉤状突起(UP)と膜様部(PF)まで開放したUP+PFモデルの伝達関数では、開放前と比べて600~800 Hz付近におけるピークのレベルが上昇するか1000 Hz付近に新たなピークが生成され、約1~2 kHzにディップが生成されることを報告した。
- ◆本稿では、生成されたピーク・ディップの成因を検討した。
- ◆UP+PFモデルの上顎洞容積を約1/2に減少させたUP+PF_sMSモデルの伝達関数では、ディップがUP+PFモデルの1240 Hzから1340 Hzへ移動し、他の帯域には大きな変化が見られなかった (Fig. 1)。
- ◆このことから、ディップは上顎洞の共鳴に由来すること、上顎洞は自然口を頸部、それ以外を胴部とするヘルムホルツ共鳴器で近似できることが示唆された。
- ◆UP+PFモデルの1000 Hzに生成されたピークの音圧分布では、咽頭、右上顎洞、蝶形骨洞、前頭洞に音圧の腹が形成され、声道全体ではなく、上顎洞を含む声道の部分的な共鳴が生じていた。
- ◆よって、上顎洞開放によって生じるピークおよびディップの双方に、上顎洞共鳴が重要な役割を果たしていることが明らかになった。

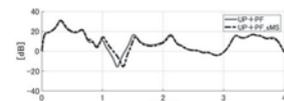


Fig. 1: UP+PF, UP+PF_sMSモデルの伝達関数

3-8-9

3-8-9 rtMRI 動画像からの話者非依存音声合成に向けた検討

Toward Speaker-Independent Speech Synthesis from rtMRI Videos

◎大谷祐人(東京理科大), 澤田隼(日本工業大),
大村英史(東京理科大), 桂田浩一(東京理科大)

real-time MRI (rtMRI)動画像を入力とする音声合成法が複数提案されており、音声器官の大部分を捉えられることから有望な結果を示している。しかし、従来の手法は話者依存のモデリングが一般的であり、未知話者のrtMRI動画像からは音声合成することができない。話者非依存音声合成の実現には、rtMRI動画像から言語情報と話者性を抽出する必要がある。本研究では、この実現に向けた検討として、rtMRI動画像から主に言語情報を抽出し、話者性を話者埋め込みで別途与える手法を提案し、未知話者への適用可能性を評価する。

提案手法では EfficientNetV2 で画像特徴量を抽出し、E-Branchformer で時系列処理を行う。話者埋め込みは X-vector を使用し、FILM により特徴量に条件付ける。実験では 75 話者の多話者 rtMRI データセットから品質の良好な 51 話者を選定し、テスト話者 5 名で評価を行った。実験の結果、読み上げ発話では WER が 3.1% から 7.6%、CER が 1.3% から 3.9% と比較的良好に発話内容が再現された。自然発話では、良好な話者で WER 4.7%、CER 3.7% が得られ、学習データ外の発話についても合成可能であることが示唆されたが、他の話者では誤り率の上昇が見られた。韻律については、Gross Pitch Error が 12.8% から 26.0% であり、音声から韻律が平坦化される傾向が見られた。

今後は言語情報と韻律の再現精度の向上とともに、話者埋め込みに依存しない話者非依存音声合成の実現に向けた検討を行う。

3-8-10

3-8-10 有限要素法を用いた声帯膜が発声に及ぼす影響の解析

Effects of the Vocal Fold Membrane on Phonation: A Finite Element Study
○吉永 司, △河原 彩(阪大・基礎工), 徳田 功(立命館大・理工学研),
△和田成生(阪大・基礎工)

- ◆声帯膜は、ヒトを除く霊長類の声帯で確認されている膜状構造であり、発声において重要な役割を果たすと考えられ、注目されている。
- ◆本研究では、有限要素法による声帯および声帯膜モデルと気流モデルとの連成解析を行い、先行研究の実験と比較して精度を検証するとともに、不規則振動における声帯膜の役割を明らかにする。
- ◆シリコン樹脂により製作された声帯膜の有無が異なる2種類の声帯モデルを再現するように形状を構築した (Fig. 1)。
- ◆シリコン樹脂の大変形を表現するため、材料モデルとして超弾性体の Neo-Hookean モデルを用いた。
- ◆声帯膜の存在により不規則振動が生じ、振動周波数の低下がみられるなど、先行研究と同様の傾向を本研究でも確認できた。

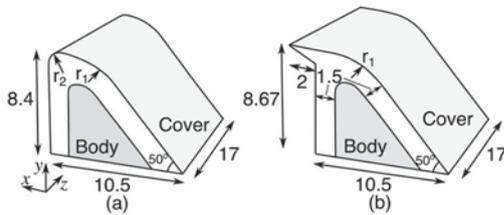


Fig. 1. Vocal fold model without membrane (a) and with membrane (b).

3-8-12

3-8-12 歌曲歌唱時のオペラ歌手における頸椎運動の統計分析

Statistical analysis of cervical spine motion in opera singers during art song singing

◎戸田菜月, 竹本浩典(千葉工大), 高橋純(大阪芸大)

- ◆オペラ歌唱では、1オクターブ跳躍進行時に声区転換点を含む高音域では音高の上昇に伴い、頸椎が大きく後弯することが知られている。
- ◆歌曲では声区転換点よりも低い音域も多く、跳躍度合いも小さいことが多いため、この結果と異なる可能性がある。
- ◆本研究では、歌曲歌唱時の頸椎運動を主成分分析 (PCA) と一般化加法混合モデル (GAMM) を用いて統計分析を行った。
- ◆統計モデルでは、瞬間的な音高と頸椎の後弯度合い (zPC1) の関係を検討するためのモデルAと、頸椎の後弯度合いの音型内での時間変化と瞬間的な音高変化を同時に検討するためのモデルBを構築した。
- ◆その結果、テノールは音高と音型の進行度合いに応じて後弯度合いを変化させるが、音高が高い条件では低い条件よりも後弯度合いを減少させる傾向がみられた (Fig. 1, 2)。
- ◆一方でバリトンは、歌唱中に一定の後弯度合いを保ち (Fig. 1, 2), PCAの結果から歌い出しと歌い終わりでは大きく後弯度合い増大・減少させる傾向がみられた。

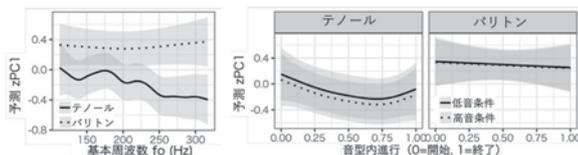


Fig. 1: Model A の推定結果

Fig. 2: Model B の推定結果

3-8-11

3-8-11 Physics-informed Neural Networks に基づく声帯振動シミュレーション

Simulation of vocal fold vibrations using physics-informed neural networks
◎横田和哉 (長岡技術科学大学)

- ◆近年、機械学習を用いた数値解析手法として Physics-informed Neural Networks (PINNs) に関する研究が進展している。
- ◆PINNs はメッシュフリーな手法であること、非線形性を自然に導入できることに加え、逆問題への適用性が高い利点を有する。
- ◆本研究では、声帯の two-mass モデルに基づき母音の生成シミュレーションが可能な PINN を構築した。
- ◆生成した母音/a/は実際の音声のフォルマント周波数範囲内であった。

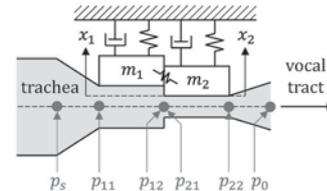


Fig. 1: Two-mass model of vocal folds.

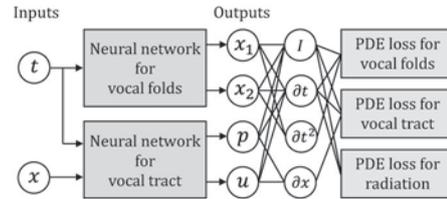


Fig. 2: Proposed PINN for vocal fold-vocal tract coupled analysis.

3-8-13

3-8-13 加齢性音声障害に対する発声訓練と視覚的提示が音響特微量に与える影響の検討

Impact of Vocal Training and Visual Feedback on Acoustic Characteristics in Age-Related Dysphonia

☆三藤早葵, 石光俊介 (広島市立大学院)

- ◆加齢性音声障害による QOL 低下や誤嚥リスクに対し、機能維持は喫緊の課題である。
- ◆視覚的な訓練成果提示を伴うリハビリ支援システムを構築し、発声訓練による改善効果を音響特微量に基づき定量的に検証することを目的とする。
- ◆本報では、VFE (Vocal Function Exercise) 訓練時の音響特微量を含めた発声機能の変化や発声持続の経過時間の提示の有無での違いを調査する。
- ◆訓練において、周波数および振幅の周期性変動を示す Jitter と Shimmer が抑制され、振動の規則性向上を定量的に実証した。
- ◆音声の清濁度を示す HNR の上昇と、全エネルギーに対する雑音成分の占有率を示す NNE の顕著な低下が確認された。
- ◆音響特微量解析の観点からも VFE が加齢性音声障害の根本的な改善に有効であると考えられる。

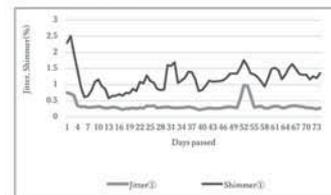


Fig. 1: Jitter and Shimmer in Patients with Age-related Dysphonia

3-8-14

3-8-14 日本語連続母音と中国語二重母音の音響的・調音的特徴の分析

Acoustic and Articulatory Analysis of Japanese Vowel Sequences and Chinese Diphthongs

◎孫静(神戸大/甲南大), 林良子(神戸大),
△金佳(関大), 北村達也(甲南大)

- ◆本研究では、日本語母語話者 11 名と中国語母語話者 10 名を対象に、日本語の連続母音と中国語の二重母音を音響的・調音的に比較した。
- ◆音響分析の結果、日本語の連続母音は各要素が単母音に近く、中国語の二重母音は両要素が単母音とは異なり(図1)、副次母音は「到達目標」ではなく「発音の方向」として機能していた。
- ◆超音波計測では、日本語で舌の高さ・前後位置の変化が中国語より有意に大きく、各母音が独立した調音目標を持つことが示された(図2)。この差異は、日本語がモーラ単位、中国語が一音節内で滑らかに二重母音を実現する音韻構造に起因すると推察される。

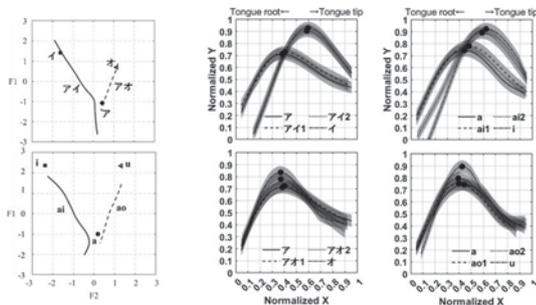


図1: 母音フォルマント図 (上図: 日本語の母音, 下図: 中国語の母音) 図2: 舌輪郭 (左図: 日本語母音, 右図: 中国語母音)

3-8-15

3-8-15 超音波装置とリップカメラを用いた中国語音節における調音器官運動の可視化とオープンデータベース構築の試み

Visualization of articulatory movements in Chinese syllables using ultrasound imaging and a lip camera: an attempt to build an open database

○李璋昱(上智大), 溝口愛(前橋工科大学),
守本真帆(中央大学), 荒井隆行(上智大学)

- ◆本研究の目的は、中国語母語話者による普通話に現れる全音節について、音声および調音器官運動を捉えた動画データを収集し、インターネット上でオープンデータベースとして公開することである。
- ◆その第一段階として、超音波装置とリップカメラを用いて話者1名による中国語単母音音節発話の収録を試みた。その成果および現時点での課題を報告する。Fig 1は超音波装置で録画した中国語a.gāng/「阿剛(人名)」を発話する際のキーフレームの舌の形を示している。

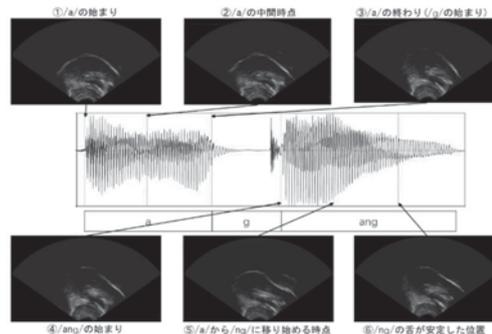


Fig.1: Tongue shapes recorded with an ultrasound system and the corresponding acoustic waveforms (the right side of each ultrasound image represents the tongue tip, while the left side represents the tongue root)

3-8-16

3-8-16 韓国語を母語とする日本語学習者において「/u/にウではなくウをあてはめる」発音指導の有効性 —音響音声学・聴覚音声学的分析をとおして—

Evaluating the Effectiveness of eu-Based Instruction for Japanese /u/ Production in Korean Learners: Acoustic and Auditory Phonetic Evidence

☆梁 津祐(国土館大院), 齋藤 孝滋(フェリス女学院大)

- ◆韓国語母語話者は、日本語の/u/を本来の非円唇母音[u]ではなく母語ウに由来する円唇母音 [u] で発音し、不自然に聞こえることがある。本研究では、この改善を目的に、/u/ をウの [u] で発音した場合と、韓国語ウの [u] に置き換えた場合を比較し、音響分析と日本語母語話者による聴覚評価を行った。
- ◆その結果、/u/を[u]に置き換えた5母音体系はNHKアナウンサー10名の平均的体系と一致し、聴覚的にも標準語として自然と評価された。
- ◆以上より、「/u/にウではなくウをあてはめる」指導が有効であることが示された。

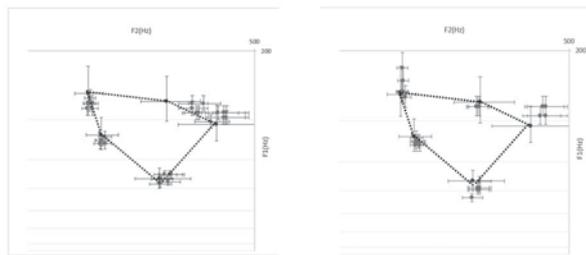


Fig.1: F1-F2 Vowel Space Plot —/u:/, [u]— Fig.2: F1-F2 Vowel Space Plot —/u:/, [u]—

3-8-17

3-8-17 「聞き取れないから話せない」の真偽について —日本人英語学習者における知覚と生成の音声認識のギャップ調査分析—

On the Truth of "I Can't Speak Because I Can't Listen" —An Analysis of the Gap Between Perception and Production in Speech Recognition Among Japanese Learners of English—

○浅野恵子(順天堂・医)

- ◆英語習得で、「聞き取れないから話せない」と言われることがあるが、その真偽は不明で、知覚と生成の音声機能の混同が遠因であろう。
- ◆日本人学習者の両機能の音声認識の困難さギャップを調査・分析する。
- ◆対象者は、日本人英語学習者の大学生1,2年生37名。TOEFL ITPを受験し、500点±10点の英語習熟度群に属している。
- ◆調査はアンケート形式で、①性別、②年齢、③海外滞在経験の有無、④各発音及び聴取困難の有無、⑤有の場合、各困難さの具体例を3つ(単語もしくは個別音素など)、⑥自由記述で感想を回答している。
- ◆対象者が発音は94%、聴取は81%が困難であると回答しているが、各困難要因の項目と個別対象者により相違が大きい。

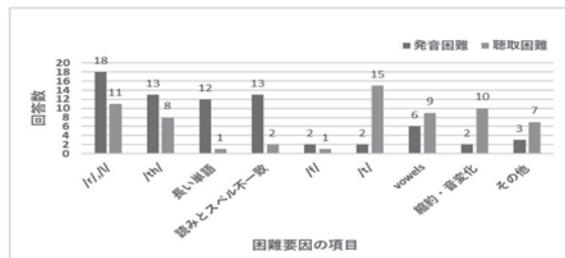


Fig.: Differences in Individual Difficulty Factors in Speech Recognition among Intermediate Japanese Learners of English.

3-9-1

3-9-1 (招待講演) 音響性難聴のメカニズムと具体的事例

The mechanism of noise-induced hearing loss and specific cases

○和田 哲郎(筑波大)

- ◆音は聴覚という感覚器にとって適刺激であるが、強すぎる音は蝸牛に障害を与え難聴が発生する。
- ◆音による難聴は、ばく露音響レベルと発症に至る時間経過から、急性音響性聴器障害と慢性音響性聴器障害の2つに分けられる (Table 1)。両者は経過が異なり、分けてとり扱うことが望ましい。
- ◆100 dB を越える強大音では急性音響性聴器障害が起こりうる。そのメカニズムは機械的障害と代謝障害で説明される。
- ◆85 dB 以上の音に長時間、長期間繰り返しばく露されることによって、5~15 年以上の経過で騒音性難聴が起こりうる。近年では、職業性騒音以外にも、レジャー騒音による非職業性騒音性難聴も注目されている。これらの音負荷により代謝障害が起こり、特に解剖学的に最も虚血に陥りやすい高周波数帯域に閾値上昇が起こりやすい。
- ◆急性と慢性に分け、具体的事例を示しながら、音響性聴器障害について概説する。

Table 1 Classification of noise-induced hearing loss.

分類	負荷音響レベル	原因となる音	ばく露時間
急性音響性聴器障害 (広義の音響外傷)			
(狭義の) 音響外傷	130 dB(A)~	銃火器、爆発など	瞬間的
(その他の) 急性音響性難聴	100~120 dB(A)	コンサートなど	数分~数時間
慢性音響性聴器障害			
職業性騒音性難聴	85 dB(A)~	職業性騒音	5~15 年以上
非職業性騒音性難聴	不明	音楽など	不明

(日本聴覚医学会編 急性感音難聴診療の手引き 2018 年版より引用・一部改変)

3-9-3

3-9-3 (招待講演) 花火師における騒音個人ばく露測定

Personal noise exposure measurement among fireworks operators

○永野 千景, △外川 舞, △石塚 洸太郎, △堀江 正知(産業医大), △井奈波 良一(岐阜大)

花火大会に従事する花火師を対象に、実際の騒音ばく露を明らかにするため個人ばく露測定を実施した。5 名の花火師に 1 分毎の等価騒音レベルを記録するドーズメータを装着し、作業中のばく露を計測した。その結果、花火大会中は継続的に 85~90 dB(A)、一時的には 100 dB(A) を超えるばく露が確認された。号砲時には 130 dB 以上の衝撃音も記録された。聴力検査では一過性閾値上昇の有無は確認できなかったが、ばく露が長期化・頻回化すれば騒音性難聴のリスクがあると推察された。花火師は副業の季節労働者が多く、安全衛生管理体制が不十分なため、少なくとも打ち上げ作業中は耳栓等の聴覚保護具の着用が推奨される。

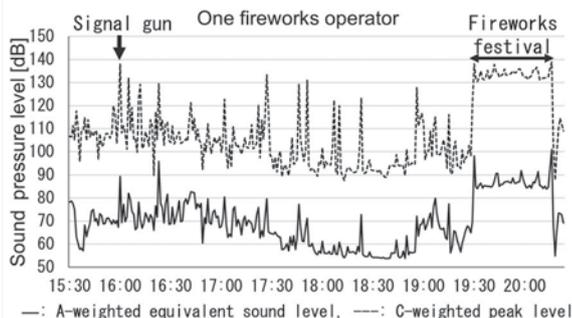


Figure. Personal noise exposure measurement result for a fireworks operator

3-9-2

3-9-2 (招待講演) 騒音障害防止ガイドライン改訂で意図する個人ばく露量の推定・制御について

Estimation and control of individual noise exposure intended by revision of Noise Hazard Prevention Guideline

○柴田延幸(安衛研)

- ◆本報では、約 3 年が経過した騒音障害防止ガイドラインの改訂で新たに取り入れられた個人ばく露測定について、その考え方や個人ばく露計を用いた測定方法を概説するとともに適切な聴覚保護具の選定方法について説明する。
- ◆屋内作業場では、作業や騒音源が移動する場合など、従来の作業環境測定では作業者のばく露する騒音レベルを正確に測定することが困難な場合に個人ばく露測定が推奨される。特に手持ち動力工具等を使用する作業場では、「個人ばく露測定による等価騒音レベルの測定」にもとづいて測定することが望ましい。
- ◆屋外作業場では騒音源からの距離による騒音レベルの変動が大きく、従来の作業環境測定における B 測定のような定点測定によって作業者のばく露騒音レベルを把握することが困難であるため、このような場合は個人ばく露測定が推奨される。
- ◆作業場における個人騒音ばく露の測定方法としてジョブごと、タスクごと、一日全体の 3 つの測定方法がある。実際の測定現場では作業形態に応じてこれら 3 つの測定方法を上手に組みあわせることになる。
- ◆聴覚保護具を選定する上で留意すべき点は、聴覚保護具の使用によって、1) 耳に届く実効騒音レベルが 85 dB 未満となること、2) 音声コミュニケーションや警告音を阻害しないようにするために、耳に届く実効騒音レベルが 70 dB 以下にならないようにすることである。

3-9-4

3-9-4 (招待講演) 歯科医療従事者の音響ばく露一現状と課題一

Noise exposure among dental healthcare workers: Current status and challenges

○山田朋美(大阪大院・歯)

- ◆歯科診療室では、歯科用切削機器や吸引装置などにより、歯科医療従事者は日常的に騒音へ曝露されている。
- ◆COVID-19 パンデミック以降、感染予防対策として口腔外吸引装置の使用が常態化し、診療室の音環境が変化している。
- ◆厚生労働省による「騒音障害防止のためのガイドライン」が 2023 年に改訂され、すべての職場が対象となっている。
- ◆本研究では、騒音ばく露および両耳マイクロホンを用いて歯科医療従事者の騒音ばく露を測定し、その現状と課題について報告する。

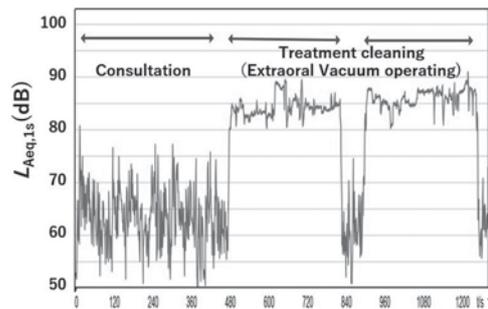


Fig.1 : Noise exposure levels during dental cleaning with an extraoral vacuum in operation

3-9-5

3-9-5 (招待講演) 保育・学童施設における 保育者・指導員の音響ばく露

Personal sound exposure levels of nursery school teachers and after-school care instructors during nursery/child care activities

○横山 栄, 小林知尋(小林理研), 船場ひさお(駿河台大)

保育園や学童保育施設における音環境については、騒音調査や快適性評価の事例報告はあるが、保育者の騒音性難聴に配慮した一日を通しての音響ばく露の測定事例は少ない。

本調査では、認定こども園および学童保育施設の協力を得て、保育者および指導員の保育活動中の音響ばく露の実感を調査し、音響性難聴に関する意識調査も行った。また、学童保育施設では壁面の吸音処理前後で調査する機会も得られ、吸音処理の効果も検証した。

音響ばく露の調査の結果、音響ばく露レベルはいずれも職場騒音としての許容限度値 85 dB を下回っていたものの多くが 80 dB 前後であり、快適性だけでなく音響性難聴も懸念されるばく露状況であった。また、アンケート調査の結果、耳鳴りなどの症状を自覚している回答も多く、多くが音環境による聴力への影響に懸念を抱いている状況であった。

同じ環境で過ごす子ども達への影響も考えると、何かしらの対策を必要とする状況であると考えられ、今後も検討を続けていきたい。

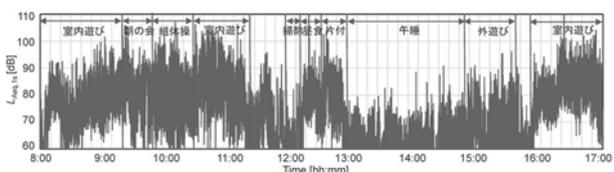


Fig. 1: Measured sound exposure levels of a nursery teacher during nursery care activities. ($L_{Aeq,9h}$: 84.7 dB, L_{Cpeak} : 136.4 dB, DOSE: 108.2 %)

3-9-7

3-9-7 音楽ホールにおけるオーケストラ奏者の音響ばく露

Personal sound exposure levels of orchestra musicians during rehearsal performance in a concert hall

○横山 栄, 小林知尋(小林理研)

令和5年に厚生労働省が「騒音障害防止のためのガイドライン」を改訂し、職場環境全てが対象となり、一部では個人ばく露計による測定も認められた。この改訂以降、筆者は騒音職場における作業員だけでなく、保育園や学童保育の保育者、楽器奏者を対象とした音響ばく露測定も試み、音響性難聴の危険性の有無について、調査を進めている。

本研究では、プロのオーケストラ奏者の協力を得て、ホールステージ上における演奏時の音響ばく露測定、および演奏音が聴力に及ぼす影響(音響性難聴の危険性)に関する意識調査を実施する機会を得た。

その結果、奏者の音響ばく露は騒音職場で聴覚保護具の使用が義務付けられる限度値 (85 dB) を超えており、音響性難聴が懸念される状況であった。また、意識調査の結果では、多くのプロ奏者が演奏音による聴力への影響を感じているものの誰にも相談できず、対策法もわからないといった声が多く寄せられた。引き続き、演奏者の音響性難聴防止のために、工学的立場からできることを探していきたい。

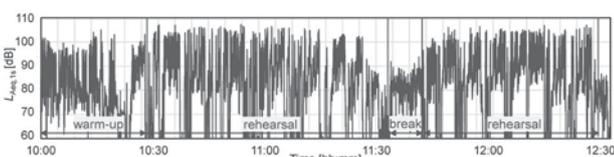


Fig. 1: Sound exposure levels of a trumpet player during rehearsal performance in a concert hall. ($L_{Aeq,2h}$: 94.0 dB, L_{Cpeak} : 123.1 dB, DOSE: 255 %)

3-9-6

3-9-6 吹奏楽部における日常的な楽器演奏による音響ばく露の現状調査と対策の考察

Sound exposure survey of university wind orchestra members in daily instrumental practice and consideration of countermeasures for hearing loss

☆加藤優稀(日大院), 橋本 修(日大), 金 政秀(武蔵野大), 横山 栄(小林理研)

◆厚生労働省は、騒音環境下で働く労働者には、騒音性難聴防止のため聴覚保護具着用などの対策を行うことを義務付けている。しかし、楽器演奏者に対する規定はなく、対策法も示されていない。

◆本研究では、日常的な楽器演奏による音響ばく露の現状を調査し、音響性難聴に関する意識調査も行った。また、演奏時の耳栓使用の適用可能性も調査した。その結果、以下の知見を得た。

◆練習時の音響ばく露レベルは平均で90 dBを超え、ピーク値は120 dBを超えていた。約3時間の練習でDOSEは160%を超えており、音響性難聴の危険性が高い。

◆音響性難聴に関する意識調査の結果、演奏音による難聴の危険性は44%が認識しており、演奏音による聴力の衰えを感じる人も16%いたが、聴覚保護具の使用はなかった。音響ばく露の状況から、演奏時には聴覚保護具の着用などの対策が必要である。

◆耳栓着用実験の結果、積極的に耳栓を着用したいという意見は得られなかったが、打楽器やフルート奏者については耳栓着用による遮音効果は実感されていた。楽器の奏法により耳栓着用の向き・不向きがあると考えられ、楽器ごとに対策を検討することも有用である。

◆同じ演奏でも、部屋や室内の演奏位置によっても音響ばく露が異なることもわかっており、演奏者が演奏音による音響性難聴の危険性を理解し、正しく対策できるよう、今後も検討を続けたい。

3-9-8

3-9-8 楽器演奏者の日常的な騒音ばく露調査-打楽器演奏者を対象として

Survey of Daily Noise Exposure Among Musical Instrument Performers: A Focus on Percussionists

☆新井一輝, 宮崎宥楓(神奈川工科大), △窪耕一郎(RCC DRUM SCHOOL), 三浦貴大(産総研), 廣江正明(小林理研), 上田麻理(神奈川工科大)

◆近年、PA システムの発展や音楽文化の普及により、音楽演奏者が日常的に接する音量は増大しており、高音量音楽への長期的ばく露によって生じる音楽誘発性難聴(MIHL)が深刻な健康課題となっている。特に打楽器奏者は、至近距離で発生する強い衝撃音に繰り返しさらされることから、他の楽器奏者と比較して高い聴覚障害リスクを有することが指摘されている。

◆先行研究では、打楽器奏者において120 dBを超えるピーク音圧へのばく露や、高音域における聴力低下、耳鳴り、音過敏などの症状が高頻度で報告されている。一方で、ドラムセット特有の放射指向性や、日本の小規模な練習空間における反射音の影響を含めた実測データは十分に蓄積されていない。

◆そこで本研究では、ジャズドラム演奏者を対象に、レッスン、ライブリハーサル、自宅練習といった日常的な演奏環境における音響ばく露量を測定し、その特徴と聴力検査結果との関連性を定量的に検討することを目的とする。



3-9-9

3-9-9 楽器演奏者における耳栓装着の効果 -聴力保護と演奏パフォーマンスへの影響-

Effects of Earplug Use in Musicians:
Impacts on Hearing Protection and Musical Performance

☆宮崎杏楓(新井一輝(神奈川工科大)), △窪耕一郎(RCC DRUM SCHOOL),
三浦貴大(産総研), 廣江正明(小林理研), 上田麻理(神奈川工科大)

- ◆演奏家は長時間・高音圧環境に曝露される職業であり、聴覚保護対策の確立が重要である。
- ◆本研究では、耳栓装着が演奏者の演奏感覚と心理評価に与える影響を、プロのドラム、サクセス、ピアノ演奏者を対象に検討した。
- ◆5段階評価および自由記述分析の結果、耳栓は安心感の向上や聴覚疲労の軽減に寄与する一方、音像認知や演奏時の一体感に影響を及ぼすことが示された (Fig. 1)。
- ◆本研究は、演奏現場における実用的な聴覚保護方策の検討に基礎的知見を提供する。

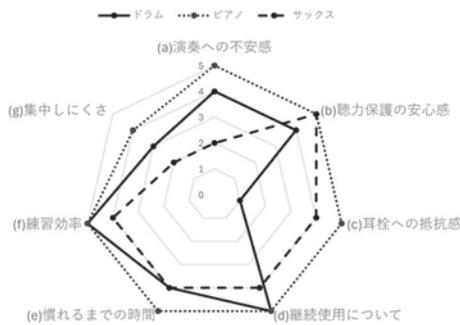


Fig.1: Individual Practice Questionnaire (Psychological Aspects)

3-9-11

3-9-11 振幅変調音に基づく回転翼機衝突回避を 目的としたマイクロホンアレイ配置の 実験的検証

Experimental Verification of Microphone Array Configuration for Rotorcraft
Collision Avoidance Based on Amplitude-Modulated Sound

☆永田健太郎(中央大院), △山本弘, 田辺総一郎, 戸井武司(中央大)

- ◆既報では、シミュレーションにより回転翼機衝突回避を目的としたマイクロホンアレイ (MA) 配置を検討した。
- ◆本研究では、MA 構成シミュレーションに基づく検証実験を行い、有効性を確認する。
- ◆ドローン機体に設置する試作 MEMS マイクロホン対面配置にすることで、機体振動に起因した固体伝搬音を抑制し、精度のよい音圧が測定できる。
- ◆MA 構成シミュレーションと検証実験の比較を Fig. 1 に示す。MA 構成パラメータであるマイクロホンの個数4個、間隔2 cm の条件が適切と分かり、本 MA 配置手法の有効性が確認できた。

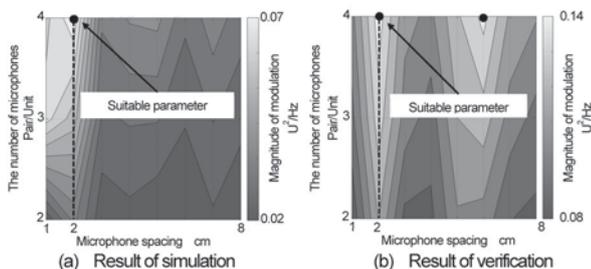


Fig. 1: Simulation and verification results

3-9-10

3-9-10 ロバスト主成分分析に基づく接近回転翼機の 振幅変調音強調による音源探知

Sound Source Detection of Approaching Rotorcraft Using
Amplitude-Modulated Sound Enhancement Based on
Robust Principal Component Analysis

☆永田健太郎(中央大院), 田辺総一郎, 戸井武司(中央大)

- ◆ドローンの普及に伴い、高度 150 m 未満を飛行するヘリコプタとの衝突の危険性が高まり、ヘリコプタ探知のためのドローン搭載用小型軽量かつ低消費電力センサが求められている。先行研究では、Fig.1(a) に示す約 2.5 km 先を飛行しているヘリコプタの音源探知を可能としたが、さらなる探知距離の延伸が必要である。
- ◆本研究では、Robust Principal Component Analysis (RPCA) を先行研究で提案されたアルゴリズムに追加し、探知距離の延伸を図る。
- ◆本研究で提案する新たなヘリコプタ振幅変調音の探知アルゴリズムにより、マイクロホンアレイを用いたヘリコプタ動作音の強調と、ヘリコプタ以外に起因する振幅変調を抑制することで、Fig.1(b)に示すように約 3.3 km 先のヘリコプタの音源探知が可能であることを確認した。

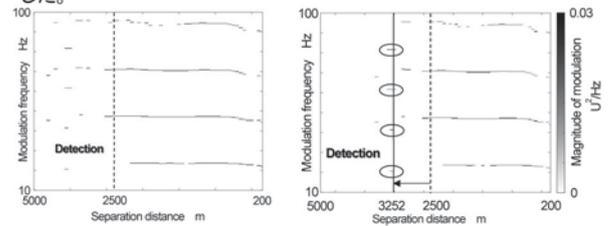


Fig.1: Relation between separation distance and modulation frequency

3-9-12

3-9-12 ナノファイバー不織布の骨格振動が及ぼす 吸音特性への影響

The effects of vibration of nanofiber non-woven sheets on sound absorption characteristics

◎吉田朋純, 赤坂修一, △浅井茂雄(東京科学大)

ナノファイバー不織布は繊維径が約 $1\mu\text{m}$ 以下の繊維であり、高い空孔率と比表面積を持つため、優れた吸音性能が期待される。そこで当研究室では、ナノファイバー不織布の吸音特性について検討し、薄くとも高い吸音性能を示すことや、特に繊維径が細かい試料では低周波数域で吸音ピークを示すことを確認した。また、音響特性モデルを用いた解析より、吸音特性には試料の振動が影響していることが示唆された。そこで本研究では、本研究では、PS ナノファイバー音響パラメータの測定を行い、ナノファイバー不織布で観測される吸音ピークのメカニズムを検討した。Figure 1 は、ポリスチレン(PS)からなるナノファイバーの吸音率の実測値と、音波が試料内を通過する際に変化する位相を示したものである。これより、吸音ピークは表面反射波と剛壁で反射し再放射した波の干渉に起因することが明らかとなった。

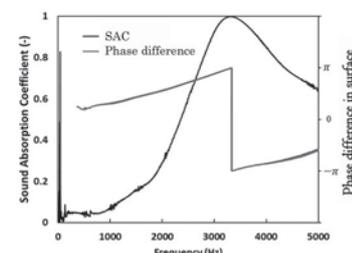


Fig.1: Sound absorption coefficient of measured and phase difference in surface of PS nanofiber.

3-9-13

3-9-13 弾性を有する繊維系中間層を用いた二重壁構造の遮音性能

Study on Sound Insulation of Double-wall Structure with Elastic Fiber Layer

◎市川友己, △中村利幸(林テレンプ), 山本崇史(工学院大・工)

- ✓ 自動車の防音手法には多孔質系繊維材の中間層と遮音層とを積層させた防音部品とボディパネルとの二重壁構造が多くみられる。
- ✓ 広く自動車の防音部品に用いられる多孔質系繊維材を中間層とした二重壁構造に着目し, 低～中周波数(200～1kHz)の遮音性能を向上させられる可能性について実験的に検討を行った。
- ✓ 繊維方向(Fig.1)を変えることで弾性を大きくした中間層について, 面密度や圧縮量などの変化による遮音性能の違いを考察する。
- ✓ 中間層の面密度や圧縮量を変えることで, 低～中周波数帯域の遮音性能を向上(Fig.2)させる可能性が示された。

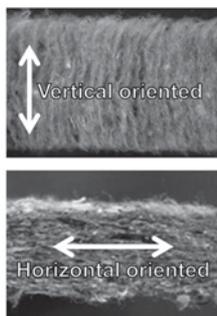


Fig.1 Fiber orientation cross-section for the sample in the layer.

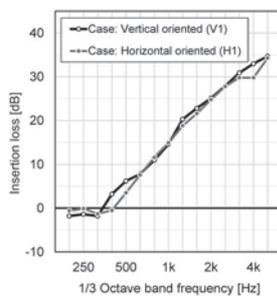


Fig.2 Example of insertion loss (transmission loss) measured for different fiber orientations.

3-10-2

3-10-2 音声の緊迫感知覚に寄与する変調周波数成分に関する検討

Study on Modulation Frequency Components of perception of urgency in speech

☆房野早希, 木谷俊介, 鷗木祐史(北陸先端大)

- ◆ 目的: 音声の緊迫感知覚の手がかりとなる時間振幅包絡線(TAE)の変調周波数帯域を明らかにすること。
- ◆ 方法: 雑音駆動音声(NVS)を用いた聴取実験を行い, (1) TAEに対する変調フィルタバンク(MFB)の利用がNVSの緊迫感知覚に影響を与えるかどうか, (2) MFBを利用した変調周波数の帯域制限により, NVSの緊迫感知覚がどのように変化するか調査した。
- ◆ 結果: 聴取実験より, 次のことがわかった。
 - (1) TAEに対するMFBの利用はNVSの緊迫感知覚に影響を与えないこと
 - (2) 0 - 64 Hzの変調周波数帯域を持つNVS ($a_{M1} \sim d_{M1}$, $a_{W1} \sim d_{W1}$)と比較して, 0 - 16 Hzの帯域を持つNVS ($a_{M2} \sim d_{M2}$, $a_{W2} \sim d_{W2}$)は同程度の緊迫感の程度を持つこと。0 - 16 Hzより狭い帯域を持つNVS, また0 - 16 Hzの範囲で一部帯域が抜けた帯域を持つNVSは緊迫感の程度が低い傾向にあること
- ◆ 結論: 緊迫感知覚に寄与する重要な変調周波数帯域は0 - 16 Hzである。

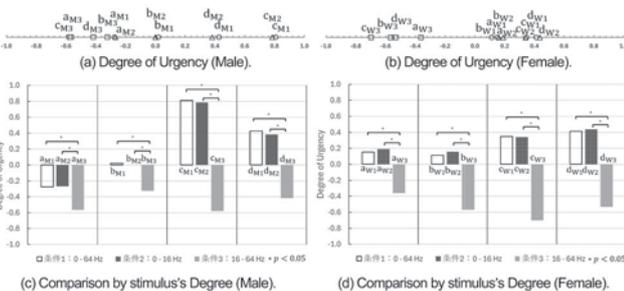


Fig.1: Urgency perception of NVS.

3-10-1

3-10-1 聴覚時間分解能検査の作成 -データの更なる検討と今後の展開-

Development of tests of human auditory temporal resolution -Additional findings and future directions

○森 周司(九州大), 森本 隆司(リオン), △岡本 康秀(済生会中央病院・慶応大), △神崎 晶(国立病院機構東京医療センター)

- ◆ 2025 年秋季研究発表会の招待講演で紹介出来なかった内容を取り上げる。
- ◆ 無音検出閾値と、振幅変調検出の測度であるカットオフ周波数とピーク感度の関係を 121 名の健聴者と 22 名の難聴者で検討した。その結果、無音検出閾値とピーク感度には関連があるが、ピーク感度とは関連がないことが分かった。
- ◆ 実験で測定した無音検出閾値より確率密度関数を推定したところ、その 99%点を超える閾値を示したのは健聴者 121 人中 2 名であったのに対し、難聴者 22 名全員の閾値が 99%点を超えた。このことは、99%点が無音検出閾値の基準値となる可能性を示唆する。
- ◆ 検査のプロトタイプのイメージ図を示した。検査では対象者の年齢や聴力などの情報を入力した後、練習と測定に進み、全測定終了後にその結果が出力される。

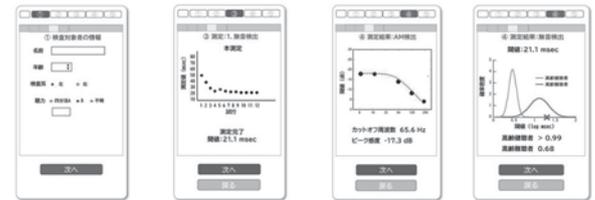


Fig.1: Images of auditory temporal resolution tests.

3-10-3

3-10-3 音の快・不快を予測するための 感覚的快さの評価実験とその結果に基づく予測モデルの検討

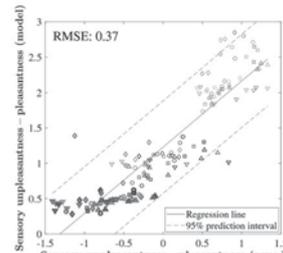
Experimental evaluation for predicting sensory pleasantness/unpleasantness and examination on prediction model based on the results

☆松尾洋社, 上江洲安史, 木谷俊介, 鷗木祐史(北陸先端大)

- 【問題】先の検討から, 快・不快を一軸で表せることが分かっている。しかし, 快・不快の評価を行うためには, 利用した刺激条件の範囲が十分ではなかった。
- 【目的】実験の刺激条件の範囲を拡張し, 感覚的快さの再評価を行うこと。結果を快・不快の評価とみなし, これを説明可能な快・不快予測モデルを構築すること。
- 【方法】
 - ① 刺激条件の範囲を拡張して作成された音刺激に対する感覚的快さの評価実験
 - ② ①の結果を用いた音の快・不快予測モデルの構築および, 主観評価値 - モデル推定値間の RMSE によるモデルの当てはまりの評価

【結果】Fig.1 から

- ① 快・不快の評価を行うために十分な刺激条件の範囲を設定できた
- ② 構築した予測モデルは評価結果を高い精度で説明できることがわかった



$$\hat{V}_p = e^{\alpha_{p,0}} e^{\alpha_{p,1}R} e^{\alpha_{p,2}S} e^{\alpha_{p,3}K} e^{\alpha_{p,5}FS}$$

五つの音質評価指標

R: ラフネス S: シャープネス

K: トナリティ N: ラウドネス

FS: 変動強度

Fig.1: Estimation results of sensory pleasantness/unpleasantness with \hat{V}_p

3-10-4

3-10-4 ノイズ環境下で生じる聞き間違いに 影響する子音の誤認知パターンの検討 Patterns of Consonant Misidentification Affecting Speech Perception in Noise

☆岩永真弥, 小川景子(広島大)

- ◆ノイズ環境下で生じる聞き間違いと子音の誤認知の関連について検討を行った。
- ◆母音よりも子音の誤認知が有意に多く、誤認知パターンには「子音の変化」「子音の脱落」「子音の付加」の3パターンが観察された。
- ◆呈示音圧差の拡大により背景音 (30 dB) と標的刺激の分離が容易になり始めるタイミング (31-39 dB) で各誤認知が特徴的な変化を示し、「子音の変化」が最も多く発生した。
- ◆誤認知の出現特徴から、ノイズ環境下におけることばの誤認知には、子音よりも母音に対する注意が先行することが影響している可能性がある。

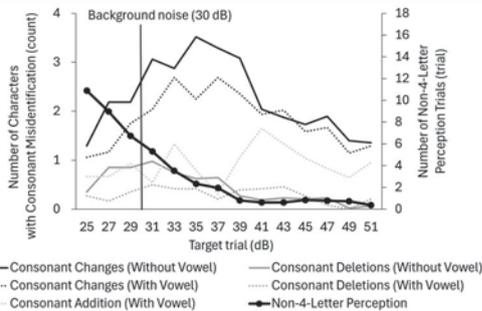


Fig.1: Changes in the Total Number of Misidentified Characters for Each Consonant as a Function of Presented Sound Pressure Level

3-P-2

3-P-2 動的圧縮型ガンマチャープ聴覚 フィルタバンクを用いたラウドネスの 計算モデルの検討

Study on computational model of loudness using
dynamic compressive gammachirp auditory filterbank

◎磯山拓都(産技高専), 木谷俊介, 鶴木祐史(北陸先端大)

【背景】我々は ISO 532-2 に基づきガンマトーン聴覚フィルタバンク (GTFB) 及び解析的ガンマチャープ聴覚フィルタバンク (GCFB) を利用したラウドネスの計算モデルを構築してきた。今後、ISO 532-3 を見据えた非定常音への拡張には聴覚フィルタの時間的挙動及びレベル依存の圧縮特性が重要な課題である。

【目的】動的圧縮型ガンマチャープ聴覚フィルタバンクを利用したラウドネス計算モデル (dcGC モデル) を構築し、ISO 532-2 で規定された定常音に対するラウドネスに整合したラウドネスを計算できるか評価すること

【方法】dcGC モデルを ISO 532-2, GTFB を利用したラウドネスの計算モデル (GT モデル) 及び、GCFB を利用したラウドネスの計算モデル (GC モデル) と比較した。

【結果】Fig.1 より、dcGC モデルは定常音に対し ISO 532-2 で規定されたラウドネスに整合したラウドネスを計算できることを確認した。

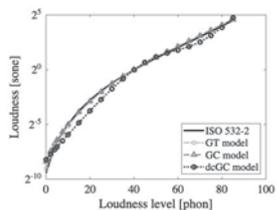


Fig.1: Relationship between loudness level and loudness

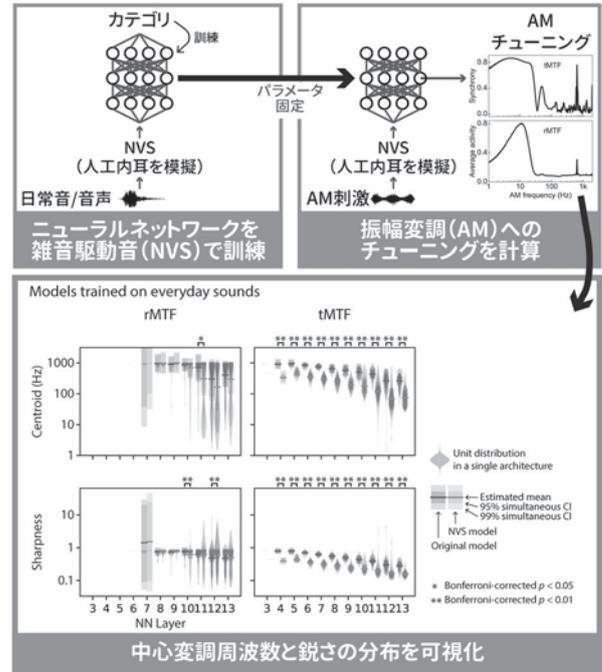
3-P-1

3-P-1 Noise-vocoded 音で訓練した ニューラルネットワークにおける 振幅変調チューニング

Amplitude Modulation Tuning in a Neural Network
Trained on Noise-Vocoded Sounds

○上村 卓也(NTT CS 研), 寺島 裕貴(NTT CS 研),

古川 茂人(静岡社会健康医学大学院大, 静岡県立総合病院, NTT CS 研)



中心変調周波数と鋭さの分布を可視化

3-P-3

3-P-3 拍節的な先行音列が時間順序判断に与え る影響について

Effects of Preceding Auditory Beat Sequences on Audiovisual Temporal
Order Judgment

◎田中紘人, 蓮尾絵美, Gerard B.Remijn(九州大芸術工)

- ◆視聴覚の時間順序判断において、判断対象となる刺激に先行した拍節的な音刺激が時間順序の判断に影響を与えるのかを調べることを目的とした。
- ◆拍節的な先行音列の有無および拍と刺激表示タイミングの関係を操作し、拍なし (No Beat) 条件、拍同期 (On beat) 条件、拍非同期 (Off beat) 条件の三条件を設けて、時間順序判断課題を行ってもらった。
- ◆主観的等価点 (PSS) については有意な差が見られなかった。一方で、丁度可知差異 (JND) の大きさは「拍非同期 < 拍同期 < 拍なし」となった (Fig. 1)。特に、拍同期条件よりも拍非同期条件において小さい JND が得られた点は、拍による時間的予測が単純に拍との同期性によって決まるものではない可能性を示唆している。

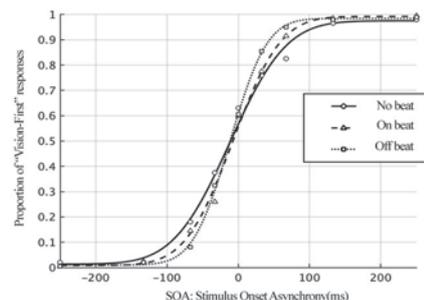


Fig. 1 Psychometric functions fitted over the average results (n=20)

3-P-4

3-P-4 ピアノの打鍵法が倍音構造と音色知覚に及ぼす影響 ～ストレート打鍵とスライド打鍵の比較～

Effects of Piano Keystroke Techniques on Harmonic Structure and Timbre Perception

☆阿部愛美香(龍谷大), △丸尾祐嗣(ピアニスト), 永瀬純也(龍谷大)

- ◆本研究では、同一ピークサウンドレベル条件下において打鍵法の違いのみでピアノ音の音色が変化するかという課題に対し、ストレート打鍵およびスライド打鍵を用いて実験的に検証した。
- ◆グランドピアノの中央ド(C4)の単音を録音し、ピークサウンドレベルが93 ± 0.3 dBの音のペアを5組用意した。
- ◆FFTによる周波数解析の結果、第13倍音において約7 dB、第15および第16倍音において約5 dBの音圧レベル差が確認された。
- ◆知覚実験(2AFC)を3名の参加者に対して実施した結果、正答率は100%、85%、100%となり、チャンスレベル(50%)を大きく上回る結果が得られた。
- ◆以上より、同一音圧条件下であっても、ストレート打鍵とスライド打鍵の間に、倍音構造および音色知覚の顕著な差が生じることが明らかとなった。

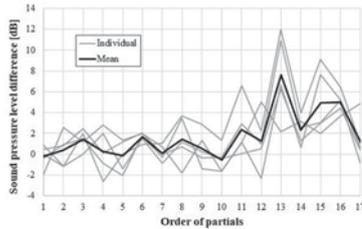


Fig.1 Level difference between straight and slide keystrokes

3-P-6

3-P-6 個人の特性を考慮した背景音が課題遂行に与える影響

The effect of background sound based on individual differences on task performance

☆伊藝雄太郎, 程島奈緒(東海大)

- ・背景音のテンポが作業効率と心理状態に与える影響を明らかにする
- ・条件1: 刺激音なし, 条件2: 固有BPM(題解答にかかる時間である作業時間間隔から算出), 条件3: 固有BPM × 1.3を背景音に設定し、数字を昇順に選択する課題(図1)を行った
- ・本課題における作業効率(時間あたりの正答数)と定義する
- ・すべての条件間で作業効率に有意差はなかった(図2)
 - ◇ 作業ペースに合ったBPMは作業効率を向上させない
- ・作業時間間隔の分散と作業効率に負の相関がある(図3)
 - ◇ 背景音によって作業時間間隔の分散を小さくできれば作業効率を向上させられる可能性がある



図1 課題出題の様子

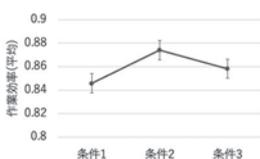


図2 条件ごとの作業効率の平均

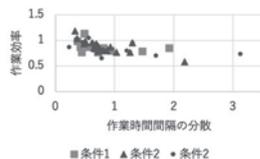


図3 作業時間間隔と作業効率の関係

3-P-5

3-P-5 高周波域における聴覚フィルタの推定

Estimation of an auditory filter in the high-frequency range

☆出山嘉俊, 大塚 翔, 中川誠司(千葉大)

- ◆ヒトの聴覚が超音波にもある程度の感度を持つことが示されているものの、未だに多くの不明点が残る。例えば、14 kHz以上の高域における聴覚フィルタに係る報告は極めて限定的で、その特性の理解は曖昧である。本研究では、Flecher(1940)の方法を用いて、16 kHzを中心周波数とする聴覚フィルタ形状の推定を試みた。
- ◆マスクには16 kHzの気導トーンバースト、マスクには16 kHzを中心としホワイトバンドノイズ(帯域幅: 300, 600, 1200, 2400, 3600, 4800, 6000, 7200 Hz)を用いた。ホワイトノイズの強度は16 kHzにおけるスペクトルレベルが10 dB SLとなるように設定した。
- ◆ノイズの帯域幅が300 Hzから2400 Hzへと増大するに伴い、平均マスクング量は14 dBから21 dBへと増加した。しかしながら、ノイズの帯域幅が3600 Hzを超えると、マスクング量の増加傾向は鈍化し、飽和する傾向が観察された。
- ◆聴覚フィルタの臨界帯域幅が平均して2400-3600 Hzの間に存在することが示唆されるが、これはMoore(2003)の等価矩形帯域幅のモデル式から算出される1752 Hzよりも大きい値となった。

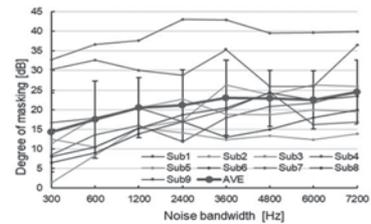


Fig.1 Degree of masking for each band width masker at 16 kHz

3-P-7

3-P-7 咀嚼音特性と食感表現との対応に関する基礎的検討: 擬音語の清濁に寄与する特徴量

A preliminary study on the relation between chewing sound characteristics and texture expressions: Acoustic features contributing to the voicing contrast in onomatopoeias

☆種石まりあ, 大塚 翔, 中川誠司(千葉大)

- ◆嚥下障害者にはとろみをつけた“嚥下食”が提供されるが、嚥下食は食感に劣り、食事の満足度の低下が懸念される。そのため、咀嚼音を疑似的に提示することで食感を改善する試みがなされている。
- ◆臨場感のある咀嚼音の再現には、知覚される食感と咀嚼音の音響特徴量との対応関係の検討が不可欠である。前報では、食感の微細な描写に適するオノマトペを用いて主観評価を行い、咀嚼音の心理音響指標との関係を調べたが、有意な相関傾向やオノマトペ間の明瞭な差異は認められなかった。
- ◆本研究では、清音/濁音の対をなすオノマトペを清濁の知覚に影響を与える咀嚼音の音響特徴量の推定を試みた。
- ◆線形混合モデルによる解析の結果、外耳道内音圧(ECSP)では全てのオノマトペ対において、また咬筋の振動加速度では一部のオノマトペ対において、RMSピークが有意な効果を示した(Fig.1)。
- ◆得られた結果は、清濁の知覚においては振幅が重要な要因であることを示す。また、咀嚼音の聴取に加え、咬筋の振動知覚もオノマトペの清濁の判断に寄与する可能性も示唆された。

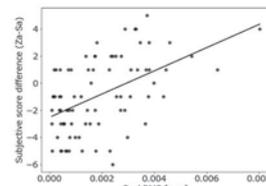


Fig.1 Scatter plot of ΔZa-Sa vs Peak RMS (ECSP). The line represents the approximate regression line based on the fixed effects from the linear model.

3-P-8

3-P-8 音像距離知覚要因統制下において 聴覚空間注意が 音像距離弁別限に及ぼす影響

The effect of auditory spatial attention on auditory distance discrimination thresholds measured under controlled conditions of auditory distance perception cues
☆加藤智世, 坂本修一(東北大通研/院情科研)

- ◆本研究では、音圧と両耳間時間差およびレベル差、輻輳角を統制した条件下で、聴覚空間注意が音像距離の弁別限に及ぼす影響を検討した。
- ◆聴取者を起点として音源を直線状に配置し、弁別刺激に先立って提示する先行音の位置を制御することで特定の位置に注意を向ける条件と注意を向けにくい条件を設定し、丁度可差異を指標として注意効果を分析した。
- ◆すべての要因が利用できた条件(Natural 条件)と音圧のみが利用できた条件(Natural-70cm 条件)の結果を比較すると、140 cm では両条件ともに効果が確認されたが、35 cm および 70 cm では Natural 条件のみに効果が確認された。このことから、身体近傍空間にて距離に注意を向けることで音像距離に対する感度が向上する可能性が考えられる。

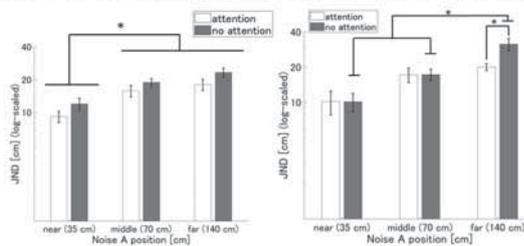


Fig. 1: JND (Left: Natural Condition, Right: Natural-70cm Condition)

3-P-10

3-P-10 加齢による認知・知覚機能への影響の評価

Assessment of the impact of aging on cognitive and perceptual functions
☆若根正太郎, 大塚 翔, 中川誠司(千葉大)

- ◆競合音存在下での聴取能力は、中年期以降に低下することが知られており、その要因として複数の認知・知覚機能の関与が指摘されている。
- ◆しかし、これらの機能を同一被験者に対して網羅的に評価し、加齢の影響を機能横断的に比較した研究は限られている。
- ◆本研究では、若年者および中高年者を対象として、競合音存在下での聴取能力、脳幹が司る時間情報処理、大脳皮質が司るワーキングメモリ (WM)、及び、音韻修復機能を同一被験者に対して計測し、加齢による影響を比較した。
- ◆競合音存在下での聴取では、聴取が困難な状況(刺激音と競合音の音響的差異が小さい、あるいは、雑音のレベルが高い条件)での聴取において加齢による影響がより顕著に現れた (Fig. 1 左)。
- ◆時間情報処理能力(両耳間時間差弁別限)および WM 容量(リスニングスパンテストスコア)は加齢により顕著に低下した (Fig. 1 中央, 右)。それゆえ、脳幹から大脳皮質に至る広範な神経ネットワークの機能が加齢により低下し、競合音存在下での聴取能力低下に寄与していると推察される。一方、音韻修復機能には加齢による影響が認められなかったことから、加齢に対する感受性が機能によって異なることも示唆された。

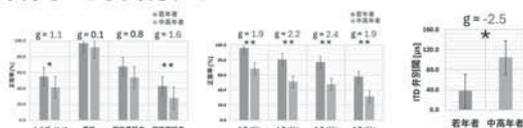


Fig. 1 各実験課題の結果 (* p < 0.05, ** p < 0.01)
左: 競合音存在下での聴取課題 (競合音条件別), 中央: リスニングスパンテストの正答率 (スパン条件別), 右: 両耳間時間差 (ITD) 弁別限

3-P-9

3-P-9 強度アクセントと旋律アクセントの顕著性の比較

Comparison of the saliency between intensity accents and melodic accents
☆後藤光, 松井淑恵(豊橋技科大院)

- ◆音楽におけるアクセントは、拍知覚を左右する重要な要因である。
- ◆異なる種類のアクセントの拍知覚への寄与を比較する場合、アクセントの知覚的な顕著性を揃えた上で実験を行う必要がある。
- ◆旋律アクセント(跳躍アクセント)と強度アクセントを2音ごとおよび3音ごとに配置した際に、知覚的な顕著性が等価となるアクセント量を明らかにするための実験を実施した。
- ◆実験の結果から、等価な顕著性を持つ跳躍アクセント量と強度アクセント量が明らかになった。
- ◆一方、2音ごと、3音ごとのアクセント間で、等価な顕著性を持つアクセント量を推定するには個人差が大きく、統制の必要がある。

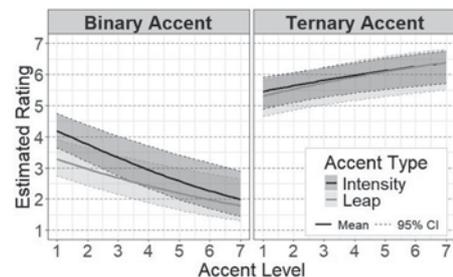


Fig. 1: Relationship between the amount of accent and perceptual evaluation estimated by the model. The left panel shows accents every two notes, and the right panel shows accents every three notes.

3-P-13

3-P-13 高齢者における音声弁別課題と時間分解能・時間微細構造処理の相関

Correlation between Speech Discrimination Tasks and Temporal Resolution / Temporal fine structure Processing in elderly listeners
☆森渡斗(豊橋技科大院), 河原英紀(和歌山大), 松井淑恵(豊橋技科大院)

- ◆高齢者は、若年健聴者、高齢者の純音聴力を模擬した模擬難聴者よりも音声弁別閾値が上昇することが明らかになっている。この弁別閾値の上昇に、純音聴力閾値より高次の処理能力が与える影響を調査した。高齢者および若年健聴者を対象に時間微細構造処理 (TFS1)・時間分解能 (Gap 検出閾値, 2点 TMTF)の測定を実施した。
- ◆時間微細構造感度の測定の結果、若年健聴者 (NH) 群は全員が課題を完了したのに対し、高齢者 (EH) 群は7名が測定不能であった。測定可能であった2名も閾値はNH群より大きい値を示した。
- ◆ギャップ検出閾値の測定、時間変調伝達関数の測定の結果、両方において、EH群はNH群よりも低い成績を示した。
- ◆模擬難聴者は高齢者に見られた時間微細構造処理、時間分解能の低下を再現していない。そのため、先行研究で見られた音声弁別閾値の乖離はこの機能低下が要因の一つであると考えられる。

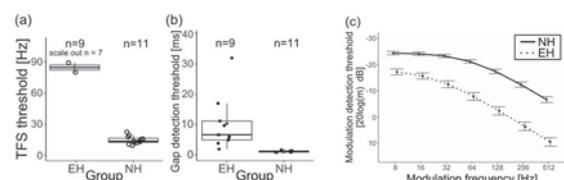


Fig. 1: (a) TFS thresholds. (b) Gap detection thresholds. Both plots show thresholds on the y-axis and listener group (EH: elderly, NH: young normal-hearing listeners) on the x-axis. (c) Temporal modulation detection thresholds. Solid lines shows NH group, dashed lines shows EH group.

3-P-14

3-P-14 強大騒音下における耳栓装用時の骨伝導音声の明瞭度: 骨伝導伝搬成分の推定に基づく音声強調の影響

Bone-conducted speech intelligibility in extreme noise with earplugging: effects of speech enhancements based on estimation of bone-conducted propagation components

☆内田知佐, 大塚 翔, 中川誠司 (千葉大)

- ◆ 骨伝導音には様々な伝搬成分が存在し、一部の成分が内耳に直接到達するため、耳栓装用時でも容易に知覚できる。
◆ 以前、強大騒音下・耳栓装用時の単音節明瞭度を調べ、強大騒音下では、耳栓しても明瞭度が50%程度まで低下したことを報告した。さらに、強大騒音下・耳栓装用時における骨伝導成分を推定するため、生体音響計測を行い、高周波成分を多く含むノイズほど音声情報が消失していることを確認した。
◆ 本研究では、強大騒音下における明瞭度を改善するため、生体音響計測を基に日本語の調音様式別、ノイズ種類毎に音声強調フィルタを作成し、単音節明瞭度試験を行った。強調フィルタの作成方法の詳細は講演原稿を参考にされた。
◆ 明瞭度試験の結果は、マルチトーカー> ホワイト> ピンクノイズとなった。明瞭度の低下は、60%程度にとどまり、特にホワイトノイズ70,80dB、ピンクノイズ80dB下では音声強調の良好な効果が観察された。

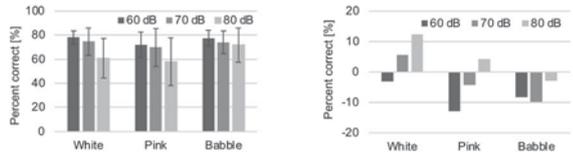


Fig.1 Percent correct of monosyllabic speech intelligibility test using enhanced monosyllables based on bioacoustic measurements under white-, pink-, and babble noise conditions. Fig.2 Differences in percent correct between original and enhanced speech under white-, pink-, and babble noise conditions.

3-P-16

3-P-16 水中における骨伝導イヤホンの検出閾測定: 外耳道深部への注水の影響

Estimation of hearing threshold using a bone-conduction earphone underwater (Part 2): Effect of water irrigation of the deep portion of the external auditory canal

○大西 廉, 大塚 翔, 中川誠司(千葉大)

- ◆ 骨伝導イヤホンには水中でも利用可能という長所があるが、水中での骨伝導イヤホンの聞こえを調べた例は極めて限定的である。
◆ 前報では水中において骨伝導イヤホンの検出閾を計測した。水中では振動子からの気導放射成分の遮断と外耳道に侵入した水による耳栓効果が生じ、高域での閾値上昇および低周波での検出閾低下が発生することを示した。しかしながら、外耳道内への水の侵入と知覚特性の関係には不明瞭な点も残った。本研究では、外耳道深部への注水を行った状態で、検出閾および単音節明瞭度を計測した。
◆ Fig. 1 に耳栓装用下における外耳道深部への注水の有無の比較を示す。外耳道深部への注水時には、耳栓装用時にも増して強力な耳栓効果および気導経路の遮断効果が発生し、低域での検出閾の低下および高域では閾値上昇が生じることがわかる。
◆ Fig.2 に単音節明瞭度試験の正答率を示す。注水時は他条件に比してすべてのカテゴリの子音の明瞭度が向上した。耳栓効果による低域のラウドネス増大に加え、外耳道に水が充満することで外耳道内放射成分が減少、慣性/圧縮骨伝導成分が増加し、聞こえが変化した可能性も考えられる。

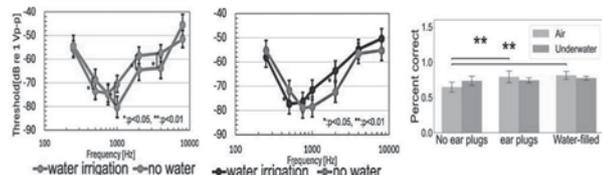


Fig.1: Thresholds in the air (left) and underwater (right) with/without water portion in the deep ear auditory canal. Fig.2: Percent correct for each condition.

3-P-15

3-P-15 一次高域・子音強調処理による骨導提示音声の了解度改善

Improving word intelligibility of bone-conducted speech by using 1st-order higher-frequency emphasis and consonant emphasis

☆中川怜奈(北陸先端大), 磯山拓都(産技高専), 上江洲安史, 鶴木祐史(北陸先端大)

- ◆ 課題: 高騒音環境下における骨導提示音声の了解度改善。
◆ 目的: 一次高域強調処理 (RT-FOE) と子音強調処理 (CE-IMP) を組み合わせた強調法 (RT-FOE+CE-IMP) が、騒音環境下における骨導提示音声の単語了解度改善に寄与するかを検討すること。
◆ 方法: RT-FOE, CE-IMP, RT-FOE+CE-IMP, および強調処理なしの4条件に対し、20名を対象とした親密度別単語了解度試験を行った。55 dB および 75 dB のピンク雑音下の条件で比較評価した。
◆ 結果: 高騒音条件 (75 dB) において、すべての強調処理条件において、強調処理なしよりも了解度が改善した。また、RT-FOE+CE-IMP は RT-FOE よりも有意に高い単語正答率を示し、CE-IMP に対しても単語正答率が高い傾向を示した。
◆ まとめ: RT-FOE と CE-IMP 単独で処理を行うよりも、組み合わせることで、単独での強調処理より了解度改善が見込めることがわかった。

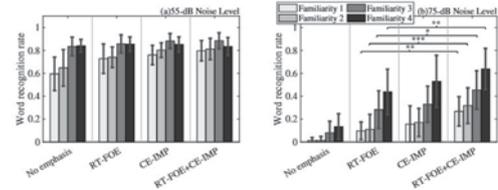


Fig. 1: Mean word recognition rate derived in word intelligibility tests for each filter type and each word familiarity rank under 55 dB and 75 dB noise level

3-P-17

3-P-17 骨導ヘッドフォンにおける音源知覚方向のズレに対する音圧依存性の調査

Perceived Sound Direction Discrepancy Depending on Sound Pressure Level in Bone Conduction Headphones

☆佐藤理生, 岩城護(新潟大院・自然研)

- ◆ 骨伝導ヘッドフォンから提示された音は、気導ヘッドフォンによる受聴に比べて音源方向が正面寄りに知覚されることが知られている。先行研究では、このズレが周波数特性に依存することが報告されているが、音圧条件の影響については十分に検討されていなかった。
◆ 本研究では、骨伝導ヘッドフォンにおける音源知覚方向のズレに対する音圧依存性を明らかにすることを目的とした。
◆ 周波数および音圧を条件とした聴取実験を行い、得られたズレ量に対して重回帰分析を適用した。
◆ その結果、周波数の主効果および基準方向と周波数の交互作用が有意であり、高い周波数ほど音源方向が正面寄りに知覚される傾向が確認された。一方で、音圧の主効果および音圧を含む交互作用はいずれも有意ではなかった。

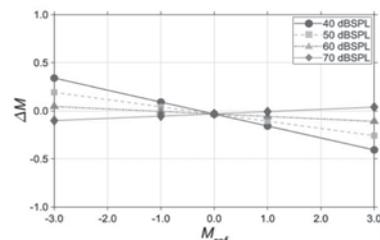


Fig. 1: Perceptual discrepancy (ΔM) versus presented sound source direction (M_ref) for different sound pressure levels.

3-P-18

骨伝導イヤホン装着時における周囲環境音の聴取特性:

検出閾および語音明瞭度の評価

Hearing characteristics of surround sounds when wearing bone-conduction earphones: Evaluation of threshold and monosyllabic speech intelligibility

☆薛 芸佳, 内田知佐, IRWANSYAH, 大塚 翔, 中川誠司(千葉大)

- ◆骨伝導イヤホンは外耳道を塞がないことから周囲環境音を聞き取りやすいとされているが、骨伝導イヤホン装着時の周囲環境音の聞こえを定量的に調査した例は極めて少ない。
- ◆骨伝導 (BC)、軟骨伝導(CC)、および気導 (AC) イヤホンでノイズを呈示した際の周囲音の聴取特性 (トーンバーストの検出閾と単音節明瞭度) を推定し、イヤホン間の比較を試みた。
- ◆無響室内で被験者の右耳に3種のイヤホンでマスキングノイズを、被験者右側の 1.0 m 離れた位置に設置したスピーカーからトーンバーストおよび単音節刺激を呈示した。
- ◆低周波域では、骨伝導イヤホンに対してより小さな検出閾および高い明瞭度が得られる傾向が認められた。一方、高周波域における骨伝導イヤホンの検出閾は気導イヤホンと同等であった。

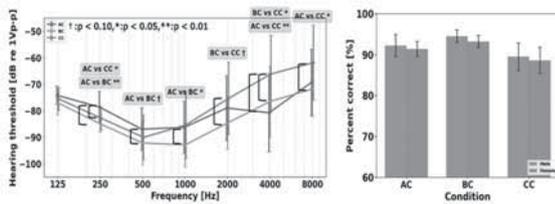


Fig.1: Hearing thresholds (left) and monosyllabic speech intelligibility (right) for AC, BC, and CC earphones.

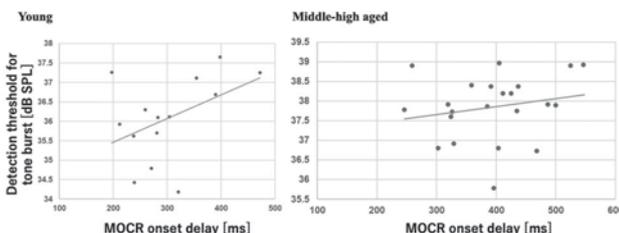
3-P-20

3-P-20 加齢による雑音下での聴取能力低下とオリーブ蝸牛束反射の時間応答特性の関係

Relationship between Temporal Characteristics of Medial Olivocochlear Reflex and Age-Related Declines in Listening-in-Noise Performance

☆堂脇悠里, 大塚翔, 中川誠司(千葉大)

- ◆脳幹から外有毛細胞 (OHC) へ接続するオリーブ蝸牛束には、音響刺激に反応し、OHC の増幅を抑制する働きがある (オリーブ蝸牛束反射: MOCR)。MOCR は、OHC のゲインを調節することで、雑音下での聴取を改善する役割 (アンチマスキング効果) を果たすと考えられている。著者らはこれまでに、聴覚的に注意に伴う MOCR の増強が、アンチマスキング機能に与える影響を検討してきた。
- ◆前報では、若年者群に比べて中高年者群は MOCR が弱いこと、および、中高年者群では聴覚的に注意による MOCR の増強効果が消失することを示した。MOCR 強度の低下が、加齢に伴う雑音下での聴取能力低下の一因であることが示唆された。本研究では、中高年者を対象として、聴覚観察中の MOCR の時間応答特性を測定し、雑音下での聴取成績との関係を調べた。
- ◆MOCR の立ち上がり時間は、若年者群に比べて中高年者群の方が有意に長かった。若年者群においては、MOCR の立ち上がり時間が短いほど、雑音下での聴取成績が高いことが確認された (Fig. 1 left)。一方で、中高年者群においては、両者の間に統計的に有意な相関は認められなかった (Fig. 1 right)。
- ◆同一被験者群を対象とした前報では、中高年者群においても、聴覚観察中の MOCR 強度と聴取成績との間に有意傾向の相関が見られた。これを踏まえ、加齢に伴う立ち上がりの遅延と比較して強度の低下の方が、雑音下での聴取能力低下への寄与が大きいことが推察された。



◆ Fig.1 Relationship between tone-in-noise detection performance and MOCR onset delay for young and middle-aged adults

3-P-19

3-P-19 咀嚼時に聴取される気導成分および骨伝導成分の推定(第2報):咀嚼音の構成成分の推定の試み

Estimation of air-conducted and bone-conducted components perceived during chewing (2nd report): Estimation of the constituent components of chewing sounds

☆至田拓史, 大塚 翔, 中川誠司(千葉大)

- ◆前報では、咀嚼時に摂食者自身に知覚される気導成分と骨伝導成分を推定し、両成分のスペクトルに有意な違いが存在することを報告した。
- ◆一方、咀嚼音には、食品の破碎、食品をすり潰す際の上下の歯の摩擦、歯同士の接触、下顎の振動、咬筋の収縮、舌の運動などの様々な発生要因が存在する。
- ◆発生要因ごとの咀嚼音成分の同定を目的として、被験者が口腔内に食品を入れずに咀嚼様運動 (空咀嚼) および 5 種の菓子の咀嚼を行っている際に、両耳の外耳孔および外耳道内に生じる音圧、および乳様突起に生じる振動の計測を行った。
- ◆外耳孔及び外耳道内音圧は特に、1,000-4,000 Hz で空咀嚼条件との差異が大きくなった (Fig.1-2)。低域にはクランチー (ザクザク/パリパリした) な食品の破碎音およびすり潰し音が、高域はクリスピー (サクサク/パリパリした) な食品の破碎音が反映されると思われる。
- ◆乳様突起振動のパワーは条件によらず低域 (~250 Hz) で大きかった。また、全帯域において咀嚼条件で増大したが、低域では条件間の差異が減少した (Fig.3)。乳様突起振動には食品の破碎やすり潰しに伴う成分よりも、下顎運動に起因する成分が強く反映されると思われる。

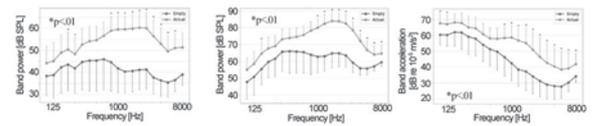


Fig. 1 Power spectral densities for the chewing and the empty-chewing conditions: Left: the sound pressure at the opening of the ear canal, Center: the ear-canal sound pressure, Right: the vibration at the mastoid for the chewing and the empty-chewing conditions.

3-P-21

3-P-21 DPOAE Suppression in Humans Under Bone-Conducted Ultrasonic Masking (26-42 kHz)

IRWANSYAH, Naoya TAKAHASHI, Sho OTSUKA, and Seiji NAKAGAWA (Chiba University)

- ◆Ultrasound is generally considered inaudible to humans, yet via bone conduction, it can somehow be "heard" (up to 100 kHz).
- ◆This suggests that ultrasonic vibration may reach the inner ear. But, which cochlear region actually responds to ultrasound?
- ◆In this study, we identify the cochlear frequency region influenced by bone-conducted ultrasonic masking (26–42 kHz) by observing distortion-product otoacoustic emissions (DPOAEs).
- ◆Our results indicate that bone-conducted ultrasound suppresses DPOAEs mainly in the basal, high-frequency cochlear region.

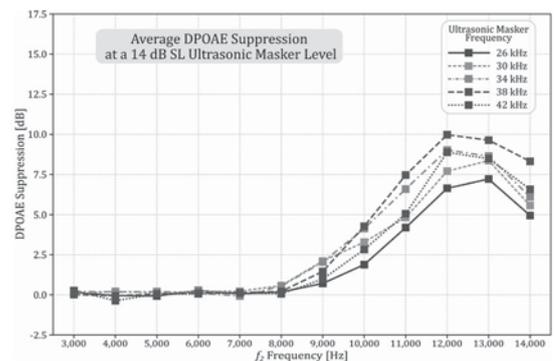


Fig. 1 Average DPOAE suppression across f_2 for five ultrasonic maskers.

3-P-22

3-P-22 復帰抑制下における耳音響放射と事象関連電位の関係

Relationship between otoacoustic emissions and event-related potential under inhibition of return

☆松毛円佳, 大塚翔, 中川誠司(千葉大)

- ◆ 左右いずれかの耳に外因的に注意を誘導すると、その効果は時間経過とともに弱まり、対側耳へ注意が向けられるようになる(復帰抑制, Inhibition of return: IOR)。
- ◆ IOR 下では、非注意耳において OAE が増大すること、及び、その増加量と後頭部の α 波の強度との間に相関が見られることを前稿で報告した。IOR 下の聴覚末梢の制御に皮質領域が関与することが示唆されたものの、皮質におけるどの処理段階が関与しているかは未解明であった。本研究では、OAE と同時に事象関連電位 (Event Related Potential: ERP) を観察し、各潜時における ERP と OAE の関係を調べることで、聴覚末梢の制御に関わる皮質領域の特定を試みた。
- ◆ 注意耳に刺激音が提示されると ERP に陽性の電位変化が生じ (Fig.1), その電位変化が大きいほど注意耳の DPOAE が小さくなった (Fig. 2 中で負の相関)。両者の相関は潜時の増加と共に減少したことから、IOR 下では聴覚皮質周辺の局所的な領域が聴覚末梢の制御に関与する可能性が示された。一方、IOR に先行する外因的注意下では、両者の相関は長潜時においても観察されたことから (Fig. 2), IOR とは異なり、高次の連合野に至る広範な皮質ネットワークが聴覚末梢の制御に関与することが示唆された。

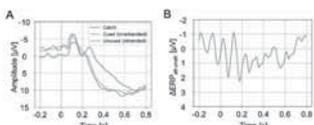


Fig.1 (A) A typical example of grand average ERP recorded during IOR. (B) Difference waveforms obtained by subtracting the ERP wave of the non-attended from the attended condition ($\Delta ERP_{att-unatt}$).

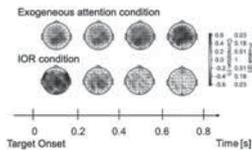


Fig.2 Inter-individual correlations between $\Delta ERP_{att-unatt}$ and $\Delta DPOAE_{att-unatt}$ when a target was presented to the left ear.

3-P-24

3-P-24 経頭蓋直流刺激がオリーブ蝸牛束反射および雑音下での聴取に及ぼす影響

Effects of transcranial direct current stimulation on the medial olivocochlear reflex and listening in noise

☆葦重温輝, 大塚翔, 中川誠司(千葉大)

オリーブ蝸牛束反射 (MOCR)は、外有毛細胞の増幅機能を抑制し、雑音下での信号検出を促進する働き(アンチマスキング効果)をする。また、大脳皮質からの遠心性神経を經由して MOCR が変調されることを支持する知見も蓄積されつつある。本研究では、経頭蓋直流刺激 (tDCS)を用いて、MOCR の制御への関与が示唆されている聴覚野の興奮性を非侵襲的に変調し、MOCR に及ぼす影響を調べた。

1.0mA の直流陽極刺激を 10 分呈示後、一時的に MOCR が増大し、雑音下での聴取成績も向上した。また、tDCS による MOCR の増強が大きい被験者ほど聴取成績の改善が大きいことが示された (Fig.1)。一方、立ち上がりの方に電流が流れる擬似刺激では両者に変化は見られなかった。これらの結果から、tDCS によって、MOCR のアンチマスキング効果を高められることが示唆された。効果を持続する手法を確立することで、聴覚情報処理障害等で見られる典型的な症状である雑音下での聴取困難のリハビリテーションへの応用が期待できる。

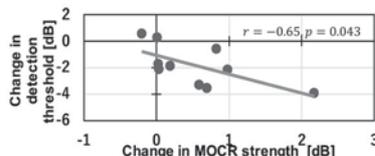


Fig.1 Inter-individual correlation between the change in MOCR strength and tone-in-noise detection threshold after anodal tDCS.

3-P-23

3-P-23 脳波による競合音存在下でのリスニングエフォートの評価

Evaluation of Listening Effort under Competing Noise Conditions Using Electroencephalography

☆福島圭佑, 大塚翔, 中川誠司(千葉大)

- ◆ 電車内やレストランなどの競合音環境において、純音聴力検査では正常と診断されるにもかかわらず、音声聴取に困難を示す症例が報告されている。この症例は隠れた難聴と呼ばれ、その診断手法はいまだ確立されていない。
- ◆ 聴取困難に伴うリスニングエフォート(聴取するために割く注意資源の量)の増大は、日常生活でのストレスや疲労を増加させる要因として認知されつつある。リスニングエフォートを客観的に推定することができれば、従来の聴力検査では捉えにくい、隠れた難聴に伴う心的な負担を評価できるようになる可能性がある。
- ◆ 本研究では、競合音下での音声聴取課題と記憶課題を同時に課す二重課題法を用い、リスニングエフォートを行動学的に評価するとともに、脳波指標との対応関係を調べた。
- ◆ 統計的な有意性は確認されなかったものの、記憶課題の成績が下降、つまり、聴取課題に割かれる注意資源が増加すると、 δ 帯域の振幅が減少する傾向が見て取れた。今後はリスニングエフォートの評価指標としての有効性を統計的に精査したい。

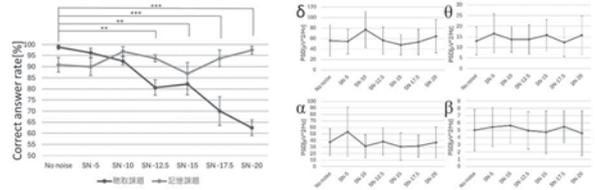


Fig.1 聴取課題、記憶課題の正答率(左)、各周波数帯域の後頭部電極のパワースペクトル密度(右)。**p<0.01, ***p<0.001

3-P-24

3-P-25

3-p-25 脳波の律動成分の時間安定度を用いた睡眠評価の試み

Sleep assessments using the temporal stability of EEG oscillations

☆入交陽斗, 大塚翔, 中川誠司(千葉大)

- ◆ 聴覚刺激による睡眠促進効果に関する研究は以前から行われているが、先行研究ではアンケートを用いた主観指標と脳波等の客観指標の乖離が存在するなど、既存の睡眠評価指標には改善の余地がある。新たに脳波の“時間安定度”という指標を導入し、睡眠の評価を行った。
- ◆ 以下の 4 条件の楽曲の呈示下で睡眠時脳波および眼電図の計測を行った。(1) 好きでも嫌いでもない (Neutral), (2) 少し好ましい (Low), (3) 中程度好ましい (Middle), および (4) 十分に好ましい (High)。
- ◆ 覚醒後の被験者に対し、OSA 睡眠調査票簡易 MA 版 (OSA-MA) による睡眠内省の評価を行った。また、計測された脳波の各帯域成分の自己相関関数から時間安定度を算出した。
- ◆ 楽曲の好ましさに伴って OSA-MA のスコアが有意に向上した。一方、好ましさに伴う時間安定度の有意な変化は見られなかったものの、 δ 帯域、 θ 帯域では睡眠ステージの進行に伴って時間安定度が有意に増大した。
- ◆ 律動成分のパワーでは区別が難しい睡眠ステージ 1 と 2 であっても、時間安定度は有意に変化した。時間安定度は睡眠の深さを正確に捉える新たな指標になり得る。

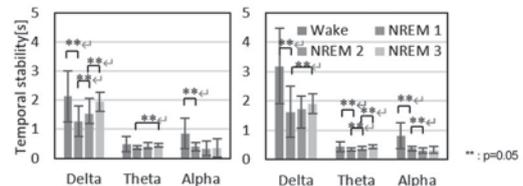
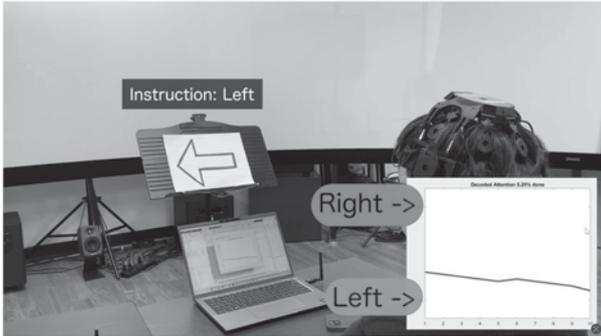


Fig.1: Temporal stability for each condition and each frequency band. **: p=0.05

3-P-26

3-P-26 Dilated CNN model for real-time spatial auditory attention decoding from a dry-electrode EEG device

☆竹内朗 (RIT, KAIST), △Hwan Shim (RIT), △Inyong Choi (Ulowa), Sungyoung Kim (RIT, KAIST)



Auditory attention decoders (AADs) allow the decoding of listeners' spatial attention from EEG signals under naturalistic listening conditions. This study aims to develop a concise, data-efficient AAD framework suitable for dry-electrode EEG systems with limited channels. Using EEG data from 16 participants in the KULeuven dataset, training was restricted to 18 channels to emulate a DSI-24 setup. A modified CNN-loc architecture with dilated convolutions slightly improved static accuracy; however, the original model provided more stable real-time attention tracking, highlighting a trade-off between offline accuracy and real-time robustness.

3-Q-2

3-Q-2 生成モデルに基づくゼロショット音響イベント検出における分布外検出モデルを用いたクラス間バイアス補正

Bias Correction between Seen and Unseen Classes Using an Out-of-Distribution Detection Model in Generative-Model-Based Zero-Shot Acoustic Event Classification

☆上原康平(神戸大), 高島遼一(立命館大), 滝口哲也(神戸大)

- ◆未知(未学習)の音響イベントクラスを、クラスの説明情報(属性情報)を元に分類・識別するゼロショット学習の手法を提案。識別対象の学習データが収集できない異常検知等への応用が期待される。
- ◆ゼロショット学習では、未知クラスのイベントが既知(学習済み)クラスに誤分類されやすい問題がある。
- ◆誤分類を減らすため、分布外検出モデルを用いて既知クラスか未知クラスかをスコア化する。従来手法では、このスコアを閾値で0/1に二値化していたが、提案手法では0-1の実数値として、よりソフトな表現に変換し、クラス分類器の出力と統合。
- ◆Fig. 1 に示す Seen-Unseen accuracy 曲線(下部の面積が大きいほど良い)より、提案手法は従来手法より高い性能を示した。

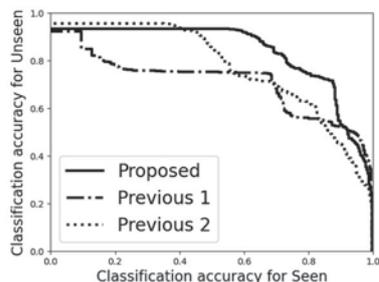


Fig.1: Seen-Unseen accuracy Curve (SUC) for each method.

3-Q-1

3-Q-1 教師なし異常音検出における RAG を活用した差分キャプション生成

Retrieval-augmented difference captioning to explain unsupervised anomalous sound detection

◎小倉稜也, 西田智哉, 川口洋平(日立製作所)

- ◆機械設備の教師なし異常音検出において、異常音が正常音と比較してどのように異なるかを説明するキャプション(差分キャプション)を生成する手法を提案
- ◆検索拡張生成(Retrieval-Augmented Generation: RAG)に基づき、異常音と類似している正常音を学習データから検索し、異常音と正常音のキャプションを比較
 - ⇒ 正常音と異常音の差異を捉えた差分キャプションを生成
- ◆提案手法によって生成した差分キャプションが、異常の要因を記述できていることを確認

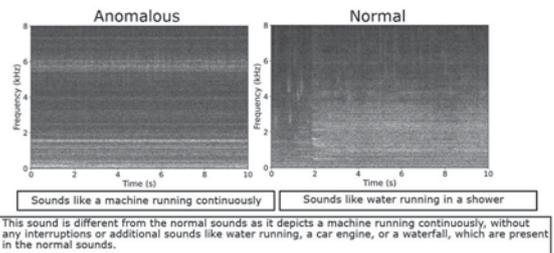


Fig.1: Spectrograms of anomalous and reference normal pump sounds and corresponding captions. The boxes in the first row denote individual captions for each sample. The box in the second row denotes the difference-caption.

3-Q-3

3-Q-3 環境音と説明文の主観的意味関連性 データセットにおける「音響 or 言語」的分析

Acoustic or linguistic analysis of environmental sounds and captions in subjective semantic alignment

◎岡本 悠希(東大院), 砺波 紀之(NEC)

- 環境音と説明文の主観的意味関連性の分析を実施する。
- 実験では、音響と言語的なもつれを解きながら意味関連性を予測する。
 - 結果より、音響特徴のみの入力からでも比較的高い意味関連性予測精度を達成することがわかった。
 - 心理音響学的な分析により、音圧や突発性雑音との意味関連性の相関も確認された。

音の可能性に気が付く



音とキャプションを独立的に分析した時、

— 観的意味関連性データセットの分析について —

3-Q-4

3-Q-4 Flow matching を用いた多様なオーディオキャプションニングの検討

Diverse Audio Captioning via Flow Matching

◎藤田直明, 中村拓紀, 板倉光佑(パナソニック)

- ◆オーディオキャプションニングは音環境を自然言語で記述するタスクであり重要な指標として正確性と多様性がある。正確性と多様性の両立する手法として拡散モデルを用いた手法が提案されているが、デノイズングによる高計算量が課題である。
- ◆正確性と多様性の性能を維持しながら計算量を低減するため Flow Matching を用いた手法を提案。
- ◆拡散モデルと同等のアーキテクチャを用いて計算量の変化による正確性、多様性への影響を比較した。実験の結果、提案法は拡散モデルに比べて高い性能を示し、計算量を下げても性能劣化が小さいことが確認された。

Method	Accuracy					Diversity			
	BLEU4	ROUGE	METEOR	CIDEr	SPIDEr	Vocab	mB4	div-1	div-2
Baseline									
250 step	0.141	0.364	0.171	0.375	0.247	2374	0.120	0.578	0.787
30 step	0.139	0.363	0.169	0.376	0.248	2782	0.112	0.581	0.794
10 step	0.135	0.363	0.168	0.367	0.253	3543	0.096	0.597	0.810
Proposed									
250 step	0.153	0.379	0.176	0.402	0.263	2256	0.200	0.507	0.721
30 step	0.154	0.380	0.176	0.403	0.264	2257	0.189	0.507	0.726
10 step	0.152	0.377	0.173	0.393	0.257	2453	0.188	0.506	0.725

図1: 従来法と提案手法の正確性と多様性の客観的評価指標

3-Q-6

異常スコア分散最小化に基づくバイアス調整を用いたドメイン汎化型識別的異常音検知

Adjusting Bias in Anomaly Scores via Variance Minimization for Domain-Generalized Discriminative Anomalous Sound Detection

☆松本昌亮, 藤村拓弥, HUANG Wen-Chin, 戸田智基(名古屋大学)

- ◆識別的異常音検知は高性能だが、ドメインシフトのため汎化性能が劣化する。
- ◆従来手法は、異常スコアを調整するバイアスを導入することでドメインシフトにアプローチしたが、各埋め込み空間に性能が依存する。
- ◆本研究はバイアスに重み付けパラメータを導入し、異常スコア分散最小化に基づくバイアス調整を用いたドメイン汎化型識別的異常音検知を提案した。
- ◆実験では、従来手法よりも高い性能を示し、重み付けパラメータは埋め込み空間に適応的に性能の高いものが選ばれることが明らかになった。

Table 1 Mean of official score [%] for DCASE2022-2024 datasets.

	DCASE2022	DCASE2023	DCASE2024
Baseline	63.79	57.00	51.59
Previous	62.68	65.32	51.47
Proposed	66.34	67.41	54.10

3-Q-5

3-Q-5 珈琲焙煎におけるハゼ音のための非負値因子行列分解を用いた音響モデリング

Acoustic Modeling of Coffee Roasting Sounds for Identifying Crack Events

◎新西 和也(広島市大院), 向井 健悟(広島市大), 藤村 秀弥, 大島 風雅, 中山 仁史(広島市大院)

- ◆入力信号は空回し時の装置音と焙煎中の観測音(装置音・焙煎音・ハゼ音を含む)を対象とする。
- ◆観測音に対してHPSSを適用し、連続成分とパルス成分に分解する。
- ◆パルス成分からハゼ候補の時間系列(周波数方向の集約・平滑化・時間フレーム整形・正規化)を行う。
- ◆装置音を既知成分としてSS-NMFを適用する。
- ◆結果より焙煎音とハゼ音で異なる特徴が基底として表現され、ハゼ音は離散的イベントとして抽出されることを確認した。

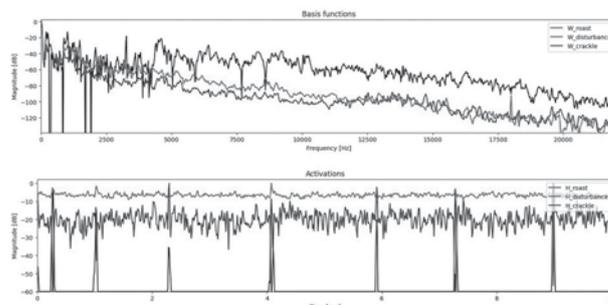


Fig.1: SS-NMF-based crack events extraction results

3-Q-7

3-Q-7 無作為連結環境音をモデル学習に用いたE2E型擬音語認識の性能評価

Evaluation of E2E Onomatopoeia Recognition using Randomly Concatenated Environmental Sounds for Model Training.

☆森山 颯太, 西村 竜一(和歌山大学)

- ◆提案したEnd-to-End型擬音語認識モデルの学習データに関する検討
 - 環境音信号を連結(反復型、無作為型)して学習データを用意
 - 反復型の環境音を含めた学習によって、不自然な繰り返しを出力
- ◆仮説:
 1. 擬音語の種類を意図的に制限、典型的なパターンの学習を強化
 2. 反復型を学習から除外することで不自然な繰り返しの出力を抑制
 3. 無作為型のデータ数を増やすことで、反復パターンを再現
- ◆4条件の学習データで擬音語認識モデルを構築、主観的に妥当率を評価(条件1, 2, 3では、使用する擬音語の種類を削減済)
 - 条件1: 学習データ数 30,000 (無作為型 100%)
 - 条件2: 学習データ数 120,000 (無作為型 100%)
 - 条件3: 学習データ数 30,000 (無作為型 80%+反復型 20%)
 - 条件4: 学習データ数 30,000 (無作為型 80%+反復型 20%)

◆実験結果(表1)

(表1) クラス別の妥当率の平均(「○(妥当)」と回答した数/回答数)

環境音のクラス	条件1	条件2	条件3	条件4
缶を開ける音	38.7	28.0	25.3	28.0
ガラスが割れる音	30.7	38.7	22.7	6.7
ドアがきしむ音	14.7	8.0	17.3	13.3
掃除機の音	28.0	33.3	33.3	16.0
掃除機の音	56.0	61.3	64.0	66.7
時計のアラーム音	57.3	46.7	56.0	72.0
タイピング音	34.7	13.3	44.0	42.7
ドアのノック音	46.7	50.7	28.0	70.7
マウスのクリック音	28.0	17.3	16.0	18.7
時計の針の音	32.0	22.7	38.7	28.0
全体平均	36.7	32.0	34.5	36.3

3-Q-8

3-Q-8 深層学習に基づく
多重音高推定と音高単位楽器推定による
自動採譜に関する検討

A study on deep learning-based automatic music transcription via multi-pitch estimation and pitch-wise instrument recognition

☆小倉真了, 坂野秀樹, 旭健作(名城大院・理工学研)

- ◆タスク: 楽器別音高推定
 - 各時刻にどの楽器がどの高さの音を演奏しているかを推定
- ◆提案手法: 楽器別音高推定を多重音高推定と, 検出された各音高に対応する楽器を推定するタスク(音高単位楽器推定)に分解(Fig. 1)
- ◆ベースライン手法: End-to-end 楽器別音高推定モデル
- ◆実験結果
 - Average precision (AP) が低下 (66.4~75.3% → 48.9~56.7%), F1スコアは同等 (61.2~69.8% → 59.9~70.1%)
 - ◇ Precision-recall 曲線の歪みに起因
 - 条件変更により AP は向上したがベースライン未満, F1 は低下
 - ◇ 無楽器ラベルの導入で AP, F1 が向上したがベースライン未満

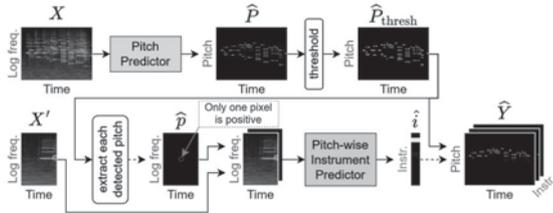


Fig. 1: Proposed framework

3-Q-10

3-Q-10 CNN を用いた音響による接近車両検出における音データの時間伸縮によるデータ拡張に関する検討

A Study on Data Augmentation via Time Warping of Acoustic Data for Approaching Vehicle Detection Using CNN

☆加藤朋暉, 堀江拓矢, 旭健作, 坂野秀樹(名城大院)

- ◆出会い頭衝突事故防止を目的に, 道路上で収録した音を用いた接近車両検出を検討している. 遠方の到来車両と通過後の車両の音が混在したデータを排除するようにデータセットを再作成したところ, データの質は向上したものの, データ数は約3分の1と大幅に減少した.
- ◆本稿では学習用データ数増加を目的としたデータ拡張手法として, リサンプリングによる音データの時間伸長の適用を検討する.
- ◆11種類の伸縮倍率*r*において, 検出精度に与える影響を調査する.
- ◆接近の検出について, すべての*r*において, 既存データのみで学習した場合と比較して精度の向上を確認した.
- ◆特に, *r* = 1.1~1.5の範囲において, 10pt以上精度が向上した.

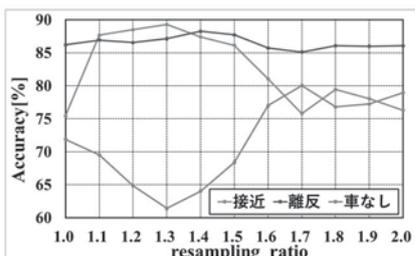


Fig. 1: Accuracy results for each resampling ratio.

3-Q-9

3-Q-9 CNN を用いた音響による接近車両検出における着目する周波数帯変更による検出精度改善

☆堀江拓矢, 加藤朋暉, 旭健作, 坂野秀樹(名城大院)

- ◆出会い頭事故の防止を目的に, 見通しの悪い交差点において収録した道路環境音を用いて接近車両検出を検討している. 従来に入力データにおける位相情報は, 609~1382 Hzの帯域を使用しており, 600 Hz以下の低周波数帯は学習に使用していなかった.
- ◆本研究では, 位相情報の周波数帯域を46~1570 Hzに拡張した入力データを使用する. 入力データが周波数方向に約2倍大きくなったため, CNNモデルの周波数方向のカーネルサイズも2倍にする.
- ◆以上の提案手法を評価した結果, 最接近時刻(0.0秒)の5.0秒前から4.0秒前までの精度が, 従来モデルよりも約15pt向上した.
- ◆Fig.1に提案手法の結果例として, 時刻毎の判定割合を示す.

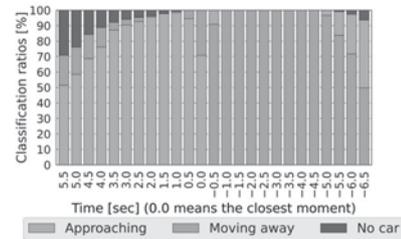


Fig. 1: Classification ratios at each time in approaching and moving away situations

3-Q-11

3-Q-11 音色特徴量に基づく機械の異常音説明

Anomalous sound explanation based on timbral features

○西田智哉, Purohit Harsh, 土肥宏太, 遠藤隆, 川口洋平(日立製作所)

- ◆教師なし異常音検知において **異常-正常差分を説明する枠組みを提案**
- ◆「**音色属性の客観指標 (Sharpness, Brightness, Boominess, Depth, Roughness) の増減を識別**」するタスクとして定式化
 - 異常音データを学習不要 / 機械音の微小差分説明に有効
- ◆近傍正常音との比較により, 正常内のばらつきを除外して異常起因差分のみを特定
- ◆MIMII-DG系データセットで **ベースラインを上回る識別精度を確認**

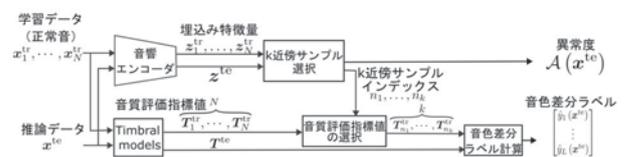


Fig. 1: Proposed timbre difference labeling method

3-Q-12

3-Q-12 自動運転中の生体信号フィードバック音による音響的マインドフルネスの効果

Effects of Acoustic Mindfulness on Drivers During Autonomous Driving

☆福田千聖, 須藤季砂, 石光俊介(広島市大・情報科学研究科)

- ◆自動運転技術の実用化が進んでいるが、未だドライバの必要性は失われておらず、事故防止に向けた取り組みが必要である。事故の大きな原因のひとつは疲労の蓄積であり、この疲労を軽減する効果的な方法を見いだすことが重要である。
- ◆本研究では、呼吸・心拍に同期したフィードバック (FB) 音を自動運転環境下のドライバに呈示することで、音響的マインドフルネスによる疲労軽減・その後のパフォーマンス向上を行うことを検討した。(Fig.1)
- ◆4種のFB音(呼吸FB:ピッチ変更C音3種・心拍FB:bpmメトロノーム1種)による影響を脳波(β/α 比)・運転パフォーマンスを用いて検証した。
- ◆結果としてFB音(C5)聴取時に運転パフォーマンス向上傾向、呼吸FB音(C5)聴取によるマインドフルネス効果が確認された。

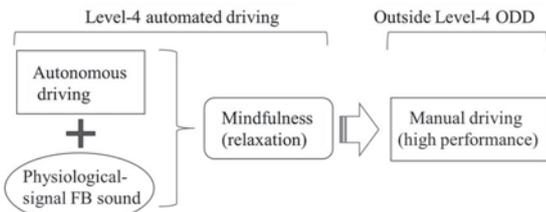


Fig.1: Conceptual flow of the proposed approach.

3-Q-14

3-Q-14 光波マイクロホンによる音圧絶対測定の精度向上

Improving the accuracy of absolute sound pressure measurements with the optical-wave microphone

○山田桂輔, 高橋弘宜, 平野琴(産総研)

- ◆光波マイクロホンは光学的音響測定法の一つであり、音圧による空気の屈折率の時間変化をレーザ断面における光強度分布の時間変化として検出する。我々は光波マイクロホンによる音圧の絶対測定の実現と高精度化を目的とした研究開発を進めている。
- ◆本研究では、測定システムの機械的振動に起因するノイズの影響の抑制法の検討を目的とし、測定システムの受音部レーザ長さの増加によるSN比の改善と測定精度の向上について評価した。
- ◆受音部長さを変化させて音圧を測定し、コンデンサマイクロホンによる測定値と比較すると、4 kHz以上ではどの受音部長さ条件による測定値もコンデンサマイクロホンによる測定値と1 dB程度の範囲で一致した。一方3.15 kHz以下ではコンデンサマイクロホンとの差が拡大したが、受音部長さ大の条件では差が小さくなることを確認した。

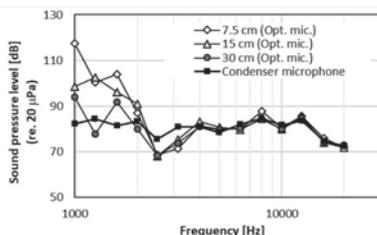


Fig.1: Sound pressure levels measured by the condenser microphone and the optical-wave microphone.

3-Q-13

3-Q-13 音楽が筋力トレーニング時の筋疲労や筋回復に与える影響の初期検討

Preliminary study on the effects of music on muscle fatigue and recovery during strength training

☆角田清香, 瀧迅, 矢田部浩平(農工大)

- ◆客観的指標の不足: 適切な負荷や休息の判断が難しく怪我のリスクや効率低下が懸念
- ◆主観評価の限界: 音楽の疲労軽減効果は知られているが多くが主観評価

音楽鑑賞・無での筋トレ過程を記録

数理モデルを用い、音楽鑑賞が筋疲労・回復に与える影響を客観的に検証



3-Q-15

3-Q-15 想起音声分類のための脳磁図データの outlier 除去手法の改良

Improvement of an Outlier Rejection Method for MEG Data in Imagined Speech Classification

☆能勢幸樹, 矢野肇, 滝口哲也(神戸大), 中川誠司(千葉大)

- ◆脳活動データは、画像や音声データと異なり、データの妥当性を確認することは困難である。そこで、次元圧縮手法によりデータを低次元空間に写像し、分布から逸脱したサンプルを外れ値として扱う方法が用いられている。
- ◆前稿では、音声想起時の脳磁図に対し t-SNE 空間に圧縮を行い、外れ値検出・除去を行った。
- ◆さらに本研究では、脳磁図に対して電流源推定を行い、得られた電流源信号に対し信号処理をすることでデータ拡張を実施した (Fig. 1)。この拡張データは、外れ値分類器の事前学習と分類器の学習に利用した。
- ◆実験の結果、外れ値除去後の学習データに拡張データを加えることで、分類精度の向上が確認された。

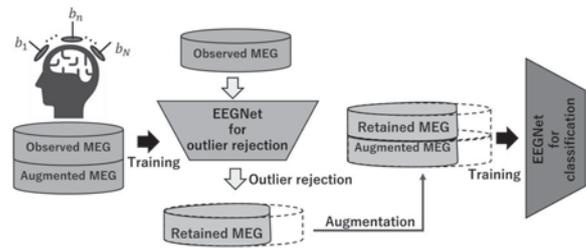


Fig. 1 Overview of the proposed method

3-Q-16

3-Q-16 少視点光学的音場復元に向けた 拡散モデルに基づく音場投影画像生成

Sound-projection-image synthesis based on latent diffusion model for few-view acousto-optic tomography

○谷川理佐子(NTT/早大理工), 石川憲治, 原田登(NTT), 及川靖広(早大理工)

- ◆ 背景: 光学的音響計測における三次元復元
 - 複数方向から計測した音場の線積分データの取得が前提
 - より少ない観測データからの復元を目指す
- ◆ 提案: 潜在拡散モデル (LDM) により任意方向のデータを合成
- ◆ 結果: 数値実験により音場データの合成可能性を確認

(正規化平均二乗誤差: -12.19 dB)

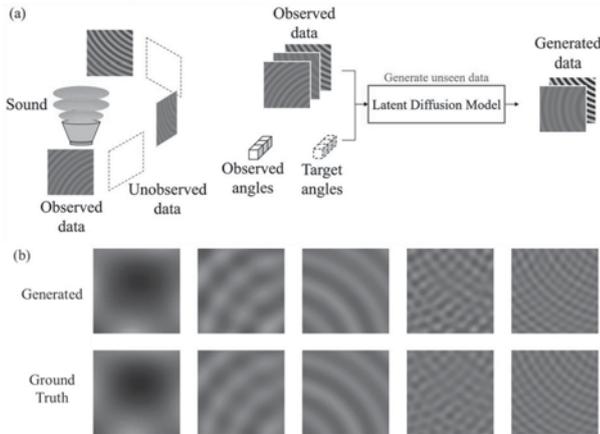


Fig. 1: Proposed method (a) Conceptual diagram of proposed approach. (b) Generated and ground truth data of five different data.

3-Q-18

3-Q-18 FDTD法における吸収境界最適化手法の評価

Evaluation of optimization methods for absorbing boundary conditions in the FDTD method

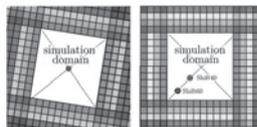
☆ 犬飼朋樹, 矢田部浩平(農工大)

前回 PMLの減衰係数最適化手法 Parametric, Fullを提案



単純な条件での最適化 & 評価

今回



PMLの配置向き, 音源位置を変更し最適化 & 評価

結果

いずれの条件においてもFullによる最適化がより有効であることが示唆された。

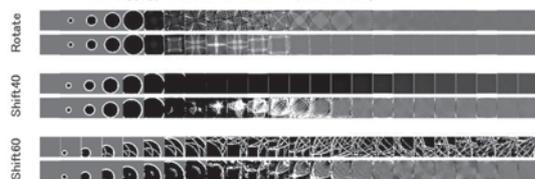


Fig. 1 Simulation results. Time progresses from left to right. In each group, the upper row shows the "Parametric" results, and the lower row shows the "Full" results.

The peak pixel value of the initial Gaussian pulse is set to 1, and the dynamic range is narrowed for better visibility. A faster transition to gray indicates greater wave absorption.

3-Q-17

3-Q-17 再生核ヒルベルト空間を用いた 分散多チャンネルアレイ音場収音法の検討

Sound field recording method using distributed multichannel microphone arrays based on kernel Hilbert spaces

☆ 柳屋啓太, 坂本修一(東北大通研/院情科研)

- ◆ 広がりをもった領域における音場収音の実現により, 再現音場内での聴取者の移動が可能となる。
- ◆ 多チャンネルマイクロホンアレイは多数のマイクロホン小さな筐体で扱うことができ, 実用性と規模のバランスに優れる。
- ◆ そこで多チャンネルアレイを分散配置し, 取り囲まれた領域の音場を収音する分散多チャンネルアレイ収音法を検討した。
- ◆ 収音系の信号から音場を再構成するにあたり, 音場を再生核ヒルベルト空間 (RKHS, reproducing kernel Hilbert space) の要素とみなし, ある有限次元部分空間への射影に基づき再構成を行った。
- ◆ シミュレーションの結果, 再構成音場 (Fig. 1 左) は原音場 (Fig. 1 右) の波面を十分に再現した。

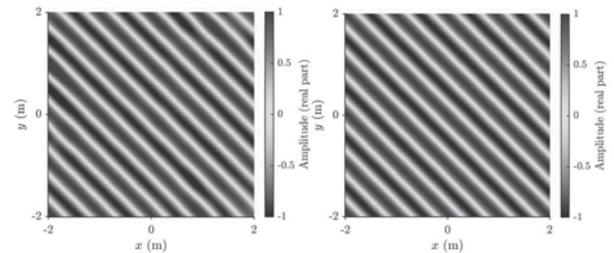


Fig. 1: Comparison of reconstructed and original sound fields (Left: Reconstructed; Right: Original)

3-Q-19

3-Q-19 等価音源法に基づく高速かつ汎用的な 光学的 3 次元音場再構成

Fast and versatile three-dimensional acousto-optic sound-field reconstruction based on equivalent source method

○石川憲治, 谷川理佐子, 原田登(NTT)

- 光学的音響計測における音場再構成
 - 観測値である音圧線積分から元の音圧分布を復元する問題
- 従来手法: 物理モデルベース, 深層学習ベースなど
 - 精度・速度・汎用性等の間にトレードオフ
 - 大規模かつ複雑な条件で速度・精度低下が課題
- 提案法: 柔軟かつ高速な再構成手法を提案
 - ✓ 測定空間に合わせた等価音源配置による精度・汎用性
 - ✓ 座標回転とルックアップテーブルによる高速化
- 時間変化を含む大規模 4 次元音場再構成を数十秒で実現

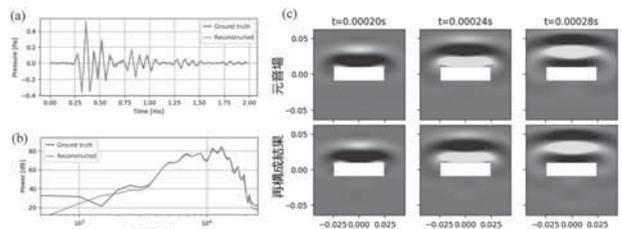


Fig.1 再構成結果 (a) 時間波形, (b) 周波数スペクトル, (c) 2D スライス可視化

3-Q-20

3-Q-20 胃腸電図と腸蠕動音の相関に関する基礎的検討

A basic study on the correlation between electrogastrogram and intestinal peristalsis sounds

○油谷壮真, 大塚 翔, 中川誠司(千葉大)

- ◆腸蠕動音は経験的に腸機能の診断に用いられており、近年は排便予測への応用も試みられている。その発生部位や発生メカニズムの詳細は不明だが、主に腸の内容物やガスの移動によって発生することが知られており、腸の蠕動運動を直接的に反映するわけではない。
- ◆一方、胃や腸管の平滑筋の電気的活動は胃腸電図として計測される。胃電図は微弱で周波数の遅い信号であり、呼吸や腹筋の動き、心電図などのアーチファクトとの分離が難しいという課題を有しているが、胃や腸管の運動そのものを反映している。
- ◆本研究では、腸機能診断や排便予測における相補的利用のための基礎検討として、腸蠕動音と胃腸電図の同時計測を行い、両者の相関について検討した。
- ◆一部の電極データ(胃腸電図)において聴診器データ(蠕動音)との間に負相関の有意傾向が認められたが(Fig.1のE2)、他の電極データにおいては同様な相関は認められていない。電極と聴診器の計測位置や計測時間長を変化させて、再検討する必要があると思われる。

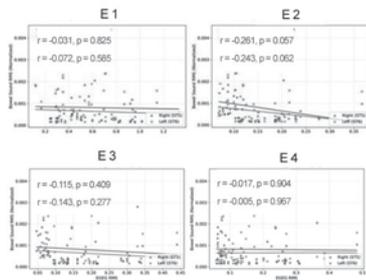


Fig.1 Scatter plots of RMS for EGE and bowel peristalsis sounds (blue: right, red: left)

3-Q-22

3-Q-22 ヘッドフォン VR サウンドの再生において頭内定位を生じる要因に関する研究 —本人の HRTF を使用した場合—

A study on the factors that cause for the lateralization in headphone-based VR sound reproduction - using the subject's own transfer function-

○坂爪亮介, 工藤彰洋(苫小牧高専), 武居周, 坂本真人(宮大工)

HRTF を用いた VR サウンド再生において、ITD、ILD、回折効果が音像の頭内定位化に与える影響を解明する聴覚心理試験

刺激音の条件として ITD、ILD、回折効果(インパルス応答をデルタ関数でモデル化)の有無、および両耳に同じ信号を再生するダイオティック信号を基準に、7種類の刺激音を生成した。

提示方位は 0° (真正面)、45°、90° (真右)、135°、180° (真後ろ) の 5 方向とした。

一対比較法

- すべての刺激音を対にして提示(方向固定、条件ランダム)、距離の違いを 5 段階(2, 1, 0, -1, -2)で採点した。
- Scheffe の方法により条件間比較を実施。

単音源の絶対判断

- 7 種類の刺激音を 5 回ずつ提示(方向、条件共にランダム)し、座標系を表示した GUI で距離定位を絶対判断させた。
- 定位した座標位置を距離値で換算して評価。

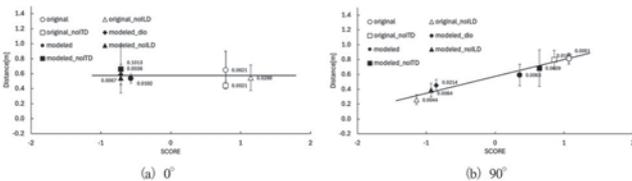


Fig.1 Correspondence between two types of tests

0° 方向では、回折効果の影響はあるものの、それが支配的な要因とはならない

90° 方向では、ILD の有無が頭内定位を引き起こす支配的な原因である

3-Q-21

3-Q-21 組み合わせ最適化に基づく HRTF の個人化手法の検討

HRTF Customization Method Based on Combinatorial Optimization

☆角野真弓, 豊岡祥太, 梶川嘉延(関西大)

- ◆HRTF の個人差による定位違和感を解決するため、実測を伴わない個人化手法を提案する。
- ◆HRIR を t-SNE で次元削減し、正中面上で仰角ごとに階層的クラスタリングを実施した。各クラスタ内の HRTF 巡回順序を組み合わせ最適化問題として定式化し、連続試聴による効率的なチューニングを可能にした。
- ◆21 名による聴取実験の結果、前後混同率は 14.4% 改善し有意差を確認した(Fig.1(a))。一方、垂直誤差は有意に改善しなかった(Fig.1(b))。
- ◆今後は HRTF 探索構造の見直しや、方向の知覚基準の提示を含めた改良を進める。

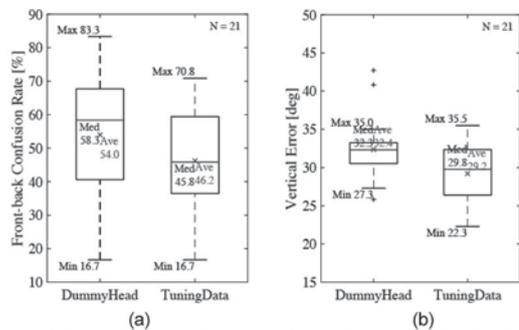


Fig.1: Box plots. (a) Front-back Confusion Rate and (b) Vertical Error.

3-Q-23

3-Q-23 実在感のある仮想音源の生成のための室内インパルス応答に含まれる残響成分の切り分けに関する検討

Separation of Reverberant Components in Room Impulse Responses for the Generation of Highly Realistic Virtual Sound Sources

☆小番涼音, 渡邊貫治, 西口正之, 安倍幸治(秋田県立大)

- ◆実在感のある仮想音源の実現を目指している
 - 室内インパルス応答の残響成分を抽出する方法を検討
 - 先行研究: 異なる聴取位置の残響成分を識別できた
 - 残響成分の切り出し位置を変更し、識別への影響を検討
- ◆インパルス応答の合成
 - 実測したインパルス応答をある切り分け位置 (CP と呼ぶ) で直接音・初期反射成分と残響成分に切り分ける (Fig.1)
 - 残響成分のみ異なる聴取位置のものに置換
- ◆主観評価実験の結果: CP に対する識別率 (Fig.2)
 - CP を変えても聴取位置の違いを識別できた
 - CP を大きくする(初期反射から残響までの遷移を含まないようにする) と識別しにくくなる

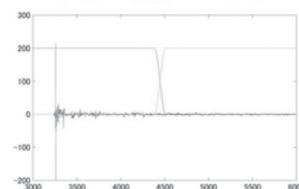


Fig. 1: Example of window hanging

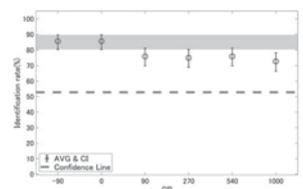


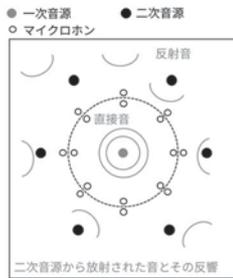
Fig. 2: Identification rate for each cut position

3-Q-24

3-Q-24 円筒調展開に基づく音場分離解析による 反射音除去の試み

An Attempt to Reflection Cancellation Based on
Sound Field Separation Analysis Using Cylindrical Harmonic Expansion
◎久代連太, 尾本章(九大・芸工)

- ◆比音響インピーダンスの能動的制御から発展して、音源放射特性の事前モデルが不要で、かつ厳密な反響除去を試みた。
- ◆まず、円筒調和展開を用いて、二重の円形マイクロホンアレイで計測された混合音場を「内部音場(反響成分)」と「外部音場(直接音成分)」に分離・推定する。
- ◆モードマッチング制御に基づき、推定された内部音場の展開係数を対象に二次音源によって相殺し、反射音成分のみを選択的に除去する。
- ◆シミュレーションの結果、一次音源の位置や放射指向性に依らず、制御後の音場が自由音場に近しい状態になることが確認された。
- ◆制御性能は音場分離解析の精度に強く依存しており、サンプリング精度が反響除去性能の支配要因であることが示唆された。



図：制御の概念図

3-Q-26

3-Q-26 複数同軸スピーカを用いた パーソナルスポット再生制御の検討

A study on personal sound zone control using multiple coaxial loudspeakers
☆遠藤要(電通大), 加古達也(NTT), 羽田陽一(電通大)

- ◆近傍では音が聞こえ、周囲では音が聞こえないといったパーソナルスポット再生の需要が高まっている。
- ◆MVDR ビームフォーマを用いて、2つの同軸スピーカを制御することで所望点での音色を保ちつつ、周囲での消音性能を高める制御手法を検討した。
- ◆実測 IR を用いた計算機シミュレーションで2つの通常スピーカと2つの同軸スピーカをそれぞれ MVDR ビームフォーマで制御した際の半径 30 cm の円周上における音圧を比較した。
- ◆所望点をスピーカの正面 2 cm に設定すると、低域において比較対象と同等の消音性能を発揮し、高域において提案手法がアレイ背面方向で比較対象に対して高い消音性能が得られた。

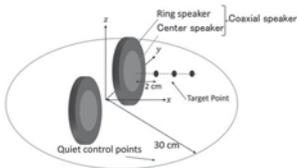


Fig.1: Coaxial loudspeaker array and control point layout.

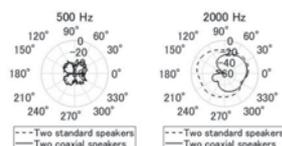


Fig.2: Sound pressure at the quiet control points when target point is 2 cm.

3-Q-25

3-Q-25 有限要素法を用いた マルチゾーン音場制御のロバスト性検討

A Study on the Robustness of Multi-Zone Sound Field Control Using FEM
◎原崇人, △竹久修平(旭化成エレクトロニクス)

- ◆マルチゾーン音場制御技術を応用・実用化する上では、環境変動に対するロバスト性が求められる
- ◆本発表では車室への応用を想定し、低域再生に適した既存ドアスピーカを用いた前後座席間の鳴らし分けを対象とする
- ◆有限要素法解析で得られる複数条件による音源～空間ノード間の伝達関数に基づき、設計時に用いた伝達関数とは異なる条件に対する空間平均音響コントラストを定義することで、乗員条件の差異などの環境変動が音場制御性能に及ぼす影響を評価した

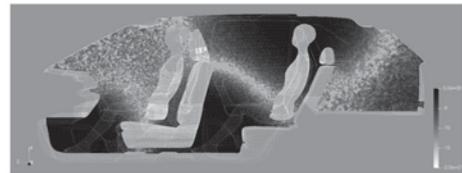


Fig.1: Example of acoustic contrast distribution at 160 Hz under sound field control

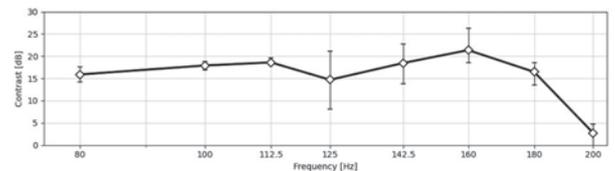


Fig.2: Example of acoustic contrast under sound field control. Error bars indicate variations across different occupant conditions.

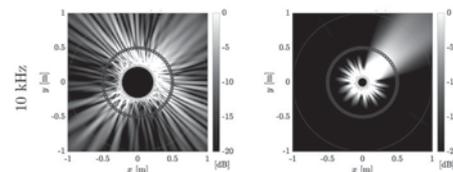
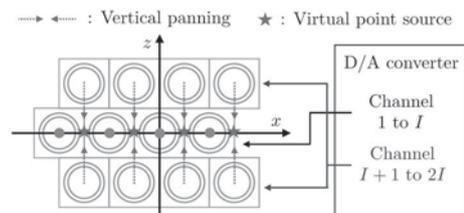
3-Q-28

3-Q-28 メッシュ型3段スピーカアレイを用いた マルチスポット再生

Multiple sound spot synthesis
with meshed triple-layer loudspeaker arrays
◎岡本拓磨(情報通信研究機構)

「こちらオカモト、空間ナイキスト周波数を倍増できるマルチスポット再生方式ができたぞ!! Over!!」
『Good job オカモトさん、欲しかったのは、それだけです!!』

- ・メッシュ型3段スピーカアレイを提案
- ・スピーカ間隔を変えずに空間ナイキスト周波数を倍増
- ・上下のアレイはパニングするためD/Aのチャンネル数は2倍
- ・計算機シミュレーションにおいて提案法の有効性を確認



(a) Simple linear with 24 loudspeakers & 24 ch D/A (b) Proposed meshed 3-layer with 3x8 loudspeakers & 16 ch D/A

3-Q-29

3-Q-29 JMS : NICT日本語複数話者 複数スタイル音声合成コーパス

JMS: Japanese multi-speaker multi-style speech synthesis corpus provided by NICT

○岡本拓磨, 大谷大和, 小椋志忠, 河井恒
(情報通信研究機構)

・日本語音声合成用公開コーパスJMS

- ・ CC BY-NC-SA 4.0ライセンスにて2026年3月に無償公開
- ・ 高品質: サンプル周波数48 kHz, 24 bitリニアPCM
- ・ 3スタイル
 - ・ 通常発話100文, 子役発話100文, 老人役発話100文
- ・ 500話者
 - ・ 女性プロ話者: 20名
 - ・ 男性プロ話者: 20名
 - ・ 女性アマチュア話者230名
 - ・ 男性アマチュア話者230名
- (ナレーター専門学校2年生以上)
- ・ 合計15万発話 (= 3スタイル x 100文 x 500話者)
- ・ 日本語標準語アクセント



<https://ast-astrec.nict.go.jp/en/release/jms/>

3-Q-31

3-Q-31 wav2vec 2.0 の中間表現を用いた 第二言語英語音声の発音誤り検出・診断モデル

Mispronunciation detection and diagnosis for second language English speech with wav2vec 2.0 intermediate embeddings

☆前田幸洋, 宮原絃造(同志社大院), 加藤恒夫, 田村晃裕(同志社大)

- ◆発音誤りの検出・診断(MDD)は, 第二言語学習者に発音誤りに対するフィードバックを行うシステムである。
- ◆MDD の研究において, 発音誤りを含む発音ラベル付きのデータ不足の問題に対して, 音響特徴量と音素特徴量, 言語特徴量の3種類の特徴量を用いることで性能の向上が見られる。
- ◆本研究では, 音素特徴量の代わりに wav2vec 2.0 の中間表現を用いた MDD モデルを提案する。
- ◆L2-ARCTIC における実験結果において, wav2vec 2.0 中間表現を用いたモデルが, 従来の音素特徴量を用いたモデルよりも性能を上回った。

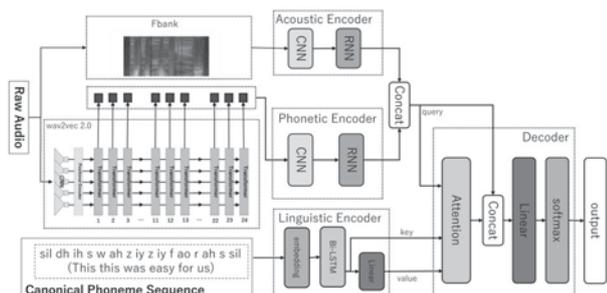


Fig.1 Architecture of MDD with wav2vec 2.0 intermediate embeddings

3-Q-30

3-Q-30 NICT 日本語複数話者複数スタイル音声 合成コーパスのための知覚表現および 話者印象語データセットの構築

Build of perceptual and impressive expression data sets for Japanese multi-speaker multi-style speech synthesis corpus developed by NICT

○大谷大和, 岡本拓磨 (NICT), 戸田智基 (名大/NICT), 河井恒 (NICT)

- ◆知覚表現や話者印象語による直感的な声質制御音声合成の実現を見据え, NICT 日本語複数話者複数スタイル音声合成コーパス (JMS) 向けの知覚表現および話者印象語データセットを構築。
- ◆アノテーション作業ではアンケート形式で各音声種別の声質に対する年代, 知覚表現対のスコアおよび話者表現語の選定を実施。
- ◆本データセットは知覚表現対スコアおよび話者表現語の頻度情報, ならびに作業者が記述した話者の印象のまとめが含まれており, JMS に同梱する形で公開。

Table 1: Perceptual representation pairs.

明るさ (暗い - 明るい)
硬さ (柔らかい - 硬い)
明瞭さ (くぐもった - 明瞭な)
流暢さ (たどたどしい - なめらか)
透明感 (かすれた - 澄んだ)

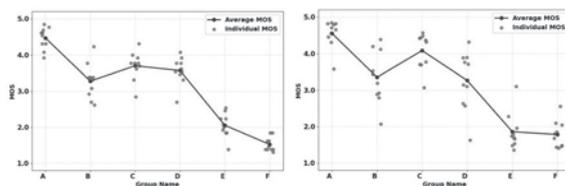
3-Q-33

3-Q-33 少量データ学習に基づく 合成音声品質の自動評価

Automation of Speech Quality Assessment Based on Small-Data Machine Learning for Application to Synthetic Speech

☆劉 正源, 志賀 芳則(東京電機大院・工学研)

- ◆近年, 音声合成技術の発展とともに, 合成音声の品質が向上し, 合成音声に対する品質評価の需要も高まっている。
- ◆こうした合成音声の品質の評価は, 人間の感覚に依存した主観評価によって行われることが多いが, 労力と時間がかかる。
- ◆一方, 従来の品質評価の自動化モデルでは, 合成音声を対象として適用する際に, 性能が低下することがある。
- ◆本研究では, 合成音声に特化するため, 短時間かつ少量データでも学習可能な音声品質評価の自動化方式を提案する。
- ◆提案手法では音声の MFCC と F0 の標準偏差を入力特徴量とし, SVM をベースに, Bagging Ensemble を取り入れた学習手法を採用した。
- ◆提案手法による評価結果は, 主観評価実験による結果と比べ, 高い一貫性を示した。



(a) Subjective Evaluation Experiment

(b) Proposed Model

Fig. 1: Evaluation Results (MOS) by the Subjective Evaluation Experiment and the Proposed Model

3-Q-34

3-Q-34 AESCA: AudioMOS Challenge 2025 のための生成音の美的スコア予測システム

AESCA: A generative audio aesthetics score prediction system for the AudioMOS Challenge 2025

©山本克彦, 宮崎晃一, 関翔悟(サイバーエージェント)

◆国際コンペ: AudioMOS Challenge 2025 (AMC25) Track 2

- 目的: 生成音声, 生成音楽, 生成音のための評価指標の開発
- 予測対象: 『音の美的スコア』に基づく4軸の主観評価値
 - ✦ 制作物の品質と複雑度 (PQ/PC)
 - ✦ 内容の楽しさと利便性 (CE/CU)

◆提案手法 (AESCA): 2種類の予測器によるアンサンブルモデル

- KAN ベース Audiobox Aesthetics: ベースラインモデルを Kolmogorov-Arnold Network で拡張
- VERSA ベース予測器: リファレンスフリー評価指標の出力を XGBoost で回帰モデル化

◆結果: AMC25 Track 2 において『世界一の予測性能』を達成

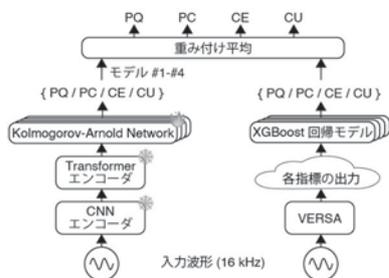


Figure 1: Overview of the proposed system (AESCA) for AMC25 Track 2.

3-Q-36

3-Q-36 選択的状態空間モデルの学習データ長に対する文脈考慮能力獲得に関する検証

Study of the acquisition of context-awareness ability for the selective state space model depending on the length of training data

©松井清彰, 三村正人, 松浦孝平, 森谷崇史, 浅見太一(NTT)

- ◆選択的状態空間モデルの一種であるMambaを音声認識のエンコーダとして活用したMambiformerについて long-form 音声に対して時間方向に対する精度変化を調査。
- ◆long-form 音声の評価用に, Librispeech の評価セットの元となるチャプター全体の音声を LibriVox から取得し, 評価セットとして整備。
- ◆実験の結果, Mambiformer は, Conformer と比較し発話中のほぼ全ての区間において精度面で勝る一方, Conformer と類似した精度変化傾向を示した。Mambiformer も一定幅の文脈までを考慮して推論を行っている可能性が考えられる。

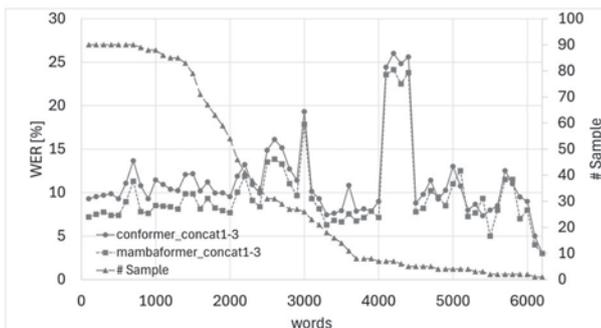


Fig.1: Word Error Rate(WER) for each 100-word segment on the long-form speech test set

3-Q-35

3-Q-35 音素単位の音声認識と LLM による文字変換を用いた脳性麻痺者のカスケード型音声認識

Cascaded Speech Recognition for Individuals with Cerebral Palsy Using Phoneme-Level Acoustic Recognition and LLM-Based Phoneme-to-Character Conversion

☆大谷越¹, 相原龍², 高島遼一³, 滝口哲也¹, 矢野藤辰彦²

(¹ 神戸大学, ² 三菱電機, ³ 立命館大学)

- ◆本研究は脳性麻痺者を対象とした文字単位の音声認識を目的とする。
- ◆構音障害者の音声認識モデルの学習を行う際, 構音障害者の音声を大量に収集することが困難であるという課題がある。
- ◆また, 日本語の多様な文字体系により, 文字単位の音声認識には大量のデータが必要であり, 構音障害者音声認識への適応が困難となる。
- ◆本研究では, 音素単位の音声認識モデルと音素から文字へ変換を行う LLM を組み合わせた音声認識システム (Fig. 1) について検討する。
- ◆LLM はテキストのみで学習でき, 音素単位の音声認識モデルは少量の音声データで学習可能なため, この手法は構音障害者音声認識に有効と期待される。
- ◆さらに, データ不足の課題に対処するため, 音声合成によるデータ拡張を行い, 音声認識モデルの段階的なファインチューニングを行う。
- ◆本稿では, 特定話者音声認識モデルにより検証する。



Fig. 1: Our proposed system architecture

3-Q-37

3-Q-37 シングルチャネルマルチトーカー音声認識と音声翻訳の統合モデリングの検討

Joint Modeling of Single-Channel Multi-Talker Overlapped Speech Recognition and Translation

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

- ◆一般に, 音声認識 (ASR; Automatic Speech Recognition) や音声翻訳 (ST; Speech Translation) のモデルは1話者用であるが, 実用的な会議や対話のシーンではオーバーラップが発生
- ◆言語の壁を越えた効果的なコミュニケーションの実現には, 精度の高いマルチトーカー音声認識 (m-ASR; Multi-Talker ASR) やマルチトーカー音声翻訳 (m-ST; Multi-Talker ST) のモデル構築が不可欠
- ◆m-ASR と m-ST を個別に実行することにより, 複数人の書き起こし文と翻訳文を獲得可能だが, 個別に学習・推論した場合, 書き起こし文と翻訳文は意味的に同じであるにも関わらずその関係を生かしきれない
- ◆本稿では音声認識と音声翻訳をシングルモデルで実行可能なm-ASRとm-STの統合モデリング (m-ASR-ST) を提案
- ◆m-ASR と m-ST を個別に実行する場合と比較して, 提案手法は複数人の重畳音声に対して ASR と ST を高精度に実行可能であることを確認

Lang.	Method	PT	# speakers in test							
			1		2		3		ALL	
			WER(L)	BLEU(T)	WER(L)	BLEU(T)	WER(L)	BLEU(T)	WER(L)	BLEU(T)
Ja→En	Separate m-ASR & m-ST	✓	4.84	24.17	7.67	20.68	12.26	18.72	9.49	20.16
	m-ASR-ST	✓	4.84	25.45	7.67	22.90	12.26	21.63	9.49	22.59
En→De	Separate m-ASR & m-ST	✓	17.03	23.04	23.86	20.89	30.11	16.15	25.02	19.41
	m-ASR-ST	✓	17.08	22.87	23.67	21.03	29.44	16.49	24.77	19.37
			16.82	23.00	22.84	21.16	27.90	16.85	23.64	19.68

Table1. Word error rates (WER) and BLEU scores in different methods. "PT" denotes initialization with a pre-trained m-ASR model.

3-Q-38

3-Q-38 wav2vec 2.0 の中間表現を用いたストリーミング Break Index 検出
Streaming Break Index Detection
Using wav2vec 2.0 Intermediate Representation

☆宮原絃造, 加藤恒夫, 田村晃裕(同志社大院)

- ◆Break Index は、韻律情報のラベリングスキームである ToBI システムに含まれる。単語・句境界などを表す音声セグメント情報であり、音声の基本周波数やピッチに基づいて付与される。
- ◆中でも、イントネーション句境界は発話の意味的な区切りと一致する傾向にあり、意味的な区切りを音響的情報から直接予測できる可能性があるため、ストリーミング音声翻訳への応用が期待できる。
- ◆Break Index 検出の既存研究は、許容誤差±40ms で 77% の F1 値を示しているが、ストリーミング処理には対応していない。
- ◆本研究では、wav2vec 2.0 のストリーミング実装である wav2vec-S を用い、その中間表現(各 Transformer 層の中間出力)を特徴量とするストリーミング Break Index 検出モデルを提案する。

Table 1: Event detection accuracy for Break Index=3 (intonation phrase)
・Previous Study:wav2vec 2.0+F0→BiLSTM
・Baseline:wav2vec-S(Final Layer Output)→TransformerEncoder
・Proposed:wav2vec-S(1-24 Layer Outputs) →TransformerEncoder

Tolerance	0ms			40ms			80ms			100ms		
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Previous Study	0.61	0.56	0.58	0.90	0.68	0.77	0.90	0.69	0.78	0.90	0.69	0.78
Baseline	0.56	0.36	0.44	0.79	0.43	0.56	0.80	0.44	0.56	0.80	0.44	0.56
Proposed	0.61	0.45	0.52	0.82	0.53	0.64	0.82	0.53	0.64	0.82	0.53	0.65
Proposed(FT)	0.64	0.59	0.61	0.82	0.68	0.74	0.82	0.68	0.75	0.83	0.69	0.75

3-Q-40

日本語音声の音響特徴に基づく
フィラー・吃音の検出

Detection of Fillers and Stuttering Based on Acoustic Features in Japanese Speech

☆峰野 侑也(豊橋技科大), 山岡 大志(株式会社ナレッジワーク)

- ◆本稿では、コミュニケーションの質的評価に不可欠なフィラーおよび吃音を検出するため、音響特徴量と分類器を用いた検出と、時系列情報を保持する End-to-End モデル (FluentNet) の比較検討を行った。
- ◆ベースラインとして MFCC 等の統計量を用いた実験の結果、時間情報を圧縮する従来手法では、動的な変動を持つ吃音の検出精度に限界があることを確認した。
- ◆提案手法は実環境データにおいて、二値分類では良好な精度を示したが、4クラス分類では複合クラスの曖昧性などにより有効な精度向上には至らなかった。
- ◆フィラーを音韻特徴に基づき詳細に分類して評価した結果、タイプによって認識されやすさに顕著な違いが見られることを明らかにし、実環境特有の課題として整理した。

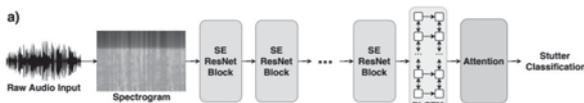


Fig.1: Architecture of the proposed FluentNet model based on SE-ResNet and Bi-LSTM with an attention mechanism.

3-Q-39

3-Q-39 言語尤度を考慮した
End-to-End 音声認識モデルの信頼度推定
Confidence Estimation for End-to-End Automatic Speech Recognition
Models Considering Language Likelihood

☆峰野 侑也, 北條 圭悟, 高城 翼成, 若林 佑幸, 西村 良太(豊橋技科大), 太田 健吾(阿南高専), 北岡 教英(豊橋技科大)

- ◆本稿では End-to-End 音声認識における過信の問題を解決するため、トークンの事後確率から言語尤度を減算した「擬似音響スコア」を提案し、信頼度推定への有効性を検証した。
- ◆医療現場等では認識困難な箇所の特定が不可欠だが、言語モデルの影響により、誤認識に対しても高いスコアを与えることが課題である。
- ◆事後確率をそのまま信頼度とした場合と比較して、提案スコアは誤認識のピークを低スコア側へ移動させ、正解と誤りの分布を明確に分離させている。
- ◆また、疑似音響スコアを用いて信頼度推定モデルを構築した結果、言語モデルのスコアとは信頼度推定においてそれぞれ異なる役割を持つことが示された。

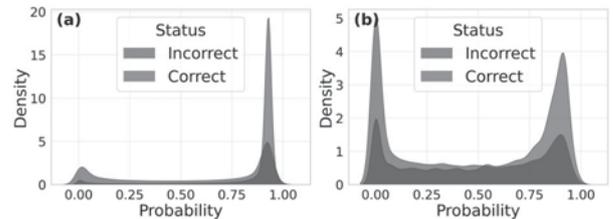


Fig.1: Comparison of histograms of ASR score distributions: (a) ASR score, (b) Pseudo-acoustic score.

3-Q-41

3-Q-41 音素ごとに異なる重みを持つ
パラメータ加重平均モデルによる
脊髄性筋萎縮症者の音声合成

Speech Synthesis for Individuals with Spinal Muscular Atrophy Using Parameter Weighted Averaging Models with Phoneme-Level Weight Control

☆長久保 諒¹, 八木 雄大¹, 高島 遼^{2,3}, 佐々木 千穂⁴, 滝口 哲也¹

(¹神戸大学, ²立命館大学, ³JST さきがけ, ⁴熊本保健科学大学)

- ◆本研究では、構音障害者の声質でかつ明確なテキスト音声合成システムの作成を目的として、健常者の子音との組み合わせを用いる手法を提案する。
- ◆構音障害は、言語理解に問題はないものの、発声発語器官の異常により発話が不明瞭となる障害であり、その音声を学習した合成音声も同様に不明瞭なものとなる。
- ◆提案手法では推論時に話者性が強く現れる「母音」と不明瞭になりやすい「子音」で使用するモジュールの重みを動的に切り替え、「母音」には構音障害者に近い特徴量を、「子音」には健常者に近い特徴量を用いることで、声質を維持しつつ明瞭性を改善する。

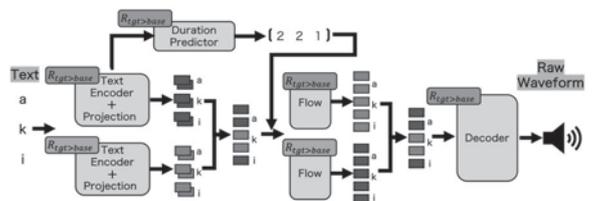


Fig.1 Overview of the proposed method

3-Q-42

3-Q-42 発話障害者の自発音声を対象とした非参照型明瞭度予測

Reference-Free Intelligibility Prediction for Spontaneous Speech of People with Speech Disorders using ASR-based Text Estimation

OHALPERN Bence Mark, 戸田 智基 (名古屋大学)

- ◆ 構音障害者の自然発話評価において、取得が困難な正解テキスト(参照)を必要としない客観的評価手法が求められている。
- ◆ 特性の異なる2種類のASRモデルを用い、「意図されたメッセージ(擬似正解)」と「知覚された音声(認識仮説)」の両方を自動推定して比較する手法を提案した。
- ◆ 人手による書き起こしなしに、推定された擬似正解と認識仮説の間の乖離(音素誤り率など)を計算することで明瞭性を算出する。
- ◆ YouTube Oral Cancer Corpus(自然発話データセット)を用いた実験において、提案手法は人間の主観評価と高い相関(0.82)を示し、正解テキストを用いる既存手法を上回る有効性が確認された。

3-Q-44

3-Q-44 LoRAによる多言語LLM-TTSの対象言語発音制御における適応方法と性能の分析

Analysis of adaptation strategies and performance for target-language pronunciation control in multilingual LLM-TTS using LoRA

©加藤集平(独立研究者)

背景

多言語LLM-TTSは高品質だが、日本語では語彙ごとのピッチアクセント誤りが残りやすい。単に対象部をカナにしても読みは改善できる一方、アクセント制御は解決しづらい。

提案

UtterTuneの枠組み(音素タグ+LoRA)を踏襲し、LoRAの挿入対象(自己注意のQ/K/V/O)と層位置(全層/下位6/上位6)を系統的に比較した。

結果

- ・ 難読語セット(difficult)でアクセント正解率が0.455 → **0.975**に上昇(all_qk)。同セットおよび非ターゲット入力セット(easy)で、客観指標(自然性・話者類似度・明瞭度)が劣化しないことも確認。
- ・ 難読語をカナ入力したkatakanaとの比較およびeasyのCERでは一部条件(all_qkvo, all_vo, lower6_*)で非劣化を満たさず、挿入箇所の選択が重要であることを示唆。

3-Q-43

3-Q-43 拡散確率モデルを活用した構音障害者音声の生成と音声認識性能向上の検討

Improving Dysarthric Speech Recognition with a Diffusion Model for the Training Data Generation

☆養毛悠晟, 陳金輝(和歌山大院), 陳訓泉(県立広島大)

- ◆ 構音障害とは、中枢神経や運動機能の障害により発話が不明瞭になる状態の総称である。構音障害者の音声認識はデータ不足が大きな課題となっている。
- ◆ 本研究では、拡散確率モデルであるGrad-TTSを活用して、構音障害者の音声を生成することでデータ不足に対処し、音声認識精度の改善を試みた。具体的には、健常音声による事前学習後、構音障害音声および合成音声を用いた転移学習を行った。
- ◆ 下図は実験結果を簡易的にまとめたものである。「重度」および「中程度~重度」の構音障害者に対して有意な改善が確認された。一方で、軽度の構音障害者では改善が限定的であり、重症度に応じた適用の重要性が示された。

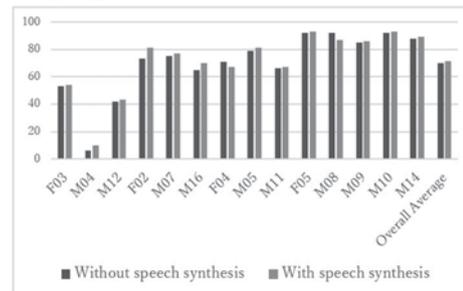


Fig. 1: WRA-based speech recognition accuracy for dysarthric speakers

3-Q-45

3-Q-45 潜在表現からの部分的推論による音声対話システム向け低遅延テキスト音声合成

Low-latency text-to-speech synthesis for spoken dialogue systems using partial inference from latent representations

☆白井成彦, 森大毅(宇都宮大学・工)

- ◆ VITSのText Encoderが推定したパラメータに基づいてサンプリングされた潜在表現zの列をセグメントごとに分割し、それぞれを独立にDecoderへ入力して音声を生成する「部分的デコード」を行う。
- ◆ 聴取実験の結果、提案手法は従来のテキスト全体の一括生成に迫る品質を保ちながらレイテンシを低下させることができた。

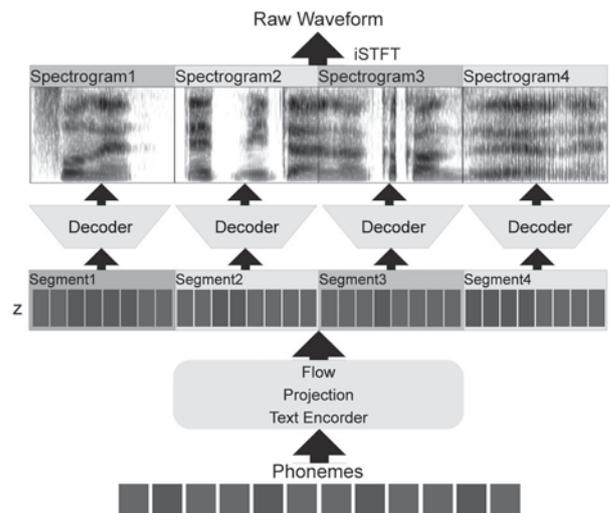


Fig. 1: Inference flow applying partial decoding from latent representations

3-Q-46

3-Q-46 コーパスにおける発声方法の違いがテキスト音声合成の残響環境下での明瞭性に与える影響の調査

Investigation of the Effect of Speaking Style Difference in Corpora on Speech Intelligibility in Text-to-Speech Synthesis for Reverberant Environment

☆栗本康太郎(名城大院・理工), 坂野秀樹, 旭健作

- ◆本研究では、通常の発声音声と模擬残響環境下での発声音声をそれぞれコーパスとして使い、テキスト音声合成を行うことで、発声方法の違いによって明瞭性がどのように変化するかを評価した。さらに、生成された合成音声に対して、明瞭性向上手法をそれぞれ適用し、残響環境下における効果についても同様に検証した。
- ◆実験の結果、通常発声合成音の正答率は42.0%であったのに対し、講義室発声合成音および体育館発声合成音ではいずれも48.0%を示した。また、明瞭性向上手法の効果は発声方法に依存し、通常発声合成音では一部の手法において明瞭性の向上が見られた一方、模擬残響環境下で発声された音声に対しては明確な改善が確認されない場合もあった。

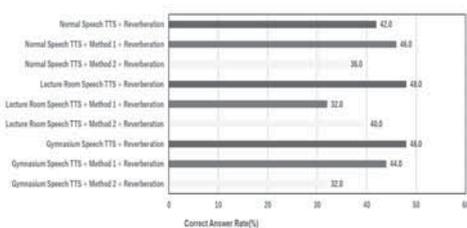


Fig.1: Word correct answer rate.

3-Q-48

講演取消

3-Q-47

3-Q-47 笑い声合成における call および吸気音の自動アノテーションの有効性

Effectiveness of automatic transcription of calls and inhalation sounds for laughter synthesis

☆上田 拓人, 森 大毅 (宇都宮大)

- ◆本研究では、笑い声合成に call および吸気音の自動書き起こしで作成されたデータを用いることが有効であることを示す。
- ◆自動ラベルを用いて作成した多数話者笑い声合成モデルの自然性の評価は人手アノテーションに基づくものに迫る自然性であった。
- ◆合成笑い声の自然性とラベリングのコストを考慮すると笑い声合成に自動ラベルを用いるのは有効な選択肢である。

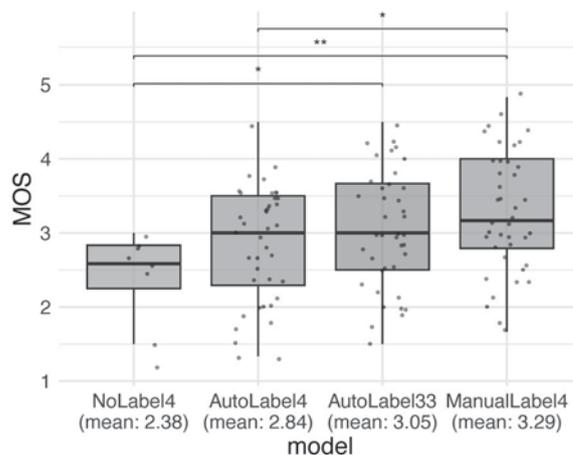


Fig.1: Naturalness MOS scores of synthesized laughter by labeling condition

3-Q-49

3-Q-49 笑い声合成に向けた特定話者自発対話音声コーパス構築の試み

Construction of a speaker-specific spontaneous conversation corpus for laughter synthesis

☆黛大誠, 有本泰子(千葉工大)

- ◆背景: 話者ごとに音響的特徴が変動する笑い声を合成するためには、特定話者の多くの笑い声を収録した音声資料が必要
- ◆目的: 特定話者の笑い声を大量に含む自発対話コーパスの構築
- ◆収録: 4名の特定話者による雑談を4か月に渡って毎日収録し、各話者の音声に対して笑い声をアノテーション
- ◆収録された笑い声数
 - 12月26日経過時点で34回の対話を行い、16時間の対話を収録
 - 男性話者Bから1,243個、総時間35.9分の笑い声を獲得
 - 女性話者Cから803個、総時間24.6分の笑い声を獲得

Table.1: Average number of laughs per speaker in each corpus

Corpus	Speaker[num]	laughter [num]	
		Ave.	Max.
Laughterscape	584	19	496
OGVC	36	197	354
Chiba3Party	36	31	57
DUEL	79	55	200
DisfluencySpeech	1	607	607
Proposal	4	576	1,243

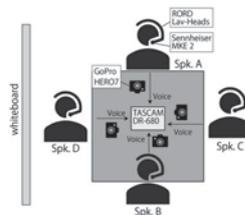


Fig.1: Recording Environment

3-Q-52

3-Q-52 音声感情認識と画像変換を用いた日本語感情字幕生成

Japanese emotional subtitle generation using speech emotion recognition and image translation

☆和田航次郎¹, 相原龍², 高島遼一³, 滝口哲也¹, △齊藤辰彦²
(神戸大¹, 三菱電機², 立命館大³)

- ◆感情を連続値(×クラス)として扱って音声から感情字幕を生成する
 - 音声感情認識 + 画像変換モデル の統合
 - 利点: 感情の強弱と複数感情の混合の潜在空間上で表現可能
- ◆日本人20名を対象とした7段階リッカート尺度による主観評価実験により, 平均評価が4.0を上回り有効性が示された

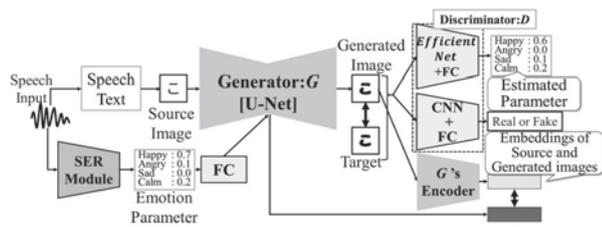


Fig.1: Structure of SER and Image-to-Image transformation model for generating subtitles

そ れ 様 を き き い お ほ え て い ま す か
(a)Happy 0.87, Angry 0.08, Calm 0.05 (b)Angry 0.99, Sad 0.01

わ か り ま し た そ う な ん で す か
(c)Angry 0.04, Sad 0.96 (d)Happy 0.05, Angry 0.06, Sad 0.05, Calm 0.84

Fig.2: Subtitles synthesized by our proposed method

3-Q-54

3-Q-54 歌唱音声の基本周波数に含まれる大局的変動の制御手法の改良とその自然性の評価

A study on improving the control method for global fluctuation of fundamental frequency in singing voice and evaluating its naturalness.

○鈴木千文, 坂野秀樹, 旭 健作(名城大)

- ◆これまで以下の方法でノンビブラート音声の基本周波数に含まれる微細変動と大局的変動の大きさを変化させた合成音を作成した
 - 基本周波数から回帰直線の傾きである微分信号を算出
 - 微分信号からハイパスフィルタとローパスフィルタにより微細変動と大局的変動を分離する
 - ◇ Fig.1(左)の基本周波数の時系列の比較より, 合成音の基本周波数は元のノンビブラート音声の基本周波数に含まれる微細変動がやや除去されている
- ◆以下の方法により微細変動の除去がされない合成音を作成した
 - 元のビブラート音声の基本周波数の移動平均を算出し, それを引くことで直流成分を除去した基本周波数を得て, 上記のフィルタリング処理を行う
 - ◇ Fig.1(右)の基本周波数の比較より, 合成音の基本周波数は元のノンビブラート音声の基本周波数をほとんど再現したものである

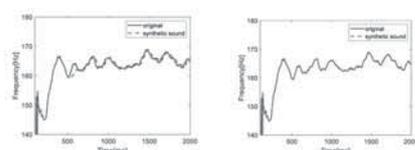


Fig.1. Comparison of fundamental frequency.

3-Q-53

3-Q-53 歌唱者間の相互作用と個人性を考慮した斉唱音声合成のための音高制御モデルの検討

Investigation of Pitch Control Models for Unison Singing Synthesis Considering Inter-singer Interaction and Individuality.

☆馬場開仁, △斎木陵佑, 齋藤大輔, 峯松信明(東大)

- ◆既存の歌声合成技術による斉唱の作成は, 独立に生成された複数の独唱音声を単純に重ね合わせる手法が一般的であり, 現実にかかる歌唱者間の相互作用を十分に再現できていない。
- ◆本研究では, 各歌唱者の内部的な音高基準の強さを表す「絶対音感度」や, 系統的なピッチのずれを表す「目標ずれバイアス」といった個人性パラメータを導入した, 斉唱音声合成のための音高制御モデルを提案する。
- ◆提案モデルは, 音高変化時に不連続に「ずれ」を更新する段階と, 音高持続中に周囲の歌唱者の平均音高へ向かって連続的に同調(引き込み)を行う段階の2段階で構成される。
- ◆2名の歌唱者による合成斉唱を用いた主観評価実験の結果, 提案モデルを適用した音声は, 単純な重ね合わせ(ベースライン)よりも「自然性」や「調和度」において有意に高い評価を得た。
- ◆特に, 絶対音感度に差を設けた「リーダー・フォロワー型」や, 記憶効果を高く設定した「相互聴取型」の制御が好まれる傾向が示された。
- ◆本モデルにより, パラメータ操作のみで「独立型」や「調和型」といった多様な斉唱の様態を単一のシステムで表現できる可能性が示唆された。