

2-1-1

2-1-1 音声生成系の様々な可視化手法

Visualization methods of speech production process

○北村達也(甲南大)

- ◆ 音声学の1分野である調音音声学にとって、発声・発話状態の可視化は重要な技術である。
- ◆ 発声器官の可視化手法
 - 間接的観測法: Electroglottography (EEG) など
 - 直接的観測法: ファイバースコープや硬性側視鏡など
- ◆ 調音器官の可視化手法
 - 磁気共鳴画像法 (MRI)
 - 超音波画像診断法
 - その他: 磁気センサシステム (EMA), エレクトロパトグラフィ (EPG), ビデオカメラなど
- ◆ 様々な可視化手法を体験できるハンズオンセミナーを開催され、講義動画が動画サイトにて公開されている。
- ◆ 国立国語研究所が公募する共同利用型共同研究にて、同研究所の装置を利用した研究ができる。
- ◆ これらのセミナーや制度を起点として、音声生成系に興味を持つ研究者の協力体制の構築が期待される。

2-1-2

2-1-2 EMA を用いた日本語母音の時間的な伸長と調音運動の検討

Examination of relationship between vowel lengthening and articulatory movements using EMA

○白勢彩子(東京学芸大), 北村達也(甲南大)

- ◆ 日本語の長・短母音のモーラリズムの生成機構を明らかにすることを目的に、磁気式計測システム (EMA) を用いて日本語共通語話者の母音発声時のデータを収集し、実証的に研究を進めてきた。本稿では、口角の運動に焦点を当て議論する。
- ◆ 発声リストは、/kabu/ (下部) → /kaRbu/ (カーブ) (Rは長音拍) のような母音の短・長の対である。共通語話者 19 名を対象に、EMA システムを用いて口角左、舌上など 5 点について計測した。
- ◆ /i/, /u/, /a/ の 3 母音とその長母音について、口角(LS)と、/i/では舌上(TT)のセンサを短・長間で比較した。対象母音区間のセンサ値の最大と最小の差分を、運動の大きさとして算出した。
- ◆ LS センサでは/i-/iR間、/u-/uR間で長母音の運動がより大きかった (P=0.004, P<.000)。/a-/aR間のLS, /i-/iR間のTTでは差がなかった。
- ◆ 口角の調音運動においても母音の短・長により相違があること、口角の運動が当該母音の生成において主要な動きである場合に、母音の短・長で運動に相違が生じることが考えられた。

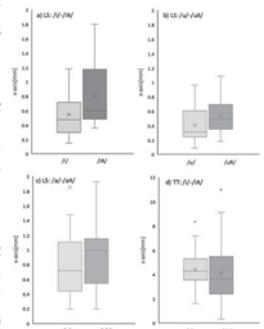


Fig.1 Results of size of sensor movement.

2-1-3

2-1-3 ポップアウトボイス話者の声帯振動観測の試み

Preliminary observation of vocal fold vibration for speakers with/without "pop-out voice"

○北村達也(甲南大), 榊原健一(北海道医療大), △内尾紀彦, △山中純太(そらいろ耳鼻咽喉科センター北駅前院), 能田由紀子(国語研), 天野成昭(愛知淑徳大)

- ◆ 喉頭ストロボスコープにより、大きな背景雑音下でもよく聞こえる声「ポップアウトボイス」の発声時の声帯振動を直接的に観察した。
- ◆ 実験条件
 - 参加者: ポップアウトボイスを有する話者 A と通らない声を有する話者 B
 - タスク: 各話者の話声区から決定した 4 つの高さで、母音/e/をポップアウトボイスとそうでない音声にて発声
- ◆ 話者 A は声帯結節の影響で緊張性の高い発声が習慣化されていることがポップアウトボイスにつながっているのではないかと。
- ◆ 話者 B の 2 種類の発声において声門開放時間率には差がなかったが、目視では声門開閉の動きが大きい様子が観察された。

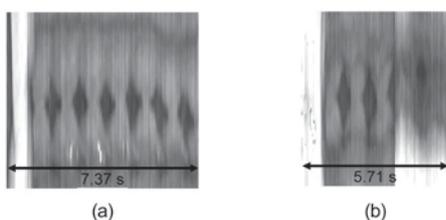


Fig.1:Kymograph of (a) normal and (b) pop-out voices at fo=167 Hz of Speaker B.

2-1-4

2-1-4 音声信号から発話器官の輪郭形状の逆推定

Inverse estimation of speech organ contours from speech signals

☆梶浦一真, △関根諒, 脇田真子, 竹本浩典(千葉工大), 平井啓之, 前川喜久雄(国語研)

- ◆ 音響特徴量から磁気センサシステムで得られたセンサコイルの位置情報を推定する試みもあるが、センサコイルの個数や位置に制限があるため調音器官全体の形状の推定は困難である。
- ◆ 本研究では、調音器官全体の形状と運動を観測できるリアルタイム MRI を用いて調音運動を動画として記録し、フレームごとに調音器官の輪郭を抽出した。また同時に録音した音声から WORLD を用いて音響特徴量を抽出した。そして DNN と CNN 2 種類のネットワークモデルを用いて音響特徴量からフレームごとの調音器官の輪郭点 (調音運動) の推定を試みた。
- ◆ 推定した輪郭点と正解となる輪郭点の MAE (平均絶対誤差) は 1.07 pixel であった (Table 1)。また、平均的な推定精度をもつフレームにおける 2 つの輪郭は目視でもよく一致した (Fig. 1)。

Table 1: Estimation accuracy (Unit: pixel)

	DNN	CNN
Tongue	1.73	1.79
UpperLip	0.56	0.80
LowerLip	1.16	1.48
Palate	0.85	1.27
UpperPharynx	0.78	1.11
LowerPharynx	1.01	1.03
Larynx	1.37	1.46
Average MAE	1.07	1.28

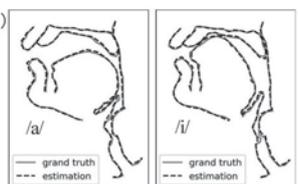


Fig. 1: Comparison between ground truth and estimation in the frame with average accuracy

2-1-5

2-1-5 舌モデルを用いた側音化構音症例の 訓練前後の舌形態の観察

Observation of tongue shape in lateral misarticulation before and after speech therapy using a tongue model

○武井良子, △森 紀美江, △山下夕香里, △多々良綾子(昭和大),
△長谷川和子(上伊那生協病院), △向井信彦(東京都市大)

- ◆側音化構音 (LM) の改善には誤った舌運動の修正が必要であるが、構音時の舌運動は視覚的に確認することが困難である。
- ◆構音時の舌形態を観察するために、2次元舌超音波画像を基に舌の3次元形状モデル(舌モデル)を構築するツールを開発した。
- ◆舌モデルを用いて健常人とLM症例の訓練前後の舌形態を比較した。
- ◆健常人の舌モデルは左右対称で、舌の正中に凹型がみられた。LM症例の舌モデルは、訓練前は左右非対称で舌の正中に凸型がみられたが、訓練後は健常人と類似した形態となった (Fig. 1)。
- ◆舌モデルは構音時の舌全体の形態を把握するツールとして有用であると考えられた。



Fig. 1: Tongue models of a normal speaker and LM patient before and after speech therapy.

2-1-6

2-1-6 口形に注目した発話訓練が音声に及ぼす 影響

Effects of voice exercise focusing on mouth shape on speech

☆設楽郁巳, 北村達也(甲南大), 牧野桃子, 山根典子(広島大)

- ◆本研究では、アナウンスの発話訓練にて口角を上げて発話するよう指導される点に着目し、発話時の口形に注目した発話訓練が音声に及ぼす影響について検討した。
- ◆発話時の口角をリアルタイムにフィードバックするシステムを開発した。システム利用者はPCのWebカメラに自分の顔を映し、口角の位置を上昇させながら発話を行う。
- ◆自宅にて2週間の間、1日1回発話訓練を行う実験を実施した。本システムを用いて発話訓練を行った実験群では、発話中の口角位置上昇が確認された。
- ◆発話訓練開始前の音声と、訓練終了後の音声を刺激音とする聴取実験を行った結果、実験群の発話訓練後の音声が高く評価された。



Fig. 1: Screenshot of proposed system

2-1-7

2-1-7 高速発話時の舌運動に関する予備的検討 —超音波診断装置を用いた観察から—

A preliminary study on tongue control during high-speed speech:
Observations using ultrasound

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

- ◆日本語母語話者3名が、徐々にスピードアップするメトロノーム音に合わせて、/pa ta ka/を繰り返し発音した。
- ◆話速が速くなると、/ka/が脱落 (Deletion) するという誤りが現れた。脱落時の舌運動は、正常時の/ka/と同様に後舌の挙上 (赤矢印) が観察された。

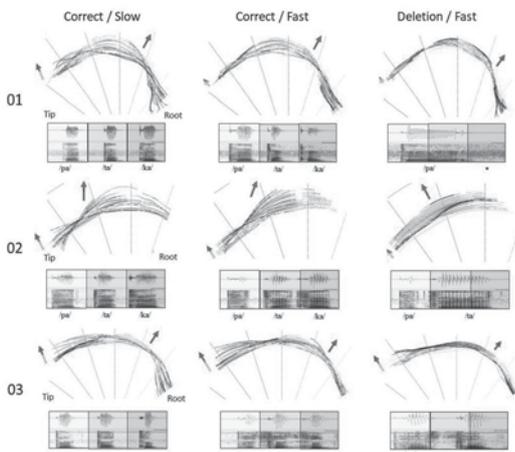


Fig. 1: Synchronized Relationship Between Tongue Configurations and Acoustic Signatures.

2-1-8

2-1-8 Using ultrasound to visualize the tongue during speech

Ian WILSON (University of Aizu)

Based on a presumed shared sense of proprioception, teachers give, and students understand, articulatory instructions such as “move your tongue higher in your mouth”. But how good is this shared proprioceptive sense? Do students even understand what teachers mean by their articulatory commands? With technological advances in imaging of the human body, teachers can now be sure of the instructions they give students, and students can get biofeedback amount the movement of their tongue and other articulators. Ultrasound imaging has been used in speech since at least as early as the late 1960s, but it has only more recently become widespread in speech applications. Does articulatory training work though?

Catford & Pisoni (1970, *Modern Language Journal*) showed that students’ production of unfamiliar sounds can greatly benefit from articulatory instruction. However, Horiguchi (2012, MSc thesis) used ultrasound to show that students could not follow simple tongue-movement instructions. But for English sounds such as /r/ and /l/, Japanese students can greatly benefit from seeing their articulation. In addition, recent ultrasound studies have shown native speakers’ preference for using bunched /r/ over retroflex /r/, something that may be very surprising to pronunciation teachers.

2-1-9

2-1-9 言語研究における調音データの活用例 —超音波を用いた舌形状分析の実践—

Utilizing articulatory data in linguistic research:
Practice of tongue shape analysis using ultrasound

○溝口愛(前橋工科大)

- ◆言語研究において超音波舌画像化 (Ultrasound Tongue Imaging: UTI) 技術の応用が進んでいる。UTIによる舌形状分析の実例を示し、分析方法について解説する。
- ◆活用例1 舌輪郭比較: 舌輪郭をモデル化しプロットすることで、音素ごとの舌形状の比較が容易となる (Fig.1)。
- ◆活用例2 舌の最高点分析: 舌輪郭の最高点および音響情報を用いて、多項ロジスティック回帰分析を実施。話者を予測するモデルと音素を予測するモデルを作成。調音と音響の関係を解明する。

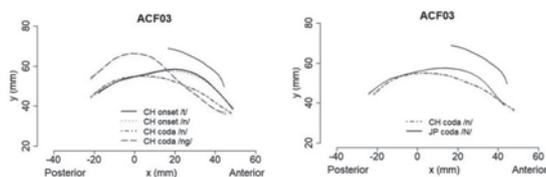


Fig.1: Tongue Contours for Chinese and Japanese Consonants by a Chinese Speaker.

2-1-10

2-1-10 リアルタイムMRI動画から抽出した 声道の輪郭に基づく音声合成

Speech synthesis based on vocal tract contours extracted
from real-time MRI videos

☆脇田真子, 竹本浩典(千葉工大), 平井啓之, 前川喜久雄(国語研)

- ◆深層学習の発展に伴い、Electro Magnetic Articulography やリアルタイムMRI (rtMRI) で観測した調音運動情報から深層学習モデルを用いて音声を合成する手法が研究されている。
- ◆本研究では、rtMRI データから調音器官の輪郭を抽出し、これらの座標データから音響特徴量 (F0: 基本周波数, mgc: メル一般化ケプストラム, cap: 非周期性指標) を予測し、音声を合成した。
- ◆客観評価実験では、テストデータ 50 文章での予測値と元音声から抽出した値との誤差を、F0, cap は RMSE (二乗平均平方根誤差), mgc はケプストラム距離で算出した。F0 の RMSE は 28.53 Hz, ケプストラム距離は 7.36 dB, cap の RMSE は 1.30 であった。(Table 1)。
- ◆主観評価実験では聴覚に問題のない 16 名がテストデータから無作為に抽出した 10 文章に対して抑揚の再現度、音韻性、品質の三項目を 1~5 で 5 段階評価した。抑揚の再現度は 3.69, 音韻性は 4.16, 音声の品質は 3.31 であった (Fig. 1)。

Table 1: Objective evaluation values for each acoustic feature

評価項目	評価値 (標準偏差)
F0 RMSE [Hz]	28.53 (5.37)
ケプストラム距離 [dB]	7.36 (0.77)
cap RMSE	1.30 (0.16)

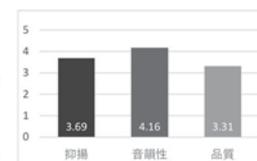


Fig. 1: Results of subjective evaluation

2-1-11

2-1-11 日本語の母音の調音における アクセント型の影響 —超音波を用いた予備分析—

The Effect of Pitch-Accent Patterns on Japanese Vowel Articulation:
A Preliminary Ultrasound Analysis

☆相葉瑛玲菜(上智大), 守本真帆(上智大/学振), 溝口愛(前橋工科大), 李瑋昱, 荒井隆行(上智大)

- ◆アクセント型が日本語の発話に与える影響を調査するため、無意味語の発話実験を行い、母音の音響的特徴(持続時間、フォルマント値)および調音動作(舌位置)を分析した。
- ◆アクセント型がこれらの音響的特徴と調音動作にもたらす影響は話者ごとに異なるが、第1フォルマントと舌の上下位置はおおむね一致する傾向が見られた。
- ◆アクセント型が直接的に舌の動きに影響している可能性が示唆された。

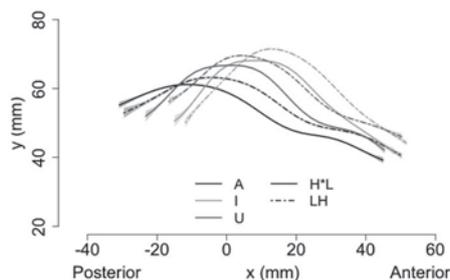


Fig.1: Example of tongue contours by vowel and pitch-accent pattern

2-2-1

2-2-1 自己教師あり学習に基づく事前学習モデルを用いた音声認識の概要と事例

Overview of automatic speech recognition using pretrained models based on self-supervised learning

○河原達也(京大・情報学)

- ◆枠組みと構成
 - トランスフォーマーと量子化器 (コードブック) の組合せ (図1)
 - 一部のフレームをマスクして復元するように学習
- ◆代表的なモデル
 - wav2vec 2.0 (6万時間) → XLS-R (多言語; 44万時間)
 - HuBERT (6万時間) → wavLM (デノイジング) → data2vec (量子化なし; 自己知識蒸留)
- ◆適用例
 - 少数言語 (クメール語・アイヌ語) の認識、感情認識
- ◆アダプタによる適応 (図2)
 - 雑音環境適応 (最下層)、話者適応 (上の方の層)

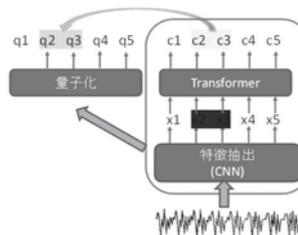


図1 wav2vec 2.0 / HuBERT の枠組み

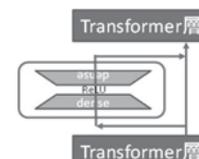


図2 アダプタ

2-2-2

2-2-2 (招待講演) 自己教師あり学習を用いた日本語事前学習モデルと音声認識・合成への応用

Japanese pre-trained models using self-supervised learning and their applications to speech recognition and synthesis

◎沢田慶, 法野行哉, 三井健太郎 (rinna)

- 近年のディープラーニングは、大量の学習データと大規模なモデル構造を用いることで性能が飛躍的に向上している。
- 自己教師あり学習とTransformerを組み合わせたwav2vec 2.0, HuBERT, WavLMなど多くの事前学習モデルが提案されている。
- 公開されている事前学習モデルの多くは英語の音声データから学習されているため、日本語の音声データからHuBERTを学習して公開した。
- HuBERTを用いた日本語音声認識において、日本語音声データで事前学習したモデルは英語音声データで事前学習したモデルよりも高い性能を示した。
- HuBERTを用いた音声合成・声質変換の手法を紹介し、日本語音声合成タスクで日本語音声データから事前学習したモデルが高い性能を示す例をあげた。
- 今後の音声研究では、強力な事前学習モデルを活用した手法を検討することが重要であろう。

2-2-4

2-2-4 自己教師あり学習に基づく音声・言語モデルを統合したEnd-to-End音声認識

End-to-end speech recognition by integrating self-supervised speech and language model

◎法野行哉, 光田航, 趙天雨, 三井健太郎, 若月駿亮, 沢田慶 (rinna)

- 近年、音声認識を含む、様々な機械学習のタスクで自己教師あり学習された基盤モデルの活用が注目されている。
- 本稿では、自己教師あり学習された大規模言語モデル及び音声表現モデルを統合したend-to-end音声認識を提案する。
- 提案手法では、音声波形からHuBERTにより抽出される音声特徴量をブリッジネットワークを介して音声プロンプトとしてGPTに与え、認識結果のテキストトークンを自己回帰的に生成する。
- 実験により、提案モデルが近年の音声認識モデルと同等の性能を達成することを示すとともに、提案モデルを「Nue ASR」として商用利用可能なApache-2.0 Licenseで公開した¹。

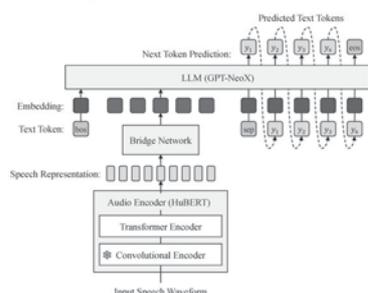


Fig. 1: Overview of the proposed model

¹<https://huggingface.co/rinna/nue-asr>

2-2-3

2-2-3 クロスモーダル表現学習の研究動向: 音声関連を中心として

Recent Advance of Crossmodal Representation Learning related to Speech

◎増村 亮 (NTT)

- ◆音声や画像、テキスト等の個々のモーダルに対する表現学習は、近年の自己教師あり学習の進展に伴い、急速に発展してきている。この潮流とともに、特定モーダルに留まらず、モーダル間の対応関係を利用したクロスモーダル表現学習に対する注目も高まってきている。
- ◆本稿では、クロスモーダル表現学習の中でも、音声に関連する研究に特に焦点を当てて、その研究動向を概説する。

2-2-5

2-2-5 AI エージェント間の自然な会話に向けたテキストからの音声対話生成

Generating Spoken Dialogue from Text for Natural Conversations Between AI Agents

◎三井健太郎, 法野行哉, 沢田慶 (rinna)

- ◆背景：大規模言語モデルの急速な進化
 - テキスト対話：AI 同士の自然な会話を生成可能
 - 音声対話：交互に音声合成するだけでは会話の自然さに課題
- ◆提案法：テキスト対話から自然な音声対話を生成するシステム
 - 音声から離散ユニット列を抽出し言語モデル (MS-DLM) を学習
 - 2 話者のユニット列 (話し手 + 聞き手の相槌・笑い) を同時生成
 - 円滑な話者交替を実現するために学習データを加工
- ◆結果：従来の音声合成ベースのシステムと比べて有意に自然性を改善
 - 実際の音声対話で発せられる相槌 (うん、んー、...) を生成可能
 - 話者交替に関連する指標でも実際の音声に近い分布を獲得

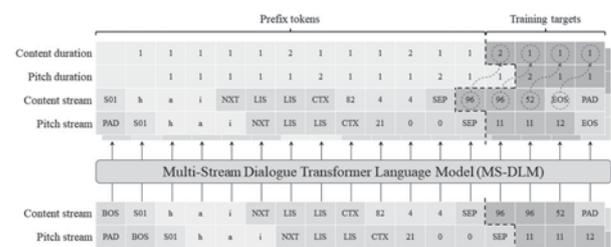


Fig.1: The proposed MS-DLM generates two channels of unit sequences based on written dialogue. The generated units are then converted into stereo waveform using HiFi-GAN vocoder.

2-2-6

2-2-6 Brain2Music: Reconstructing Music from Human Brain Activity

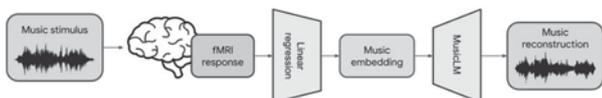
Timo Denk*†, ☆ 高木優**†,

松山卓矢**, Andrea Agostinelli*, 中井智也**,

Christian Frank*, 西本伸二**

* Google, ** 阪大, *** Araya, † 共同筆頭著者及び共同責任著者

◆ 脳活動と生成 AI の対応関係を探ることは、脳がどのように世界を表現しているかを理解するだけでなく、AI と脳がどのように関連しているかを解釈することもできるユニークなアプローチである。本発表では、近年我々が提案した、音楽生成 AI (Google MusicLM) とヒト脳活動と結びつけて脳活動から音楽を再構成した研究について紹介する。



◆ 図： 脳からの音楽再構成の概要図。

2-3-2

2-3-2 聴覚による身体近傍空間の方位依存性の検討

An investigation for directional selectivity on auditory peripersonal space representation.

◎寺岡諒(室蘭工大), △黒田尚輝, △寺本渉(熊本大)

- ◆ 身体近傍空間は、身体のすぐ周囲の空間であり、危険回避や周辺環境との相互作用において重要な役割を果たしている。
- ◆ 本研究では、身体の前後左右で聴覚刺激を呈示することで、身体周囲の身体近傍空間の様相を検討した。
- ◆ 実験の結果、身体近傍空間の様相が背後(180°)とそれ以外の方向(0°, 90°, 270°)で異なることを示した。

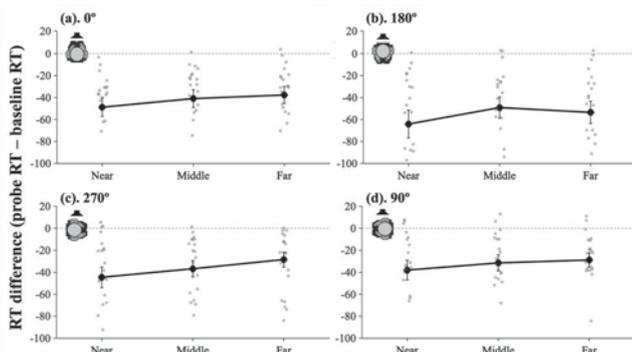


Fig. 1: Average RT difference (probe RT – baseline RT) across all listeners at each distance in each direction condition. Scatter plots show individual data. Error bars denote the standard error of the mean.

2-3-1

2-3-1 迫真型聴覚 AR による近接音響心理測定の試み

An Approach to Space-Time Psychoacoustic Measurement with Proximal Auditory AR

○李珍咏(東大), △石原茂和(広国大), △石原恵子(広国大), 北爪裕道(北教大), 伊東乾(東大)

- ◆ 我々研究グループが開発した、導電性高分子と有機ピエゾフィルムを用いた聴覚 AR ヘッドセットと、通常のスピーカーアレーを組み合わせることで、音像が遠方から耳に飛び込んできたり、あるいは耳から飛び出した音が遠方に去るような、動的な聴覚の錯覚を引き起こす、旧来存在しなかった「迫真型聴覚 AR」システムを開発した。
- ◆ このシステムである種の運動する音像を聴取すると、聴いている本人はまったく意識しないのに反射的に身体の揺動が発生する。この事実を利用して、立位の保持に難のある高齢者や疾病罹患者に迫真型聴覚 AR の動的な音像を提示することで、新たな原理に基づく「前庭リハビリテーション」を設計出来る。そこでこれを実装した。
- ◆ 完成したシステムを用いてパーキンソン病罹患者を被験者とするトレーニングを実施したところ極めて良好な初期結果が得られたので、これを報告する。



Fig. Left: Schematic view of vestibular rehabilitation system, Center: Training for a patient of Parkinson's disease, Right: Center-of-gravity sway of a patient after the rehabilitation trials.

2-3-3

2-3-3 気導音の伝達経路に基づく音像定位を活用した方向認知精度の向上

Improvement of direction recognition accuracy using sound image localization based on air-conducted sound transmission path

☆小澤風太(中央大院),

小暮知史, △丹野慶太(アルプスアルパイン), 戸井武司(中央大)

- ◆ 耳を塞がない聴覚のみによる方向指示として、本研究ではヘッドレスト等の頭部至近に設置した2ch スピーカによる方向認知精度の向上を目指した。
- ◆ 各種特性補正を行っても、頭部至近2ch スピーカでは耳位置のずれにより Fig. 1 に示す6~10 kHz 付近が弱れることを示した。
- ◆ 帯域を制限して音像定位実験を行い、頭部至近2ch スピーカでは Fig. 2 (a) に示す特性補正が弱れる帯域を削除した0~4 kHz の周波数域が音像定位に有効であり、またそのうち Fig. 2 (b) に示す0~100 Hz の再生周波域外の音を削除することで定位精度が向上することを示した。
- ◆ さらに、帯域制限において定位精度が向上した条件に、残響時間の異なるインパルス応答を畳み込みことで、Fig. 2 (c) に示す残響時間が1.0 s の場合に定位精度が高く、ヘッドホンと同等となる。これは空間特性を模擬した残響を付与することで、音像位置が明確化したためと考えられる。

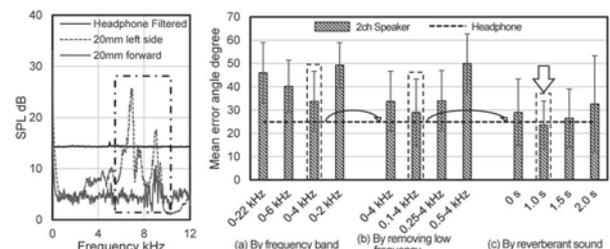


Fig. 1 Changes in characteristic correction results at the listener's ears depending on head position Fig. 2 Changes in sound image localization accuracy due to frequency band, bass range, and reverberation sound

2-3-4

2-3-4 気導音と骨導音の併用による音像定位を活用した方向認知精度の向上

Improvement of Direction Recognition Accuracy using Sound Image Localization by Combining Air-Conducted and Bone-Conducted Sound
☆小澤風太(中央大院), 小暮知史, △丹野慶太(アルプスアルパイン), 戸井武司(中央大)

- ◆聴覚のみによる方向指示として、耳を覆うヘッドホンに対して、耳を覆わない気導音と骨導音の併用による方向認知精度の向上を目指す。
- ◆正弦波に両耳間レベル差(ILD)を付与し、周波数ごとに定位精度を評価した。結果、Fig. 1に示す250 Hzと500 Hzにおいて気導音より骨導音の定位精度が劣ることを示した。これは骨導音において低音域では減衰が小さく音が反対の耳まで届くためであり、骨導音は音像定位において低音域の音像定位精度が低いと考えられる。
- ◆気導音と骨導音の併用について、頭部後方至近に設置した2chスピーカで低音域、骨導イヤホンで高音域を再生し、音像定位実験を行った。結果、Fig. 2に示す両者のクロスオーバー周波数が0~4 kHzで重なる場合に平均誤差角度が10度以下と最も小さく、気導音と骨導音の併用によりヘッドホン単体での定位精度を上回することを示した。

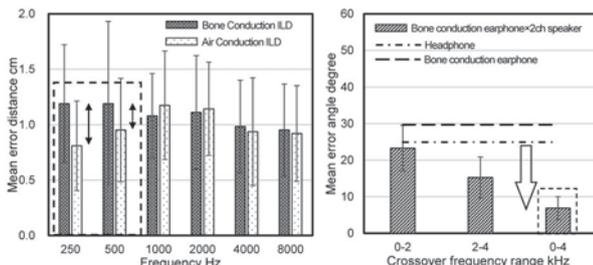


Fig. 1 Localization accuracy with ILD and ITD

Fig. 2 Changes in sound image localization accuracy when playback frequency ranges overlap

2-4-1

2-4-1 偏光高速度干渉計を用いた球面座標系における3次元音場復元

Reconstruction of three-dimensional sound field in spherical coordinates by using a high-speed polarization interferometer

☆野澤遙, 今西麻優子, 及川靖広(早大理工), 石川憲治(NTT)

- ◆背景：光学計測で得られるデータは音圧積分値。
➢物理モデルに基づく音場復元手法が近年提案されている。
➔球面座標系内で計測領域に音源を含むモデルは未提案。
- ◆目的：観測領域内に音源を含む計測データから3次元音場の復元。
- ◆提案法：偏光高速度干渉計で得られたデータに対して、外部問題を前提としたHelmholtz方程式に基づくモデルを用いた復元手法を提案。
- ◆結果：シミュレーションと実測データによる実験を行った。シミュレーションではノイズを含む状態でも精度よく復元可能なことを確認。実測データからは、3次元放射音場を復元することに成功した。

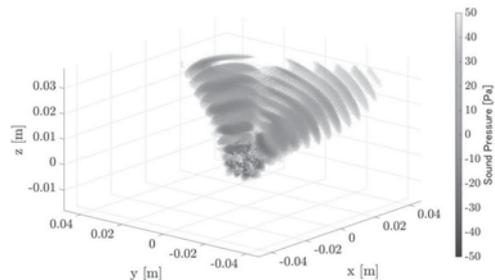


Fig.1. Three-dimensional sound field reconstructed from line-integrated sound pressure measured by a high-speed polarization interferometer.

2-3-5

2-3-5 血腫による耳介硬度の変化が軟骨伝導の伝搬成分に及ぼす影響

Effects of hematoma-induced changes in auricular hardness on propagation components of cartilage conduction

☆田村 茜, 大塚 翔, 中川 誠司(千葉大)

- ◆骨伝導の装着性の問題を解決する方法として“軟骨伝導”が提案され、補聴器やスマートホン、イヤホンなどに応用されている。
- ◆耳介の内出血による瘤状の隆起である耳介血腫は、イヤホン等の使用の妨げになる場合が多い。軟骨伝導デバイスが有効であると考えられる一方、血腫による耳介の硬度や質量の増加は軟骨伝導の聞こえに影響を及ぼすと考えられる。
- ◆血腫による硬度の増加が軟骨伝導の伝搬に及ぼす影響の調査を目的とし、血腫者(Hematoma)および聴覚健全者(Normal)の耳介の硬度、耳介/頭部振動、検出閾を計測し、それらの関係を検討した。
- ◆Hematomaの耳介/頭部振動は概ねNormalを上回った(Fig. 1)。血腫による硬度の増加が振動量の増大を招いたと思われる。しかしながら、検出閾は250-1000 Hzで上昇、4000 Hzで低下した。また、硬度と振動の関係は部位で異なった。
- ◆血腫が検出閾に及ぼす影響は部位や周波数によって変化する。メカニズムの全体像の把握は容易ではない。今後は耳介全体の硬度を含めた新たなパラメータの導入等が必要と思われる。

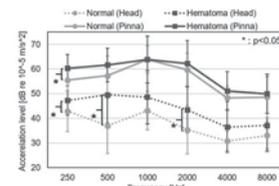


Fig. 1 Acceleration levels of the pinna and head.

2-4-2

2-4-2 音の物理特性を考慮した深層学習による音場イメージング画像雑音除去

Sound Field Image Denoising Using Deep Learning Considering Physical Characteristics of Sound

☆浦田大輔, 及川靖広(早大理工), 石川憲治(NTT)

- ◆背景：光学的手法による2次元音場の定量計測
➢音場イメージング画像に対する雑音除去が行われる
- ◆従来法：Deep Neural Networkに基づく雑音除去
➢学習の損失：正解値と予測値のみから計算される
- ◆提案手法：
➢音場を満たすHelmholtz方程式に基づく項をネットワーク学習の損失に付加
- ◆結果：推論画像のHelmholtz項が減少
➢画像としての復元品質を保ったまま、より音場に適した結果が得られた

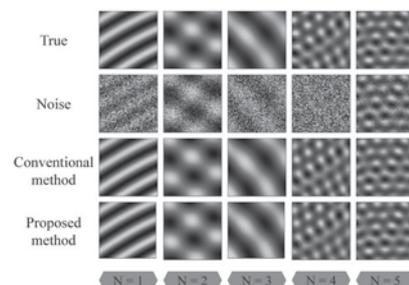


Fig 1: Comparison of denoising results for sound field simulation images. N is the number of sound sources.

2-4-3

講演取消

2-4-4

2-4-4 音響プロジェクターのための音場測定

Sound field measurement for Acoustic Projector

○松村泰河, 杉谷太誠, 須田宇宙(千葉工大), 高橋義典(工学院大)

- ◆現代では多チャンネル化など音響分野のオーディオ技術が著しく成長している。
- ◆新しいオーディオ技術として、音波をスクリーンに反射させることで仮想音源から音が出ているように錯覚させ、臨場感のある音場を作る音響プロジェクターの開発を提案した。
- ◆本研究では、音響プロジェクターの実現に向けて音場測定を行なった。
- ◆研究の結果、音波は幾何光学的に反射する成分が大きいことが分かった。この結果を活用して音響プロジェクターの構造を考え、構想の実現を目指す。

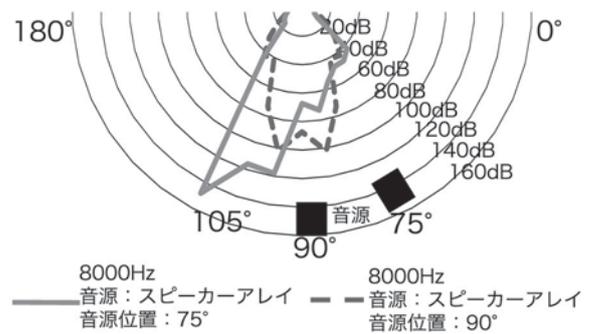


Fig.1: Sound field measurement results (Sound source position 75° and 90°)

2-4-5

2-4-5 自己運動を伴う人体に取り付けたマイクロフォンとトラッキングセンサを活用した音響イベント定位

6DoF SELD: Sound Event Localization and Detection Using Microphones and Motion Tracking Sensors on self-motioing human

◎安田 昌弘, 齊藤 翔一郎, 中山 彰, 原田 登 (NTT)

【音響イベント定位】

- 音響信号からいつ・どこで・何があったのかを推定する
- 本研究: 並進・回転 (6DoF) 運動する人間に取り付けたマイクロフォンによる新しい音響イベント定位のタスク (6DoF SELD) に取り組む

【課題】

- データセット: 無 (運動を伴うため、既存データからの生成も困難)
- 体の動きに伴ってマイクロフォンが回転運動
→ 外部音源の急速な相対移動を引き起こしパフォーマンス劣化

【提案】

- 6DoF SELDのためのデータセット提案
➢ 18chのマイクロフォンとトラッキングセンサを埋めこんだヘッドホン型デバイスを自己運動する人が装着し、データ収録
- センサによる頭部の回転・並進運動観測を利用したマルチモーダルシステムの提案 → 運動時のSELD性能向上

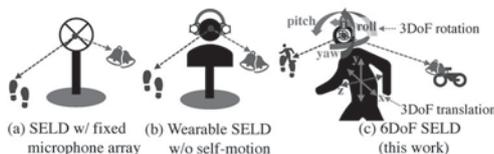


Fig.1: Conventional and our problem settings of SELD

2-4-6

2-4-6 環境音分析における事前学習済みモデルのバイアス調査

Investigation of Bias in Pre-trained Models for Environmental Sound Analysis

☆井上かほり, 井本桂右(同志社大学)

- ◆環境音分析における事前学習済みモデルがバイアスを含んでいると、目的タスクもバイアスの影響を受ける
- ◆事前学習済みモデルが人間の聴覚特性と異なる判断をすることによって生じるバイアスについて調査
➢ 人間はスペクトル包絡よりもスペクトルの微細構造を重視
➢ 環境音分析における事前学習済みモデルはスペクトル包絡と微細構造のどちらを重視しているのか調査
- ◆事前学習済みモデルの音響タグ付け性能を評価
➢ 音データの MFCC 成分のうち、低次成分または高次成分をそれぞれ他の音の同等なケプストラム次数の成分と入れ替える
➢ 実験結果: 事前学習済みモデルは人間と同じくスペクトル包絡を重視

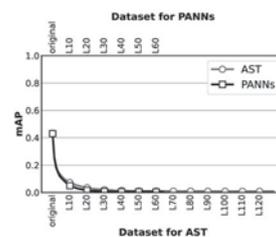


Fig. 1: Evaluation results for low-order replacement dataset (Ln is the low-order to n-dim replaced dataset)

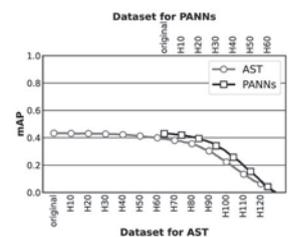


Fig. 2: Evaluation results for high-order replacement dataset (Hn is the high-order to n-dim replaced dataset)

2-4-7

2-4-7 生成的異常音検知における 識別的近傍平滑化

Discriminative neighborhood smoothing for generative anomaly sound detection

© 藤村拓弥, 戸田智基 (名古屋大学)

- ◆生成的異常音検知: 正常音の音響特徴量をモデル化し, モデルからの逸脱度により異常度を算出する手法
- ◆識別的異常音検知: 正常音の識別特徴量をモデル化する手法. 生成的手法よりも高い性能を達成するが, 識別器が学習評価の乖離の影響を受け, 性能が著しく低下する場合がある.
- ◆目的: 学習評価の乖離による性能低下リスクを排除しながら, 生成的手法よりも高い性能を達成すること
- ◆提案法: 識別特徴量空間での複数の近傍サンプルを用い, 生成的手法の異常度を平滑化によりアンサンブルする. 正常音識別特徴量モデルからの逸脱度は用いないため, 学習評価の乖離による影響は受けない.
- ◆実験: (1) 提案法が生成的手法 (AE) の性能を大幅に改善すること, (2) 識別的手法 (SerialOE) が学習評価の乖離による影響を受ける場合であっても提案法が有効であることを確認した (Table 1).

Table 1 Evaluation results. Values of each machine represent the harmonic mean of the AUC [%] and pAUC (p=0.1) [%] over all section. Values of the ToyCar-5 represent the AUC of ToyCar in section 5.

Method	fan	gearbox	pump	slider	ToyCar	ToyTrain	valve	ToyCar-5
SerialOE	78.97	63.02	76.56	81.49	49.37	55.05	63.02	33.58
AE	57.11	59.50	54.70	55.63	58.61	58.74	51.27	72.87
Proposed	60.20	63.89	68.90	69.28	57.90	67.60	55.20	77.29

2-4-9

2-4-9 騒音変化を考慮したフィードバック ANC システムにおけるバーチャルセンシングに関する検討

Study on Virtual Sensing in Feedback ANC Systems Considering Noise Changes

©豊岡祥太, 梶川嘉延(関西大)

- ◆ANC システムは騒音に対して同振幅, 逆位相の擬似騒音を生成し, 騒音と重ね合わせることで騒音低減を行うシステムである.
- ◆本稿では, 騒音に対して最適な騒音制御フィルタを切り替えることによって, 騒音変化にロバストなシステムを提案する.
- ◆提案法では, 事前に様々な帯域を持つ騒音に対して最適な騒音制御フィルタを複数設計する. そして, ANC 動作中に内部の参照信号に対して FFT を実行することで, 騒音変化を検知する.
- ◆騒音変化を検知すると, その騒音に対して最適な騒音制御フィルタを選択し, 補助フィルタを再推定することによって, 騒音変化に対してロバストなシステムになる.
- ◆Fig.1 に, 所望地点における騒音低減量を示す. 提案法では騒音が変化しても約 18dB の騒音低減効果を保つ.

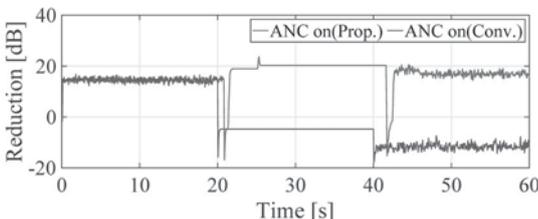


Fig.1: Noise reduction at the desired position.

2-4-8

2-4-8 音響信号を用いた時系列解析による モーターの故障予知に向けた基礎検討 -機械学習を用いた小型 DC モーターの 過電圧故障の判別-

Basic Study for Prediction of Motor Failures by Time Series Analysis Using Acoustic Signals -Discrimination of Overvoltage Failures in Small DC Motors Using Machine Learning-

☆池垣 聡司, 坂本真一 (滋賀県立大)

- ◆音響信号を用いた時系列解析によるモーターの故障予知に向けた検討として, 小型 DC モーターの過電圧故障の判別を行った.
- ◆モーターの駆動音の録音データ最大 60 分を学習データとして機械学習を行った. (Fig. 1, Fig. 2)
- ◆機械学習の結果, 判別正答率 100%を確認した.

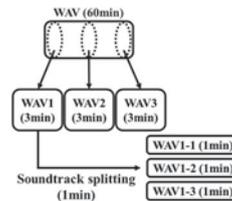


Fig. 1 Recording data analysis procedure (60 min).

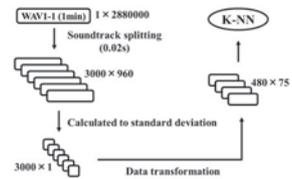


Fig. 2 Recording data analysis procedure (1 min).

2-4-10

2-4-10 評価関数から導かれる Filtered-x LMS 法の安定動作条件

A condition derived from the cost function for stably running filtered-x LMS algorithm

○藤井健作(コダウェイ研), △棟安実治(関西大・工), 首木禎史(千葉工大)

Filtered-x 法は, 騒音制御フィルタは線形フィルタとみなせるとの仮定の下で 2 次系と騒音制御フィルタの接続順序を入れ替えた構造で実行される係数更新法として導かれる. 更に, その仮定が有効となるためには, 騒音制御フィルタの係数は緩やかに更新される必要がある.

また, その構造で実行される Filtered-x LMS 法を導く評価関数は, 2 次系を 2 次系フィルタで置き換えた構造で構成される誤差マイクロホンの出力をゼロにする擬似最適解を 2 次系推定誤差がゼロである場合に誤差マイクロホンの出力をゼロにする最適解に近付けることによって構成される.

以上のことから, Filtered-x LMS 法を適用したシステムで安定した騒音制御が実行されるためには, ① 擬似最適解を最適解に近付けること, ② 係数更新が緩やかに実行されるとの仮定が成り立つことが条件として必要になる. 本報告では, ① によって導かれた評価関数から騒音検出マイクロホンの出力を印加した 2 次系と 2 次系フィルタの両出力の符号(±)の一致, ② からは騒音検出マイクロホンの出力を 2 次系フィルタと騒音制御フィルタの縦続接続に印加して得られた出力と誤差マイクロホンに達する 2 次騒音の出力との符号の一致が騒音制御を安定して実行できるための条件となることを示す. また, ② の条件が成り立つ場合には, 両者の符号の一致率は同程度となることを示す.

2-5-1

2-5-1 クラシックおよびジャズ楽曲の自動抽出されたサビ部分を利用したBGMによる作業効率向上

Work efficiency improvement by BGM utilizing automatically extracted chorus parts of classical and jazz music

○星野博之、△堂國雅揮、△的場雄太(愛知工業大)

◆本研究では、サビ部分が自動抽出できる「能動的音楽鑑賞サービス Songle」、および、RWC 研究用音楽データベース中のクラシック 40 曲、ジャズ 5 曲を用いて、サビ部分に注目した楽曲選択とサビ部分提示の作業効率向上における有効性調査を行った。

◆実験1: サビでのクラシック楽曲選択

- (1) サビ部分の好みに基づく選曲 クラシック 40 曲から被験者 10 名ごとに好みによる楽曲選択(好み・普通・好みでない)を行い、5 分間の暗算課題を行った結果(正答数)を Fig.1(a)に示す。
- (2) サビ部分の主要評価因子に基づく選曲 被験者 10 名による 14 個の形容詞対の主観評価結果の因子分析を行い、第 1 因子の最高・最低の楽曲を BGM として暗算課題を行った結果を Fig.1(b)に示す。

◆実験2: ジャズ楽曲サビ提示実験 ジャズ 5 曲からサビ部分およびサビ部分以外を平均 22 秒の長さで抽出し、被験者 6 名ずつがそれぞれの提示下(20 秒間の無音と交互提示)で暗算課題を行った結果、5 曲ともサビ部分を BGM とした方が正答数が多くなった。

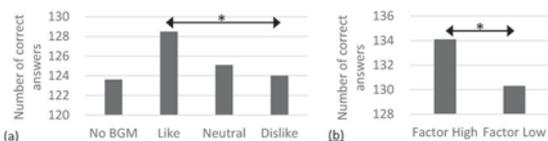


Fig.1 Results of arithmetic task (Number of correct answers, *: $p < 0.05$)

2-5-3

2-5-3 音楽聴取時の鳥肌感(Musical Chills)を効果的に生じさせるための部分的な音量増幅の検討

Investigation of partial volume amplification to effectively generate musical chills

○横山彩乃、岩城護(新潟大院)

- ◆音楽鑑賞時に鳥肌感(Pleasurable Musical Chills)が生起する場合がある。
- ◆我々は先の実験で鳥肌感が効果的に生起するための音量増幅量には限界があり音量増幅量が 3 dB 程度までが限界である可能性を示した。
- ◆本研究では最も効果的に鳥肌感を生起させる音量増幅量をさらに明らかにすることを目的とした。
- ◆音楽鑑賞時に鳥肌感を対象として、その楽曲を部分的に音量を加工して聴取実験を行った。楽曲印象(アンケート)、鳥肌感報告ボタン信号、生体信号(GSR, CSI)を計測した。
- ◆楽曲印象の傾向が異なる対象者では鳥肌感の生起と関連づけられる音量増幅量が異なる可能性が示された。

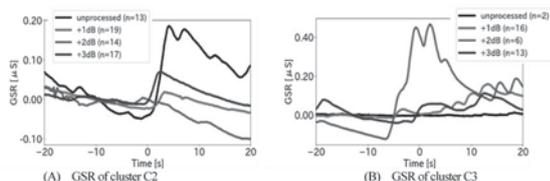


Fig.1: GSRs from two clusters of participants (The response from participants is synchronized at time zero. Each GSR is normalized so that the average before time zero should be zero.)

2-5-2

2-5-2 感性工学的手法による和音的多層構造モデルの構築

Harmonic hierarchical structure model based on Kansei Engineering

☆坂本櫻、山崎陽一(関西学院大)、片平建史(早稲田大)、長田典子(関西学院大)

◆和音印象の定量化に向けた感性工学的手法に基づく検討

- 和音の響きの構成要因が複雑な様相を持つことを明らかにした
 - ◇感情層: 快, 覚醒
 - ◇高次印象層: 明るさ, 陰気, 美的, 淡泊, 刺激
 - ◇低次印象層: 協和(-), 強さ(-), 重厚
- 和音を構成する要因を説明する物理要因を明らかにした
 - ◇ピッチの構造に基づく音響特徴量, 音質評価指標, スペクトルの帯域的情報, アタック区間の音響特徴量

◆重回帰分析による和音の響きを推定するモデルの構築

- 構成要因すべてにおいて説明力の高い($R^2 > .7$)モデルが得られた
 - ◇従来の研究で用いられていたピッチ構造に基づく音響特徴量に他の音響特徴量を加えることで, 和音の響き全体を網羅
- 構築したモデルにより, 協和度の順位を高い精度で推定
 - ◇協和の心理評価: Major>Sus4>Minor≒Dim>Aug
 - ◇本モデル: Major>Sus4>Dim>Minor>Aug

Table 1 Comparison of estimated value (Consonance)

	Evaluation value	Estimated value	
		C&F (2006)	This Model
Major	1	1	1
Minor	3	1	4
Dim	4	5	3
Aug	5	3	5
Sus4	2	4	2

2-5-4

2-5-4 BGM に対する態度に関わる要因について—女子大学生に対する質問紙調査より—

On the factors of the attitudes toward Background Music: A questionnaire study for female university students

○山崎晃男(大阪樟蔭女子大学)

- ◆女子大学生 168 名を対象に、BGM に対する態度に関する質問紙調査を実施し、その態度に関わる要因について検討を行った。その際、音楽のコントロール可能性に注目するとともに、聴覚過敏性やビッグファイブ性格検査によって測定される性格特性と BGM に対する態度との関係についても検討した。
- ◆因子分析の結果、BGM に対する態度に関する 3 つの因子を得、それぞれを「他者コントロール BGM に対する否定的態度」「自己コントロール BGM に対する肯定的態度」「BGM による注意阻害」と解した。
- ◆他者コントロール BGM に対する態度では回答者の 88.1% が肯定的であり、自己コントロール BGM に対する態度では 76.8% が肯定的と、BGM に対して多数が肯定的であった。その一方、BGM によって注意が阻害されると答えた者も 59.5% いた。
- ◆各因子得点と聴覚過敏度および 5 つの性格特性の得点との相関係数を見たところ、「他者コントロール BGM に対する否定的態度」と聴覚過敏度との間に中くらいの、「BGM による注意阻害」と聴覚過敏度との間に弱い正の相関が示された($r = .548, p < .01; r = .370, p < .01$)一方、聴覚過敏度と「自己コントロール BGM に対する肯定的態度」との間には相関が見られなかった。性格特性に関して、「他者コントロール BGM に対する否定的態度」と協調性との間に弱い負の相関があり($r = -.285, p < .01$)、協調性の高い者ほど他者がコントロールする BGM に対して肯定的な態度を有していた。

2-5-5

2-5-5 能楽師とオペラ歌手による歌唱の空間放射特性

Spatial radiation characteristics of singing by Noh and Opera singers

○木谷俊介(北陸先端大), 牧勝弘(愛知淑徳大), 饗庭絵里子(電通大), △中内茂樹(豊橋技科大)

- ◆本稿では能楽師とオペラ歌手の歌唱放射特性に違いがあるのかどうかを明らかにすることを目的とする。
- ◆42 ch 球状マイクロホンアレイを用いて発話者の歌唱音を収録した。
- ◆各マイクロホンで収録した音声信号をベクトル演算することで放射特性を求めた。
- ◆能楽師とオペラ歌手の特徴を示すために、素人の放射特性と比較した。
- ◆その結果 (Fig. 1)、能楽師の歌唱は低周波数帯が下方向に、オペラ歌手の歌唱は高周波数帯が上方向に放射していることが示された。

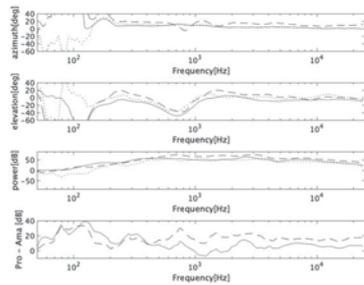


Fig. 1 Spatial radiation characteristics of singing. Horizontal axis shows frequency, vertical axis shows from the top, horizontal direction, vertical direction, power, and the difference between professionals and amateurs. The red line is for Noh singers, the blue line is for opera singers, and the black line is for amateurs.

2-5-7

2-5-7 歌声による歌唱中の体内運動の推定

Estimation of articulatory gestures during singing by singing voice

◎高橋純(大阪芸大), 志々目樹, 戸田菜月, 竹本浩典(千葉工大)

- ◆歌唱中の体内運動は外部から観察できないが、指導者は学習者の歌声から体内運動を推定して歌唱指導をおこなっている。
- ◆本研究では、プロの歌手(Pros), プロの器楽奏者(Prol), 学生(AmaS)が、1年間の歌唱指導を経た声楽未経験者の歌声から、歌声の評価と歌唱中の声道形状をどの程度推定できるのかを検討した。
- ◆歌唱指導によって歌唱中の声道形状は変化し、訓練前に比べて訓練後は口唇が開き、口腔と咽頭腔が拡大して喉頭が下降した (Fig. 1)。
- ◆ProSは訓練前後の声道形状を詳細に推定できることが示された。Prolは訓練前後の歌声の変化は評価できるが、声道形状の変化はプロほど詳細に推定していなかった。AmaSは歌声自体の評価ができず、それに伴って声道形状の変化の推定もできていないことが示された (Table 1)。

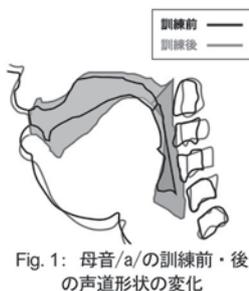


Table 1 母音/a/における訓練前・後の声道各部のピクセル数と評価スコア

	部位	Pixel	ProS	Prol	AmaS
訓練前	口唇	11	2.4	3.0	3.5
	口腔	1298	1.8	3.0	2.7
	/a/ 咽頭腔	432	1.7	1.5	3.2
	喉頭	68	1.7	3.2	3.2
訓練後	口唇	14	3.9	4.5	3.2
	口腔	2183	4.0	4.3	3.3
	/a/ 咽頭腔	985	4.2	3.8	3.5
	喉頭	87	3.8	3.7	3.2

Fig. 1: 母音/a/の訓練前・後の声道形状の変化

2-5-6

2-5-6 歌唱時における緊張と弛緩の可視化手法の検討

Study of Visualization Methods for Tension and Relaxation during Singing

☆久保寺 早紀(国立音楽大学大学院), 三浦 雅展(国立音楽大学)

- ◆歌唱時において歌唱者は自身の身体を駆使して演奏を行なう。その際に身体に力を入れたり、脱力を行ったりして演奏を行なうが、この時に過度な緊張もしくは脱力をしてしまうと演奏技術に支障が出ることもある為、適度な力加減が必要とされている。
- ◆本報告では、伸縮センサを用いて歌唱時の緊張及び弛緩に着目した可視化手法の検討を行なう。伸縮センサ装着図を図1に示す。
- ◆実験方法
測定箇所: 胸部, 腹部 演奏者: 本大学院声楽専攻修士課程3名
実験条件: 脱力した状態 (Relax), 最大限力んだ状態 (Power), 一番歌いやすい状態 (Free)
演奏課題: G.Concone 作曲『コンコーネ 50 番練習曲 No.5』
T.Giordani 作曲『Caro mio ben (邦訳: いとしい女よ)』

- ◆息を吸うタイミングで区切られた領域のセンサ信号から、胸部および腹部それぞれについて最大値を取得し、演奏者毎に実験条件間の比較と胸部と腹部の最大拡張の合計に着目して多重比較 (Tukey's HSD 法) を行なった。図2に検定結果の一例を示す。結果 S02 の R 条件と P 条件における胸部の伸縮 R 条件と、S03 の P 条件と F 条件における腹部に有意差が認められた。



図1 装着図

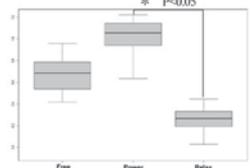


図2 多重比較の結果の一例

2-5-8

2-5-8 生演奏における演奏者と鑑賞者の相互作用

Performer-and-audience interaction in public music performance

○正田 悠(京都市立芸大)

- ◆音楽の生演奏は、演奏者から鑑賞者に向けて演奏が伝達されるとともに、鑑賞者の存在が演奏に影響を及ぼす双方向的なプロセスである。
- ◆演奏者と鑑賞者がともに同じ空間で同じ時間を共有することの効果調べるために印象評定と自律神経系の計測から実験を行なった。



Fig.1: Schematic picture of performer-and-audience interaction

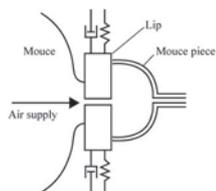
- ◆演奏の感情的ニュアンスに関する印象評定では、録音演奏よりも、生演奏において、演奏者の評定により近くなった。
- ◆「速いテンポの曲では心拍数が速くなる」というような演奏テンポへの生体テンポの引き込みは、生演奏条件でのみ認められ、録音演奏を聴取る条件では認められなかった。
- ◆生演奏において楽曲の特徴に引き込まれることにより、鑑賞者は演奏により没頭することができ、その結果、複雑な演奏のニュアンスを受け取ることができる可能性がある。

2-5-9

2-5-9 金管楽器の自励振動に対する機械振動フィードバックのシミュレーション

Regarding self-excited vibration of brass instruments Simulation of mechanical vibration feedback

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



金管楽器の吹鳴は口唇振動, 気柱振動, 楽器の機械振動からなる自励振動系 → 自励振動系に対する機械振動のフィードバックは検討が不十分

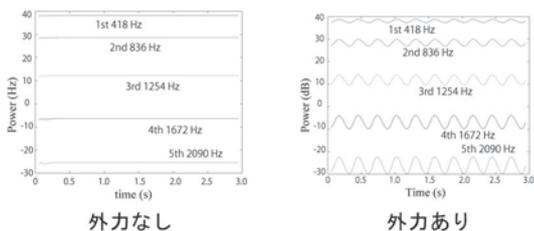
目的 シミュレーション上の金管楽器吹鳴における機械振動の影響の評価

方法

垂直振動モデルによる金管楽器シミュレーションを構成. 定常吹鳴時に 4.5 Hz 大きい正弦波の外力を口唇に与え, マウスピース内音圧の各モードのパワーの時間変化をみる.

結果・考察

マウスピース内音圧の各モードは 4.5Hz 毎に周期的な変化. 恣意的に与えた口唇運動の位相変化によって運動を助長, 阻害する効果が周期的に現れたと考察.



2-5-11

2-5-11 モーションキャプチャを用いたグランドピアノ演奏における動作解析について*

*Motion Analysis of Grand Piano Performance Using Motion Capture

○三戸勇気(日大・芸術), 塚田湧太, 柴田直樹, 山口達也(日大・理工), 駒澤大介, 川上央(日大・芸術), 篠田之孝(日大・理工)

筆者らは光学式モーションキャプチャシステムを用いた感情表現による鍵盤楽器の演奏動作の計測並びに動作解析を行ってきた. 従来の計測では上半身のみに反射体であるマーカを取り付けていたが, 今回, 下半身の動きも計測するため, 全身にマーカを取り付け, 演奏時間が1分程度の曲を用いて実験を行った. 本文はプロピアニストのグランドピアノの演奏動作における身体各部の関節部位などの軌跡長を算出し, グランドピアノの感情表現における演奏動作の解析の検討を行った.

今回, 演奏時間が1分程度の曲を用いて全身に取り付けたマーカ情報によるグランドピアノの演奏動作の解析を行った. 今後, 曲の時間的な流れと感情表現による演奏動作の関連について検討していく予定である.

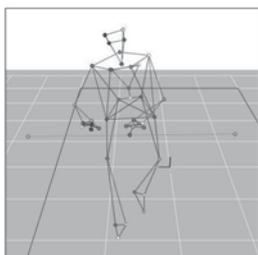


Fig.1 Marker adhesion position

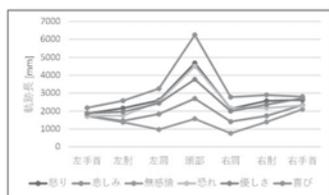


Fig.2 Trajectory length in X-axis

2-5-10

2-5-10 Analyzing posture of kotsuzumi players, hitting motion, and sound production

Saha Chaity, Tomohisa Funakura, Akihiro Matsutani, Marie Tabaru (Tokyo Institute of Technology)

- ◆ The Kotsuzumi, a renowned Japanese drum, is central in Noh and Kabuki theatre, and while its structure and sound are well-studied, its playing techniques remain largely unexplored.
- ◆ Using Deep Learning, skills of professional kotsuzumi players were deeply analyzed in our study through skeleton analysis, providing insights into their sound.
- ◆ Key parameters such as front backbone angle, hand slope, body balance, and kotsuzumi hitting angle [Fig.1(a)] as well as the hitting motion [Fig.1(b)] and audio properties like loudness, RMS, amplitude, rise & fall time, timbre [Fig.1(c) & Fig.1(d)] were analyzed in our study.
- ◆ Based on these, a user-centric GUI platform was developed, and algorithms were built into this system.
- ◆ Using this system, players compare techniques, get session-end feedback, and integrate tradition with technology.

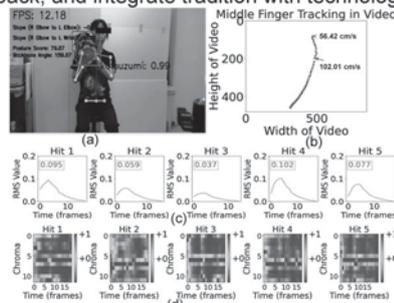


Fig.1: (a) Assessing Performance (b) Measuring hitting motion (c,d) Sound Feature Extraction (timbre, loudness).

2-5-12

演奏芸術医学と基礎科学

—演奏動作のバイオメカニクス—

Performing Arts Medicine and Basic Science — Biomechanics of the Music Performance —

○酒井直隆(医療法人アーツメディック さかい整形外科)

演奏芸術医学 Performing Arts Medicine は音楽家とダンサーを対象とした新しい医学分野である. 演者は 1980 年代からこの新分野の医学に取り組み, 臨床現場の疑問を基礎科学のバイオメカニクス研究で検討し, 臨床に応用してきた. これまで治療した自験例 6,320 例の音楽家のうち, 手のオーバークース障害 2,653 例の内訳は腱鞘炎・付着部炎・筋肉痛の, 筋腱の疼痛が 60% を占めた. またピアニストの 78% がオクターヴと和音の練習時に発症していた. これらのテクニックの動作解析を行ったところ, 手の小さいピアニストが母指をより大きく外転させており, 母指にかかわる筋腱の疼痛が多い事実を裏付けていた. さらに音楽家の手の X線像で手の大きさを計測しオーバークース障害の有無で比較すると, 障害群のピアニストの hand span や母指・中指・小指の長さが, オーバークース障害を経験しないピアニスト群より短かった.



Fig. 1 Hand motion analysis with reflective markers

2-6-1

デフスポーツにおける支援技術 ～競技特性・ルールと技術との関係～ Assistive Technology in Deaf Sports

○白石優旗(筑波技術大学), △設楽明寿(筑波大学)

デフスポーツは、ろう・難聴者のみが参加するスポーツのことを指し、聴覚的な合図に代えて視覚的な合図を使用する点を除き、一般的なスポーツと同様に行われている。また、ろう・難聴者がスポーツに参加する際の支援技術については、数は多くないものの、研究開発がなされている。しかしながら、代替技術としての支援技術がほとんどであり、各スポーツのルールや場面に特化した技術とは何かを考慮して開発された事例は我々が知る限り存在しない。そこで、我々は陸上競技短距離走のルールと使用場面に特化した方法として、スタート合図を触覚情報で直接提示するシステム“HaptStarter”を2016年から開発している。今回はデフスポーツにおける支援技術を複数の事例を取り上げるとともに、競技特性、ルール、技術との関係に着目しつつデフスポーツとはなにかを聴講者一同と共に考えていきたい。

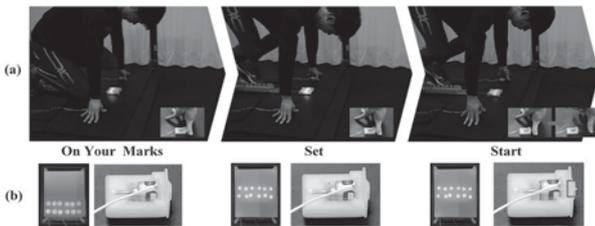


Fig.1:HaptStarter's workflow. (Left to "On Your Marks", "Set", "Start".)

2-6-3

2-6-3 競技者への応援に関する研究 —観戦者が発する声掛けに着目して— Cheering's impact on athlete performance -Analyzing spectator support-

○久永将太(慶應大・総), 浅野恵子(順天堂大・医)

- ◆ 応援(声援)の意味は、競技・試合などで声(心の声も含む)による応援をし、選手やチームを元気づけることを指す。
- ◆ 応援に関する「声掛け」に焦点を当て、アンケート調査を実施した。
- ◆ 大学でスポーツ関連科目の授業を履修する168名(18-21歳の男性48名、女性120名)に実施した。
- ◆ 各場面の声援の送り方として、「声を出して」、「小声で呟くように」、「声には出さず心の中で」といった主に3種類の声援のスタイルが存在していることが分かった。
- ◆ 観戦者はいつも同じような応援の様式で声を掛けているのではなく、シーンに応じて用いる声援の種類や、声掛けの内容を切り替えていた。
- ◆ 応援行動を行うことで「感情が高ぶる」との回答が多くみられた。

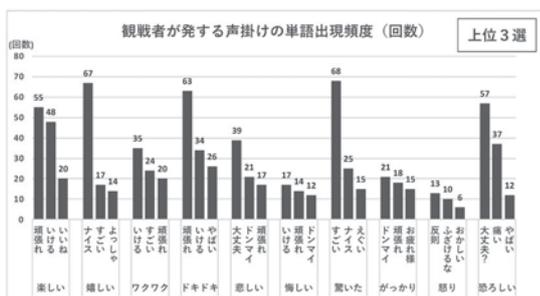


Fig.1:Focusing on the voices of spectators.

2-6-2

2-6-2 (招待講演) 聴覚情報を活用した視覚障害 スポーツ/eスポーツの競技力向上

[Invited talk] Improving the competitiveness of sports/e-sports for visually impaired players by utilizing auditory information

○三浦 貴大(産総研), △松尾 政輝(筑波技大), 上田 麻理(神奈川工科大)

- ◆ 本講演では、視覚障害のある競技者が行うスポーツ/eスポーツにおける状況を紹介の上で、分野における展望を述べる。
- ◆ 本稿の内容は、以下の解説記事・総説論文を基にしている：
 - 三浦, 視覚障害者が音で見る世界—感覚代行と支援機器—, 日本音響学会誌, 77(3):199-207, 2021.
 - Miura, T., Yabu, K., Narrative Review of Assistive Technologies and Sensory Substitution in People With Visual and Hearing Impairment, Psychologia, 65(1):70-99, 2023.
 - 三浦, 松尾, 坂尻, 大西, ゴールボールにおける投球音の定位能力を訓練するアプリケーション, 音響技術, 51(197):36-41, 2022.

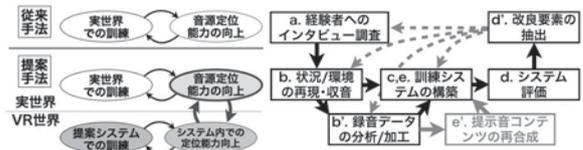


Fig.1:(Left) Auditory training scheme for visually impaired sports using acoustic VR technology, (Right) Process of studies to develop auditory training systems

2-6-4

2-6-4 スマートグラスを用いた深層学習に基づく 競技場の音の認識

Stadium Sound Recognition Based on Deep Learning Using Smart Glass

☆森岡龍健, 西田昌史, △綱川隆司(静岡大)

現在、娯楽の観点からの聴覚障がい者支援に関する研究が必要とされており、スポーツ観戦支援に関する研究が行われている。本研究ではスマートグラスを用いた競技場の音の認識を改善する手法として、従来の機械学習手法に加えて、深層学習の導入や特徴量としてメルフィルタバンクを使用することを新たに提案する。

本研究では、認識手法としてCatboost、LSTM、GRU、特徴量としてMFCCとメルフィルタバンク(MFB)を用いて、実際に二つの競技場で録音したデータを対象に単一の競技場と異なる競技場間で交差検証による評価実験を行った。実験結果(Table 1)より、浜松アリーナで学習してエコパアリーナで認識を行った場合には、提案手法であるMFBを用いた場合のCatboostが最も優れていることが示された。また、MFCCを用いた場合には、LSTMやGRUがCatboostよりも優れていることが分かった。そして、どの手法も実時間処理が可能であることが示された。

Table 1: Recognition results trained in Hamamatsu Arena and recognized in Ecopa Arena

	特徴量	Catboost	LSTM	GRU
F値	MFCC	0.50	0.60	0.59
	MFB	0.64	0.56	0.56
RTF	MFCC	0.068	0.174	0.170
	MFB	0.086	0.314	0.378

2-6-5

2-6-5 野球のバッティング練習時の環境音の影響 その2 -映像からの動作分析-

Effects of environmental sounds on baseball batting practice
- motion analysis from video -

○春日秀雄, △新田晃司, 田中博, 上田麻理(神奈川工科大)

◆音楽は身体パフォーマンスの向上に寄与するとされており、スポーツ選手にとって欠かせないものとなっている。

◆環境音の違いによって選手のバッティング動作に影響が出るか、姿勢推定ツールのOpenPoseを使用して調査を行う。

➢ スイング動作における体軸の傾きの変化を調査

✦ 体軸のブレでスイングのよし悪しを判断できる



Fig.1: Tilt of body axis during swing

➢ 3つの環境(アップテンポな「普段流している音楽」、スローテンポな「クラシック」、「無音」)で違いを調査

◆「普段流している音楽」では「無音」と比べてスイングの後半に体が上を向くという傾向が見られた

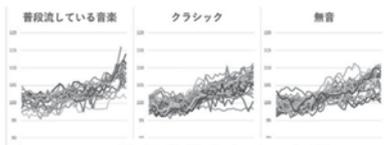


Fig.2: Shifts in body axis tilt

2-6-7

2-6-7 背面式配置を用いたトランスオーラル再生方式に関する一検討

A Two-Channel Transaural Reproduction System with Back-to-Back Speaker Drivers

○藤井優美, 村山好孝, 高岡仙(シーイヤー), 渡邊祐子(東京電機大・デザエ)

◆トランスオーラル再生方式の一手法として、2つの小型スピーカを背中合わせに配置したBack to Back Transaural (BBT) 再生方式を提案する。

◆従来の研究では理想的な点音源を仮定してきたが、実用上はスピーカの指向特性を考慮する必要がある。そこで、スピーカを横向きに配置することでBinaural 音場の堅牢性が向上できると仮説を立て、その基礎検討として、トランスオーラル再現音場の物理特性の違いや堅牢性をステレオダイポール(SD)方式と比較した。

◆伝達特性を比較すると、BBT方式はクロストーク成分の減衰に特徴があり、低域を含めて安定した逆行列を生成した。さらに、異なる受聴位置での実験により、BBT方式がSD方式と同等以上の性能を持ち、前後・左右方向の受聴位置の変化にも対応できることが確認された。

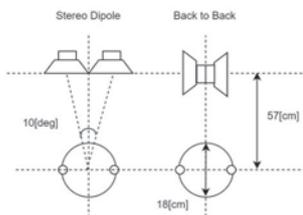


Fig. 2 スピーカの配置

2-6-6

2-6-6 バレーボール競技における聴覚訓練 アプリケーション構築の試み

An Attempt to Construct an Application for Hearing Training
in Volleyball Competition

☆高橋隼人(神奈川工科大), 三浦貴大(産総研), 田中哲雄(神奈川工科大), 山口雄大(日体大), 上田麻理(神奈川工科大)

◆バレーボール競技においても選手は様々な聴覚情報を利用して競技を行っている。本研究では、聴覚情報を活用したバレーボール競技力の向上手法を構築することを目的とし、競技中に音を正確に判別可能となるような聴覚訓練システムの開発を試みた。

◆本アプリケーションはプロフィール登録、メニュー、履歴参照の機能を提供する。現在は聴覚トレーニングと打球音の音質評価の2種類のメニューを開発した。Fig.1はトレーニングメニューの4問目の打球音を再生した後に表示される画面の例である。

◆本アプリケーションをFig.2に示すようにWebアプリケーションとして実装した。今後、日本体育大学バレーボール部員をはじめとした競技経験者を対象にログを収集していき、訓練効果について検討する。



Fig.1 : Menu Screen of the Application.

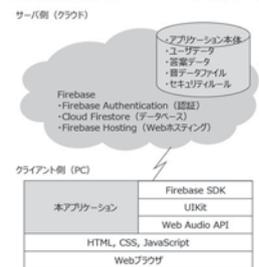


Fig.2: Application configurations.

2-6-8

2-6-8 オブジェクトベース音響のための音響メタデータ の座標変換手法

Coordinate conversions in audio metadata for object-based audio

◎岩崎泰士, 久保樹樹, 大出訓史(NHK)

◆オブジェクトベース音響の位置情報には極座標と直交座標が用いられるが、既存のレンダリング手法(勧告ITU-R BS.2127)では、5.1.4のスピーカ配置に基づいて座標変換((-1,-1.0)を110度へ変換)を行うため、22.2ch音響や7.1.4のスピーカ配置において規定((-1,-1.0)を135度へ変換)通りに座標変換されない。今回、任意のスピーカ配置に対し、規定通りに極座標と直交座標を変換可能な手法を提案した。

◆22.2ch音響のスピーカ配置の中層の各スピーカ位置について、極座標と直交座標の座標変換を行った結果、提案法では全スピーカにおいて規定通りに変換できることを確認した。

◆主観評価実験により、13地点について、変換前の位置情報と変換後(提案法と既存法)の位置情報を用いてレンダリングした音を比較し、提案法が既存法よりも高く評価されることを示した(Fig.1)。

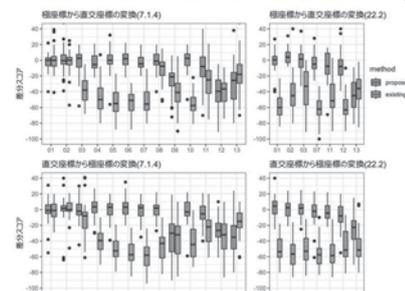


Fig.1: Results for subjective test

2-6-9

2-6-9 ステレオ音源のチャンネル間相互相関の周波数依存性

Frequency dependency of inter-channel cross-correlation for stereo sources

○水町光徳(九工大), 堀内俊治(KDDI 総合研)

- ◆ヘッドホンやイヤホンでのステレオ音源再生のために、左右チャンネル信号を加算した簡易クロストーク成分を、適切な割合でステレオ原信号に加算するステレオ音像幅縮退方式を提案した。
- ◆クロストーク成分の適切な加算割合を検討するために、各周波数帯域におけるステレオ音像幅縮退音源のチャンネル間相関値を調査する。
- ◆Fig. 1に、大編成のオーケストラ音源(左図)と弦楽四重奏音源(右図)のサブバンドチャンネル間相関値を示す。
- ◆各音源の音源数や音源配置に応じて、周波数帯域ごとのチャンネル間相互相関が多様に変化することを確認した。

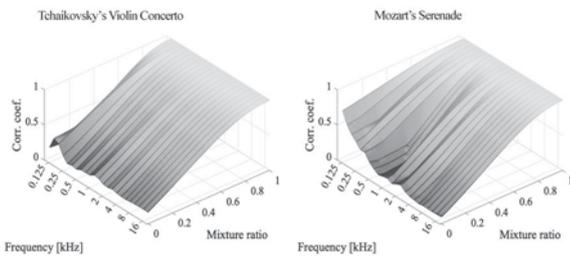


Figure 1: Inter-channel sub-band cross-correlation for Tchaikovsky's Violin Concerto in D major, op. 35 and Mozart's Serenade in G Major, K525, Eine Kleine Nachtmusik in left and right panels, respectively.

2-6-11

2-6-11 行為としてのオーディオ受聴を 実現するための BoSC システムの検討

Study on the relationship between behavior and sound impression during audio listening using BoSC system

☆西尾弘大, 伊勢史郎(東京電機大)

- ◆本稿では、スマートフォンやサブスクリプションサービスの普及によりもはや忘れ去られようとしているアナログレコードに針を落とすような喜びを体験できるようなオーディオ受聴のユーザーインターフェースを実現するためのシステム構成を検討した。
- ◆通常のオーディオルームと同様に音楽作品を聴く行為を再現するには畳み込みをリアルタイムで行う必要があることからリアルタイムオーディオルームシミュレータを作成し、従来の BoSC システムとの間で聴覚比較実験を行った。
- ◆被験者自身にレコードプレーヤーを操作させオーディオ受聴行為を体験させた。その結果、レコードプレーヤーで音楽を聴く場合には主体的に音楽に関わる心理的状況が生じる可能性が見出された。



Fig. 1: Operating the record player

2-6-10

2-6-10 HatmanDPS 技術による 鉄道車両内アナウンス改善効果の検証

The effect of improving in-train announcements with HatmanDPS technology

○鈴木理雅(TOA エンジニアリング), 中家諒(TOA),

△濱田一平(EFFECTORNICS ENGINEERING)

- ◆“音像感のコントロールにより聴こえ方の印象変更を目的としたアナログ電子回路の音声処理技術”「HatmanDPS」を用いた鉄道車両内アナウンスの改善を目指し、この技術の効果検証のため鉄道車両環境を模した高騒音環境において聴取実験を実施した。
- ◆聴取対象者は64~79歳の高齢者とし、Fig.1に示す配置で「音声」と「疑似騒音提示条件はTable1に示すように計12通り。
- ◆結果、いずれの条件においてもHatmanDPS「OFF」と比べて「ON」の場合の方が高い正答率となり、一部条件においては有意な差をもって効果があることが示された。(Fig.2)

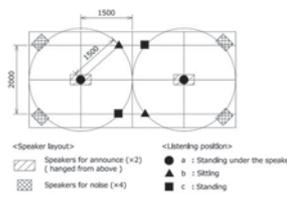


Fig.1 Layout

DPS	Listening position	Speaker	= 12
ON	a	male	
OFF	b	female	
	c		

Table1 Condition combination

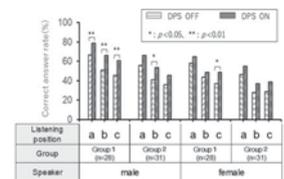


Fig.2 Word intelligibility

2-6-12

ビジョントランスフォーマーと知識蒸留を 用いた環境音分類の評価

Evaluation of Environmental Sound Classification using Vision Transformer and Knowledge Distillation

○王昌竜, 伊藤彰則, 能勢隆(東北大)

- ◆事前学習された Vision Transformer は環境音の分類で優れた結果を示しているが、事前学習を行わない場合においては、クラシックな手法と比べて優位性が見られない。
- ◆環境音分類において、現場で十分なデータを収集することが難しいため、Vision Transformer を画像データで事前学習するという手法が有効である。
- ◆データ拡張は手元のデータを加工し新しいデータセットを生成する手法である。データ拡張は学習に新しい知識を導入しないため、わずかに数パーセントの改善しかもたらさず、かつ学習には時間がかかる。
- ◆モデルサイズを減少させるために、学習済みの Vision Transformer を自己蒸留させると、若干の効果がある。ただし、知識蒸留は完璧に転移できないという制約がある。

2-6-13

2-6-13 3D オーディオにおける上層の知覚に必要な帯域について

On the bandwidth required for the perception of the top layer in 3D audio
○亀川徹 丸井淳史(東京芸大)

- ◆3D オーディオの上層の低域の周波数を制限して、オリジナルの音との違いが半断できなくなるカットオフ周波数について、PEST 方法を参考にした方法で、モノ音源や2チャンネルステレオ音源を上層のみから再生して実験をおこなった。
- ◆ピンクノイズやバーストノイズの場合は、上層の低域がわずかでもカットされるとその違いが半断できることがわかった。
- ◆実際の楽器やナレーションでは違いが区別できなくなるカットオフ周波数の第1四分位数は125Hz から1kHz と素材によって異なり、再生方向や素材の複雑さによっても変わってくることを示された。

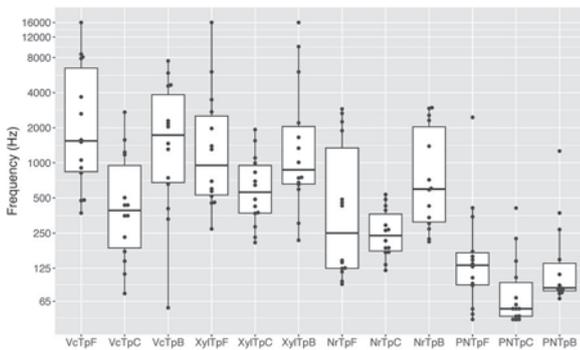


Fig.1: Boxplot of the response at frequencies where the difference from the original becomes indistinguishable when low frequencies of the top layer

2-7-2

2-7-2 FMBEM 入射指向性解析に基づく不整形残響室の拡散性評価

Evaluation of sound-field diffusion in an irregularly shaped reverberation room by FMBEM incidence directivity analysis

☆萩原諒(東大・工), 樹本貴之(サイバネット), 安田洋介(神奈川大・建築), 佐久間哲哉(東大・工)

- ◆既報では直方体室における定常音場を対象に、FMBEM 入射指向性解析に基づく拡散性指標を計算し、傾向などの基礎的検討を行った。
- ◆本報では、上記の拡散性指標の計算を新たに不整形室へ適用し、直方体室との比較を通じて室形状が指標に与える影響を分析した。
- ◆受音セルと音源の位置を固定した上で、室内の吸音率は均一とし、受音セルの大きさや吸音率などの条件を変えながら比較を行った。
- ◆不整形室における全体的な指標の周波数特性は、両指標ともに直方体室の場合と同様の傾向となった。
- ◆反射性の不整形室では平行面が存在せず、モードの方向の偏りが直方体室よりも少なくなるため、指標が全体的に直方体室の場合よりも大きくなり、特にオクターブバンドの指標で最も大きな値となった。
- ◆受音セルの大きさが異なる際に、直方体室では指標の値が異なるのに対して、不整形室では指標の違いが比較的小さく現れるなどの傾向の違いを確認した。

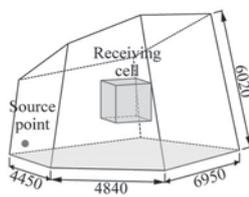


Fig.1: Analysis model.

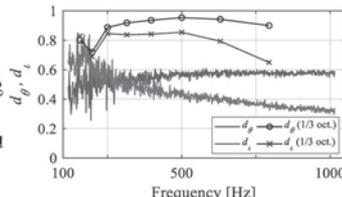


Fig.2: Diffusion indices.

2-7-1

2-7-1 有限個のランダムな平面波の重ね合わせによる局所的な領域的拡散音場

Locally diffuse sound fields generated by superimposing a finite number of random plane waves

○田中 達宏, 大谷 真(京都大院・工学研)

背景 音場の拡散性を定量的に評価するための理論的な基盤(完全ではない拡散音場のモデル)を構築したい。

目的 球面調和領域で定式化された有限次数の拡散性が、領域的拡散性とどのように関係するのかを明らかにする。

方法 有限次数の拡散音場(有限平面波モデル)で観測される音圧の統計的性質を調べる。

結果 自由音場に関する理論的検討により、局所的な領域的拡散性の成立が導かれた。数値実験では、拡散性の概念を拡張して、自由ではない音場に関して検討し、やはり局所的な領域的拡散性が成立することが分かった(Fig.1, A-C)。

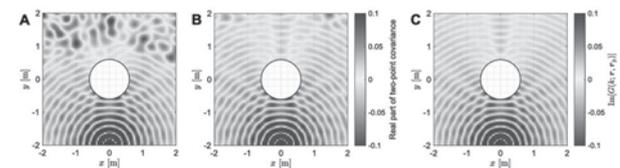


Figure 1. Real part of calculated two-point covariance between $r_p=(0,-2,0)^T$ and positions on an $x-y$ plane in discrete random-wave models of select diffuseness degrees: (A) $L=24$ and (B) 32. (C) Imaginary part of a Green's function for a point source at r_p .

2-7-3

2-7-3 交差リブ拡散体の実施例

- オフィス個室ブースの室内音響とアンケート結果の対応 -

Example of Crossed Rib Diffuser

- Correlation between room acoustics in office private booths and survey results -

○侯野祐美, △碓美友, △栢野旭代(大建工業), 中津成博, 上田泰孝(安藤ハザマ技研)

- ◆オフィスの内装木質化と室内音響調整を同時に実現するパネルとして、木製の交差リブ拡散体*を作製し、個室ブースに設置、オフィス利用者にアンケート調査をおこなったところ、「音環境を改善する効果」と「見た目の印象を向上させる効果」を確認した。
*交差リブ拡散体 (Crossed Rib Diffuser, 以下「CRD」)
- ◆本稿では、調査に用いた Fig. 1 のような個室ブース内の音響特性を測定することで、アンケート結果との対応を確認した。
- ◆インパルス応答測定結果から、各種室内音響指標、インパルス応答波形の観察、自己相関関数の算出、周波数特性の分析をおこなった結果、CRD 拡散パネル、および CRD 拡散+吸音パネルが音環境に与える影響を考察したため報告する。

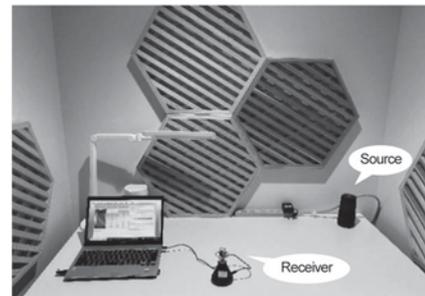


Fig. 1 : Image of Room B (CRD diffusion type).

2-7-4

2-7-4 パッチ型吸音構造の吸音原理に関する研究

Sound absorbing mechanisms of Patch-type sound absorbers
○真田 明, △藤本望夢(岡山工技セ), △小野裕行(三乗工業),
△菊田久雄(大阪公立大学)

- ◆著者らは、これまでに多孔質吸音材料の表面に非通気のパッチを貼り付けることにより、薄い部材で低周波数から比較的広い帯域で吸音効果が得られるパッチ型吸音構造を提案している。本研究では、パッチ型吸音構造の吸音原理を明らかにすることを目的に、1次元音場を仮定した簡易モデルにより垂直入射吸音率を求め、FEMによる計算結果と比較した。
- ◆低周波数域における吸収ピークは傾向がよく一致したが、高周波数域における吸収ピークは異なった。
- ◆高周波数域における吸収ピークの計算結果の乖離の原因を解明するために、吸音材料中の各部位における吸収パワーを求めた。これにより、吸音原理を明らかにした。

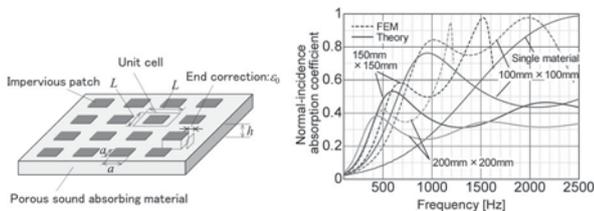


Fig. 1: Patch-type sound absorber

Fig. 2: Calculated normal-incidence absorption coefficients of the patch-type absorbers.

2-7-6

2-7-6 スリットを有する微細周期構造の幾何特性が吸音特性に与える影響に関する検討

Study on parameters and sound absorption properties of micro-periodic structures consisting of cubic unit lattice with slits
☆ 城修平(近畿大院), 菅原彬子, 平栗靖浩(近畿大)

- ◆音響メタマテリアルとして、単位格子にスリットを有する微細周期構造を設計し、モデルを作成した(Fig.1)。
- ◆モデルに対して、吸音特性と密接に関わるとされている迷路度、静的粘性迷路度、流れ抵抗の3つの輸送パラメータ(JCALP model)に着目し、吸音特性との関係を数値解析より検討する。
- ◆解析結果より、輸送パラメータの変化により吸音特性に特徴が与えられることが確認できた。流れ抵抗に着目すると、値の増加に伴って低周波数帯域で吸音率が向上する傾向が見られ、広帯域化を実現する可能性が示唆された(Fig. 2,3)。

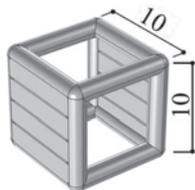


Fig. 1 Analysis model

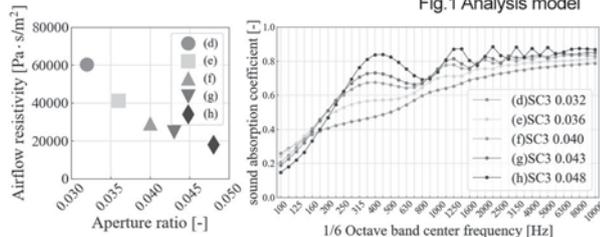


Fig. 2 Relationship between flow resistance and aperture ratio

Fig. 3 Absorption coefficients of five cases with different aperture ratio

2-7-5

2-7-5 音響メタサーフェスの吸音特性におけるユニットセルの最適化設計に係る考察

A study on the optimization design of unit cells in the sound absorption characteristics of acoustic metasurfaces
○中西伸介(広島工業大・工)

- ◆本報では、音響メタサーフェス(AMS)の吸音特性を1オクターブバンドで完全吸音にするユニットセルの個数割合を決定する手法を提案し、計算例を以て考察する。
- ◆また、計算条件においてユニットセルとするヘルムホルツ共鳴器のネック部の設計について紹介する。
- ◆ユニットセルの表面は12 mm角の正方形で、一辺が2.2 mmまたは3.2 mmの正方形開口を有する。
- ◆中心周波数を125 Hz, 250 Hz, 500 Hzとする1オクターブバンドでほぼ完全吸音になる計算例を示した。
- ◆このとき、ユニットセルの面積割合を考慮した音響アドミタンス比の実部が互いに等しくなるように面積割合を決定する関数を提案した。

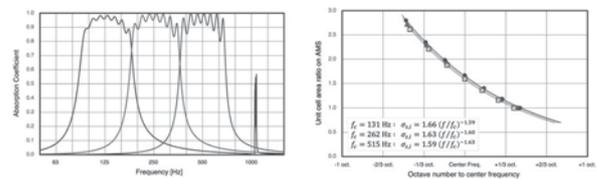


Fig. 1: (a) A nearly perfect sound absorption in one octave band by AMS consisted of unit cells having square opening holes of 3.2 mm side. (b) Unit cell area ratio on AMS for nearly perfect sound absorption.

2-7-7

2-7-7 周期配置スリット共鳴器の吸音率予測における有限要素モデルの比較

Comparison of Finite Element Models for Predicting Sound Absorption Coefficients of Periodically Arranged Slit Resonators
☆米澤 美桜, 井上 尚久(九大・芸工)

本研究では、周期配列スリット共鳴器の吸音率解析を取り扱う。共鳴器のエネルギー吸収に関して、線形化ナビエ-ストークス方程式を使った厳密性の高いモデル (VTA model)、粘性・熱伝導境界層の影響をヘルムホルツ方程式の境界条件として与えるモデル (BC model) の二つを取り上げ、単純なスリット共鳴器、スリットを内側に延長した共鳴器の吸音率を有限要素法によりそれぞれ算出した。解析結果から、遷移周波数より低域ではBC model で正確な結果が得られないことが示された。

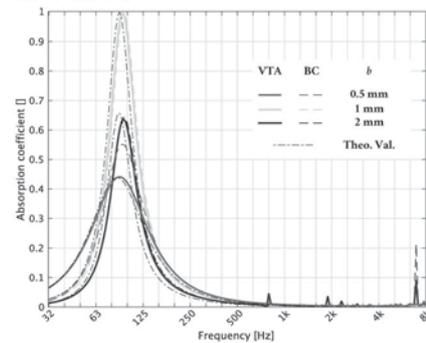


Fig. 1: Normal incidence absorption coefficients calculated by using VTA and BC models for different combinations of neck width b and cavity depth D_b .

2-7-8

2-7-8 モードマッチング法を用いた共鳴器型吸音構造の吸音特性予測

Sound Absorption Characteristics Estimation of a Resonator Using Mode-Matching Method

◎岩永尚文, 本地由和(ヤマハ)

- ◆ヘルムホルツ共鳴器等の共鳴器型吸音構造の吸音特性を予測したい。
- ◆共鳴器ネック部をキャビティ内に延長したり、多段化したり、多孔質材と組み合わせたりすることで広帯域な特性を実現したい。
- ◆このようなケースに適した開口端補正が不明なため、伝達マトリクス法(TMM)のような平面波音場を仮定した手法では特性予測が難しい。
- ◆導波路の波動解析に用いられるモードマッチング法(MMM)を多層の共鳴器型吸音構造の解析に適用した。
- ◆有限要素法(FEM)を参照解としたシミュレーション結果を示す。

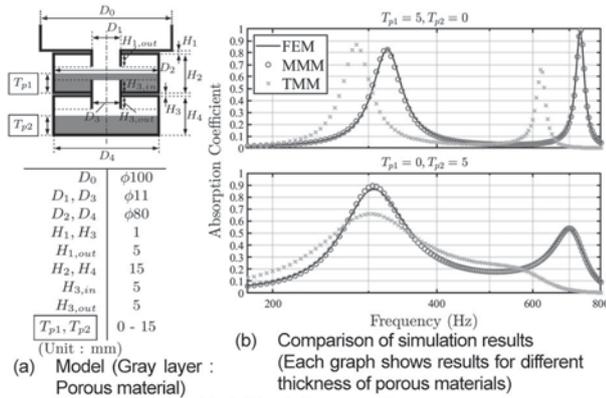


Fig. 1: Simulation example

(TMM considered Ingard's end correction at both ends of the neck)

2-7-10

2-7-10 膜・多孔質弾性材料間の連続条件が吸遮音性能に及ぼす影響の数値解析的考察

Numerical investigation on the effect of the continuity condition between membrane and poroelastic materials on sound absorption and insulation performances

◎井上尚久(九大・芸工)

- ◆膜・多孔質弾性体の接着・非接着が周期的に分布する場合 (Fig. 1) の計算例を示した。
- ◆繰り返し周期 T_s に関して以下の結果が得られた(Fig. 2)。
 - 波長よりも十分大きい場合には、各部分はそれぞれの特性を示し、全体としてはその割合で平均した値へ近づくことが示唆された。
 - 波長に比べて十分小さい場合には、均質化法の類推から、等価な連続条件が見いだせる可能性が考えられる。

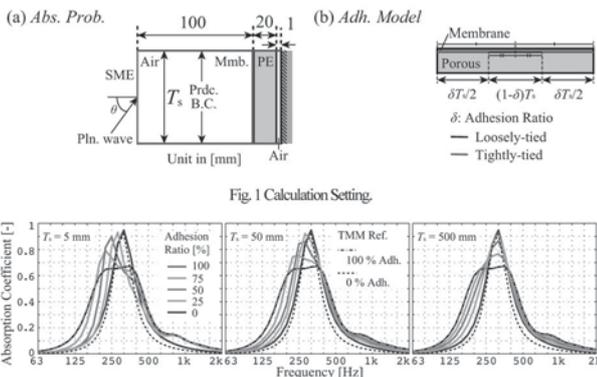


Fig. 2 Random-incidence absorption coefficients calculated for layered materials with different periods of the continuity distribution and adhesion ratios.

2-7-9

2-7-9 焼杉の表面性状が垂直入射吸音特性に与える影響に関する実験的検討

Experimental study on the effect of the surface conditions of yakisugi on the sound absorption characteristics

○菅原彬子, 平栗靖浩, △岡村健太郎(近畿大)

- ◆伝統的な外装材として利用されてきた「焼杉」を内装材として利用したいという需要がある。
- ◆本研究では、内装材としての利用を想定し、吸音材としての応用可能性を、音響管および残響箱を用いて検討する。
- ◆垂直入射吸音率の結果より、焼杉の炭化層は多孔質吸音材の特徴をもち、炭化層が厚いほど吸音率が高いことがわかった。
- ◆残響室法吸音率の比較により、塗膜性塗装の影響は小さく、塗装をしてもある程度の吸音性能を確保できる可能性が示唆された。

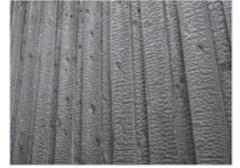


Fig. 1 yakisugi

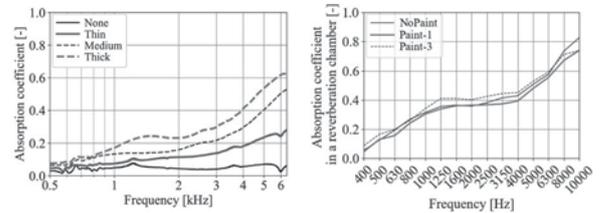


Fig. 1 Absorption coefficients at normal incidence with different thicknesses of carbonized layer

Fig. 2 Absorption coefficients in a reverberation chamber for yakisugi samples with and without coating

2-7-11

2-7-11 三重周期極小曲面の遮音特性に関する基礎的検討

Fundamental study on sound insulation characteristic of Triply Periodic Minimal Surface

☆矢野拓実(近畿大院), 菅原彬子, 平栗靖浩, △堂寺知成(近畿大)

- ◆Fig. 1 に示す曲面を三重周期極小曲面 (TPMS) という。これは昆虫の鱗粉などで確認でき、特定の光に対してバンドギャップを生成する。これを音に応用することで遮音材として有用なのではと考えた。
- ◆有限要素法を用いて TPMS の遮音性能を検討する。また、既報[1]の結果から TPMS の層を増加、さらに圧縮させる検討を行う。
- ◆解析結果から、それぞれの曲面で異なる周波数を遮音できるという結論を得た (Fig. 1)。また、層を増やすことで、バンドギャップ周波数帯域の性能がそのまま向上し、圧縮することで厚みを保ったまま遮音性能が向上するという知見を得た。

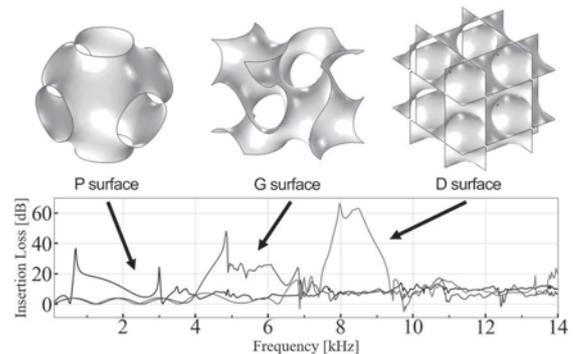


Fig. 1: Sound insulation performance of TPMS

2-7-12

2-7-12 ステージ上での音源の指向性が遮音板の 効果に及ぼす影響について

The effect of the directivity of the sound source on the stage on the effectiveness of the acoustic barrier

☆川瀬遠我, 高野靖, 大谷真(京大院・工学研)

- ◆ ステージ上の楽器の前には他の演奏者の聴覚を保護するための遮音板の効果を正確に3次元でキルヒホッフヘルムホルツ境界積分方程式(KHBE)を用いてシミュレーションを行い、遺伝的アルゴリズム(GA)を用いて遮音板の形状と配置の最適化を行い、異なる指向性や周波数を持つ音源に対して遮音板の効果がどのように変化するかを検討した。
- ◆ 複数の目的関数を用いてGAを行い、求められる音響特性に対応する目的関数を選別した。
- ◆ GAの結果から得られた遮音板を用いて、二重極音源を置いてみると、単極音源を元にGAを行った遮音板は二重極音源の場合より目的を達成できていないことが分かり、指向性を考慮する必要性が明確になった。

Table 1: Calculation conditions and results

Case	frequency[Hz]	directivity	t_x [m]	t_y [m]	X [m]	Z [m]	ΔL_{far} [dB]	ΔL_{near} [dB]
I	40	Monopole	4.0	2.5	-2.0	0.71	-2.4	-1.5
II	40	Dipole	1.8	2.8	-0.66	0.47	0.13	11.16
III	400	Monopole	2.0	1.4	-0.29	0.74	-3.0	-1.8
IV	400	Dipole	1.4	3.3	-0.72	0.35	0.14	20.1

2-8-2

2-8-2 自動車交通騒音と自動車単体騒音対策に係る現状について

Status of Road Traffic Noise and Vehicle Noise Regulation

○立田安礼(環境省水・大気環境局モビリティ環境対策課)

- ◆ 自動車交通騒音の低減に向けて、国が基準や規制を検討する際や、都道府県等が自動車騒音対策を計画的総合的に行うためには、騒音暴露状況を経年的に把握することが重要であり、騒音規制法第18条により都道府県等による自動車交通騒音の常時監視の実施及びその結果について環境省へ報告することを義務づけている。
- ◆ また、自動車単体騒音対策については、騒音規制法第16条に基づき規制が実施され、逐次強化されてきており、令和4年6月に今後の自動車単体騒音の低減のあり方(第四次答申)において、四輪車騒音の国際基準であるUN Regulation No.51 03 Seriesのフェーズ3の加速走行騒音規制値との調和等が中央環境審議会より答申された。
- ◆ 自動車交通騒音の常時監視の制度の仕組みや常時監視の結果、自動車単体騒音対策の検討経緯や第四次答申等について紹介する。

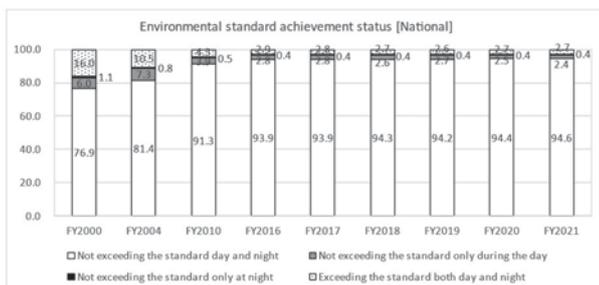


Fig.1: Achievement status of the EQSs of Road Traffic Noise

2-8-1

2-8-1 道路交通騒音の現状から将来へ

From the current situation to the future of road traffic noise

○山本 貢平(小林理研)

- ◆ 世界的なカーボンニュートラル政策の中で、変わりゆくモビリティの将来を念頭に、現在の道路交通騒音の状況、自動車単体騒音規制の強化、技術開発、環境管理の現状と将来を考える。
- ◆ 全国調査で環境基準の達成率は年々上昇しているが、苦情数は横ばい。
- ◆ ASJ RTN Modelからは、43年間に大型車からの騒音は改善されているが、小型車からの騒音は変わらない。
- ◆ 日本は、自動車単体騒音規制に国際基準を取り入れたが、将来、その効果がどのように表れるか?
- ◆ ZEV (Zero Emission Vehicles) の開発と東京都による導入促進の目標。
- ◆ 低騒音舗装の現状。Aireless Tire はゲームチェンジャーとなり得るか。
- ◆ タイヤ騒音のラベリング制度、マフラー性能等確認制度による騒音管理の取り組み。
- ◆ 台湾による改造マフラー車の自動取り締まり装置の導入。
- ◆ EUにおける騒音地図による環境管理と航空地図を用いたノイズマップの推計。

2-8-3

2-8-3 欧州 UK の環境騒音 News (WHO ガイドライン事例~ノイズマップ現状トレンド)

○平山 正充 (HBK) △Douglas Manvell (DMdB)

WHO ガイドラインの欧州環境騒音へのインパクトとノイズマップとアクションプランの現状報告をします。また欧州 Green Deal : 環境騒音と CN に言及します。このスペシャルセッションの内容は、(一財)日本自動車研究所が主催する「自動車交通に関する技術懇談会」で意見交換されたものである。

Environmental noise: news from Europe

This paper summarizes some important recent developments in Europe concerning environmental noise. The topics covered include the impact of the WHO noise guidelines on noise policy in Europe, using a few some examples. In addition, the European Green Deal regarding carbon neutrality and environmental noise is presented. The current status of the EU Environmental Noise Directive on noise mapping and action plans is given, with issues and trends highlighted. It closes with other recent topics concerning environmental noise in EU and UK.

2-8-4

2-8-4 商用車のカーボンニュートラルに向けた取り組み方針と道路交通騒音への影響に関する一考察

A study on the carbon neutrality policy for commercial vehicles and its impact on road traffic noise

○谷津幸裕(日野自動車株式会社)

- ◆地球温暖化防止のため世界規模でのカーボンニュートラル(以後 CN と略す)の議論が進められている。ここでは CN を達成できたクルマ社会での道路交通騒音について現在得られる知見やデータから考察した。
- ◆一般財団法人 日本自動車工業会(JAMA)では CN に向けての車両は BEV(バッテリーEV)に限らず多様な選択肢から模索すると公表している。また一部乗用車メーカーでは個社の戦略を示しているが、商用車メーカーでは日本自動車工業会の発表内容以外に個社の戦略を明確に示している事例は少ない。
- ◆これまでの知見やデータなどから BEV の乗用車の道路交通騒音に係る効果をもとに商用車の効果について考察した。今後、商用車は小型トラックと称する GVW4 トン以下クラスは一定程度 BEV 化が進むがその他のカテゴリーは内燃機関が残る可能性が想定される。
- ◆一方で道路交通騒音低減に対してはタイヤ路面騒音の寄与が大きくなることから、自動車単体騒音規制強化だけでなくタイヤ路面騒音対策や交通流対策、取り締まりなど含めた総合的な対策が必要。
- ◆今後の自動車単体規制強化の議論の際には CN に向けた車両の仕様の最新状況と、道路交通騒音の現状を把握したうえで議論が必要と思われる。

2-8-6

2-8-6 二輪車から発生する突発的な騒音に関する実験的検討

Experimental study on single event noise emitted from motorcycle

○田中雅文(自動車研)

- ◆道路交通騒音の課題の一つとして、一部の二輪車からの高いレベルの突発的な騒音の発生が挙げられている。そこで、走行実態等に基づき、運転方法やマフラーの交換が二輪車からの騒音の発生に及ぼす定量的な影響を調べた。
- ◆3水準(Normal, Aggressive, S-Aggressive)の運転方法による走行パターン調査結果に基づき、テストコースにて発生騒音を測定し、運転方法による騒音への影響(標準マフラー装着時の UN R41 の最終値 L_{urban} からの変化)を調べたところ、Normal や Aggressive の運転方法に比べて S-Aggressive で大きく、10 dB 以上高いレベルとなる場合のあることが確認された。
- ◆同様の結果からマフラー交換による影響を調べたところ、装着するマフラーにより幅があるが、標準装着のマフラーよりも 10 dB 以上(最大で 23 dB)高いレベルとなり得ることが確認された。
- ◆運転方法とマフラー交換の両者の組み合わせによる影響をみたところ、装着するマフラーにより大きな幅があるが、標準マフラー装着時の L_{urban} と比べて、多くの場合で +10 dB 以上、最大で +33 dB と非常に高いレベルの騒音を発生し得ることが確認された。

Table 1 Example of level differences by driving conditions and/or replacement muffler

ΔSPL, dB	①		
	Normal	Aggressive	S-Aggressive
Original muffler ^(a)	-1 ~ +2	+2 ~ +5	+6 ~ +13
② ↓ΔSPL ^(b)	↓+19 ~ +23	↓+18 ~ +22	↓+17 ~ +23
Replacement muffler ^(a)	+21 ~ +24	+22 ~ +27	+28 ~ +33

① ΔSPL by driving conditions (a) ΔSPL from L_{urban} with original muffler
 ② ΔSPL by replacement muffler (b) ΔSPL from original muffler

2-8-5

2-8-5 騒音レベルと路面騒音因子(路面性状値)の関係

The relationship between pass-by noise levels and road surface characteristics

○常松 直志(日本道路(株)), △井澤 克則((株)レインボー・コンサルタント)

- ◆車外騒音測定試験用路面として ISO 10844:2011-2021 により、路面条件等の規定がなされているものの、わずかな路面性状値の差異により車外騒音の評価値に差が生じているのが実情である。
- ◆本研究では路面特性を変更した路面を9種類構築し、騒音レベルと路面性状値(路面テクスチャ(MPD)・すべり抵抗・吸音率)との関係性について、4種類の測定用タイヤで騒音測定し、検証を行った。
- ◆測定結果から基準タイヤ(SRTT)はMPDと騒音レベルに係性は認められなかったが、他の3種類のタイヤでは正の相関が認められた。
- ◆スポーツタイヤ(Sports)についてはBPNと騒音レベルに負の相関が認められ、高い相関を有しているが、基準タイヤ(SRTT)については、関係性は低い。
- ◆いずれのタイヤも吸音率と騒音レベルは負の相関が認められ、最も支配的な路面騒音因子と判断された(Fig.1)。

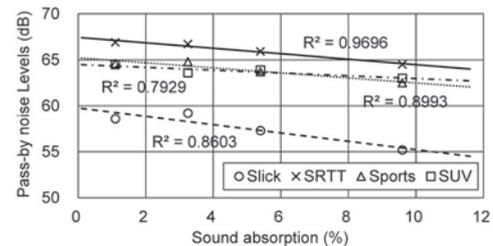


Fig.1: Relationship between sound absorption and pass-by noise levels

2-8-7

2-8-7 オープンデータと ASJ RTN-Model を用いて作るノイズマップ

A noise map created based on open data and the ASJ RTN-Model

○坂本慎一(東大生研), 章 心怡(東大大学院)

- ◆現代の社会資産であるオープンデータと、音響学会が社会に提供する公共財産である ASJ RTN-Model を組み合わせることによって、ノイズマップを作成した事例を紹介した。
- ◆ノイズマップにより広域の騒音測定等の大きな労力をかけずに環境影響を評価でき、良好な環境実現のために必要な対策を検討することができる。
- ◆カーボンニュートラル社会に向けて道路交通騒音の状況が変化する時代には、我が国の道路交通騒音を常にウォッチし続ける道路交通騒音調査研究委員会の活動とその成果である ASJ RTN-Model の重要性がますます大きくなる。

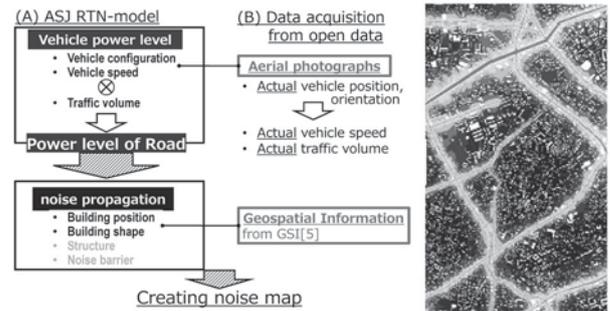


Fig. 1 Flow for creating a noise map Fig. 2 Example of noise map

2-8-8

2-8-8 道路交通騒音の予測モデル“ASJ RTN-Model 2023”
—2023年版の概要—

Road traffic noise prediction model “ASJ RTN-Model 2023”
—Outline of 2023 version

○岡田恭明(名城大・理工), 安田洋介(神奈川大・建築),
坂本慎一(東大・生研), 一木智之(ニューズ環境設計),
穴井 謙(福岡大・工), 山内勝也(九州大・芸工)

日本音響学会道路交通騒音調査研究委員会は、昭和49年に組織され、昭和50年に騒音レベルの中央値 L_{A50} を予測値とする計算方法(ASJ Model 1975)を発表した。昭和63年からは等価騒音レベル L_{Aeq} に基づく計算方法の開発にも取り組み、平成5年にエネルギーベースの予測計算方法(ASJ Model 1993)を発表した。その後、平成10年の「騒音に係る環境基準」の改正に伴い、その評価量 L_{Aeq} に対応したASJ Model 1998を平成11年4月に公表している。以降、予測の適用範囲の拡大、予測精度の向上を目的にして、本モデルの改定を5年ごとに行っており、今回、その成果としてASJ RTN-Model 2023を今年4月発行の日本音響学会誌80巻4号で発表する予定である。本スペシャルセッションでは、ASJ RTN-Model 2018から変更した点を中心に、次の4件の内容を報告し、皆様からの御意見等を伺いたいと考えている。

- ▶ 音響パワーレベル(2章): 岡田恭明(名城大学・理工)
- ▶ 伝搬計算方法(3章, 付属資料A3・A4):
安田洋介(神奈川大・建築)
- ▶ 高架構造物音(5章, 付属資料A1, 参考資料R6):
一木智之(ニューズ環境設計)
- ▶ 予測精度(参考資料R5): 山内勝也(九州大・芸工)

2-8-10

2-8-10 道路交通騒音の予測モデル
“ASJ RTN-Model 2023”—伝搬計算方法—

Road traffic noise prediction model “ASJ RTN-Model 2023”:
sound propagation calculation method

○安田洋介(神奈川大・建築), 一木智之, 福島昭則(ニューズ環境設計),
坂本慎一(東大・生研), 横田考俊(小林理研)

- ◆新たな検討結果と知見を踏まえ、ASJ RTN-Model 2018(以下、Model 2018)の伝搬計算法を2023年版(以下、Model 2023)に更新した。
- ◆形式上の主な変更点は、伝搬計算の基本式において地表面の影響を明記したことである。これに伴い、Model 2018と同じ考え方に基づくものでも表記が変更されている箇所がある。
- ◆反射音の計算方法に関しては、吸音性障壁の回折補正量を用いた鏡面反射法を基本とするものに変更し、帯状反射面及び半無限反射面での反射補正量を、矩形反射面での反射補正量の計算方法を用いて計算する形に変更した。
- ◆空気の音響吸収に関する補正量に関しては、舗装種別ごとに区別した他、高架構造物音に関するものを追加した。
- ◆伝搬経路における回折点を平面座標から近似的に求める点について注記を加えた。
- ◆その他の各種補正量や気象の影響に関して、最新の知見に基づき注記等を加え、適用範囲をより明確にした。
- ◆関連する付属資料・参考資料について、本文の変更に準ずる形に変更した。

2-8-9

2-8-9 道路交通騒音の予測モデル“ASJ RTN-Model 2023”
—音響パワーレベル—

Road traffic noise prediction model “ASJ RTN-Model 2023”
—Sound power level of road vehicles

○岡田恭明(名城大・理工),
一木智之, 福島昭則(ニューズ環境設計), 山内勝也(九州大・芸工)

ASJ RTN-Model 2023の音源特性に関して、この5年間における検討結果を踏まえ、以下の点を変更した。それらの概要について報告する。

- ▶ 自動車走行騒音のパワーレベルに関して、走行区間別ではなく、走行状態別(定常・減速走行, 非常走行, 加速走行)に変更した。
- ▶ 非常走行及び加速走行の自動車走行騒音のパワーレベルのモデル式に関して、実用性を考慮して、定数 a 、速度依存性を表す係数 b (料金所付近での係数 b を10に固定)、排水性舗装等による騒音低減効果の経年変化を表す係数 c (定常走行の係数と整合)を見直した。
- ▶ 最新の知見に基づいて、自動車専用道路におけるパワーレベル式に対する補正項(ΔL_{traf} : 日通過交通量に伴う排水性舗装等による騒音低減効果)を見直した。
- ▶ 自動車専用道路の密粒舗装における加速走行並びに一般道路における排水性舗装での自動車走行騒音のパワーレベルについては、それぞれ2.2.2項及び2.2.3項にまとめた。

その他、令和4年度までの自動車専用道路における実測データから、排水性舗装等における定常走行のパワーレベル式の有効性なども確認しており、その適用範囲を排水性舗装の場合は舗設後11年まで、高機能舗装II型の場合は舗設後8年までと注記に記載した。

2-8-11

2-8-11 道路交通騒音の予測モデル
“ASJ RTN-Model 2023”
—高架構造物音—

Road traffic noise prediction model “ASJ RTN-Model 2023”
—Structure-borne noise of viaducts—

○一木智之, 福島昭則(ニューズ環境設計), 岡田恭明(名城大・理工)

- ◆道路交通騒音の予測モデル“ASJ RTN-Model 2023”の高架構造物音に関して、以下の点を更新・追加した。それらの概要について報告する。
- ◆高架構造物音のパワーレベルに関して、最新の知見に基づいて見直した。
- ◆高架構造物音の伝搬計算方法に関して、地面反射音を考慮するように更新した。
- ◆高架構造物音の周波数特性を追加した。
- ◆高架構造物音のパワーレベル測定方法を見直した。

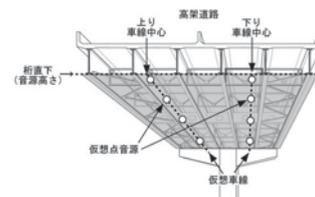


Fig. 1 Arrangement of hypothetical point sources for structure-borne noise.

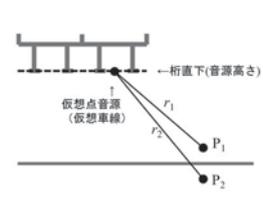


Fig. 2 Geometry for calculation of structure-borne noise.

2-8-12

2-8-12 道路交通騒音の予測モデル “ASJ RTN-Model 2023” —予測精度—

Road traffic noise prediction model “ASJ RTN-Model 2003”
—Prediction accuracy—

○山内勝也(九州大・芸工), 穴井 謙(福岡大), 一木智之(ニュース環境設計),
岡田恭明(名城大・理工), 坂本慎一(東大・生研)

- ◆ ASJ RTN-Model 2023 の予測精度(=予測計算値と実測値との対応の良さ)を検討し、それに影響する要因を考察する。
- ◆ 本稿では、自動車専用道路の道路近傍と沿道において実測したデータにより、予測精度の検討をおこなった事例を紹介する (Fig. 1)。
- ◆ ASJ RTN-Model 2023 で新たに追加されたラウドアバウト周辺部における騒音予測の精度の検討例を紹介する。
- ◆ 予測精度に影響する要因について、これまでのモデルで紹介されていた知見群を再整理した内容について紹介する。

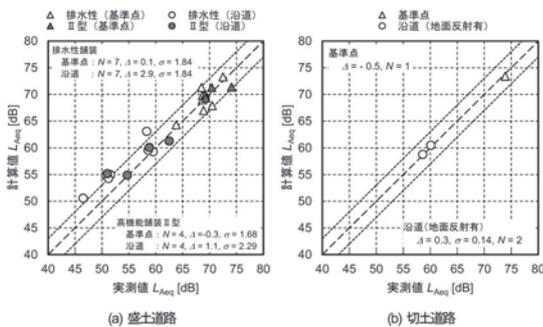


Fig. 1 自動車専用道路における計算値と実測値の対応

2-9-2

2-9-2 美術館におけるサウンドマスキングの 活用に向けた実践的研究

Practical study for utilization of sound masking in an art museum

◎松尾綾子(早稲田大/明治大), 上野佳奈子(明治大)
△竹内利夫, △亀井幸子(徳島県立近代美術館)

- ◆ 日障害者の鑑賞機会の提供に積極的に取り組んでいる徳島県立近代美術館において、鑑賞空間の環境を調整する試みとして、サウンドマスキングの導入を行った。
- ◆ 鑑賞に適したマスキング音を選定する手がかりを得るため、展示作品と音の組合せの影響に着目して、展示室での鑑賞時の音の効果を評価するワークショップを行った。さらに、その結果を踏まえて特別展にて再度マスキング音を導入し、サウンドマスキングの効果及び実用性を検討した。
- ◆ その結果、マスキング音のメロディやテンポ、拍子感等の音楽的な特徴が鑑賞に影響を与えることが示唆された。また、マスキング音が鑑賞時の居心地の良さやリラックス効果、緊張感の緩和、会話のしやすさ等に対して効果があることが分かった (Fig.1)。

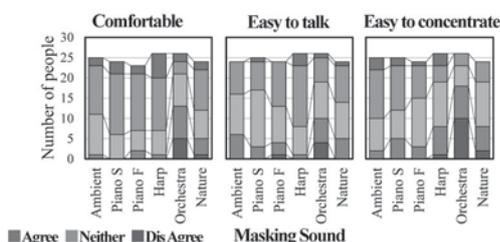


Fig.1: Results of evaluating the impression of each Masking Sound.

2-9-1

2-9-1 音環境および光環境による快適かつ 集中力向上のための作業環境の構築

Creating a comfortable and concentrated workplace
with sound and light environments

☆松本大祐(中央大院), 戸井武司(中央大)

- ◆ 近年テレワークなどの導入企業が増え、従来の作業場所から変化が生じており、作業環境の改善が求められている。本研究では、作業中の音環境と光環境に着目し、音環境と光環境の両方を考慮した快適で集中できる作業環境の構築を目指す。
- ◆ 音環境の単刺激として、周囲からの音声情報マスキング時の主観および脳血流を用いた客観評価を実施し、マスクーに音楽を付加することで、快適かつマスキングに適した音環境を構築できることを示した。
- ◆ 光環境の単刺激として、照度を変更した際の主観および客観評価を実施し、照度の高い環境で作業に集中できる光環境であることを示した。
- ◆ 音環境と光環境の複合刺激において主観および客観評価を実施し、Fig. 1 に示した脳血流結果より、音声マスキング時および照度の高いSL4の条件で他の条件と比較して最も快適で集中できることを示した。Fig. 2 に示す2桁の加減算によるパフォーマンス評価結果より、実際に作業時間が短縮し、本実験条件下では音環境より光環境の方が作業時間短縮への寄与が大きいことを示した。

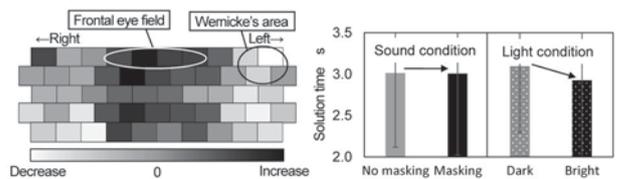


Fig. 1 Cerebral blood flow of SL4 based on result of SL1 under complex stimuli

Fig. 2 Calculation times under complex stimuli

2-9-3

2-9-3 聴覚過敏者が捉える大学キャンパスの 音環境に関する調査研究

Surveys on the Acoustic Environment of University Campus as Perceived by Individuals with Hyperacusis.

☆岩水岳喜, △伊藤大地, △富田克典, 高田正幸(九州大・芸工)

- ◆ 聴覚過敏の特性を持つ学生や教職員が、大学キャンパスの音環境をどのように捉えているかを把握するために、九州大学伊都キャンパスにおいてISO 12913-2: 2018に基づくサウンドスケープ調査を実施した。
- ◆ Khalfa Hyperacusis Questionnaire 日本語版により参加者の聴覚過敏の特性を評価したところ、参加した30名のうち、13名が聴覚過敏の傾向を持つと判断された。
- ◆ 「食堂」は、全体として「活気がある (vibrant)」「出来事が多い (eventful)」と評価されたが、特に聴覚過敏の特性を持つ参加者は、持たない参加者と比べて、快適性 (pleasant) の評価が低かった。
- ◆ 「図書館」は、全体として「快適な (pleasant)」「落ち着いた (calm)」と評価されたが、聴覚過敏の特性を持つ参加者は、持たない参加者と比べて、より「活気がある」「出来事が多い」と評価した。

Table.1: Mean scores of the soundscape quality at each survey site. (A: people with hyperacusis, B: people without hyperacusis, *: p<0.05)

Scales	Library		Dining		Classroom	
	A	B	A	B	A	B
pleasant	4.15	4.24	2.92	3.82*	4.00	4.06
chaotic	1.46	1.18	3.46	3.35	1.62	1.47
vibrant	1.85	1.18*	3.85	4.18	1.62	2.00
uneventful	3.92	4.12	2.00	2.06	4.23	3.94
calm	4.62	4.76	2.31	2.29	4.08	3.88
annoying	1.08	1.18	3.08	2.71	1.38	1.76
eventful	2.00	1.29*	4.00	3.76	1.69	2.00
monotonous	4.00	3.82	1.77	1.94	3.69	3.76

2-9-4

2-9-4 仮想空間における響きの印象に関する研究 —3DモデルによりVR表示された建築空間 の視覚情報に調和する残響時間の提案—

Study on the impression of reverberation in virtual space -Proposal of reverberation time to match the visual aspect of architectural space displayed in VR using 3D models-

☆風岡翔太, 石川あゆみ(岐阜高専)

- ◆本研究では、3DモデルによりVR表示された建築空間の視覚情報に調和する残響時間を提案することを目的とする。
- ◆本報では、マテリアル(3Dソフトで作成する質感)とオブジェクト(仮想空間内に設置する3Dモデル)が同じで容積が異なるコンサートホールの3Dモデルを対象とし、その予想残響時間(人が建築空間の視覚情報(見た目)から予想した響き)と視覚印象の評価結果から、上記の残響時間を検討する。
- ◆コンサートホールの3Dモデルの表面積と吸音率(マテリアルとして採用した表面部材の吸音率)をEylingの式に代入して求めたCRTが、予想残響時間と大きく乖離する結果となった。

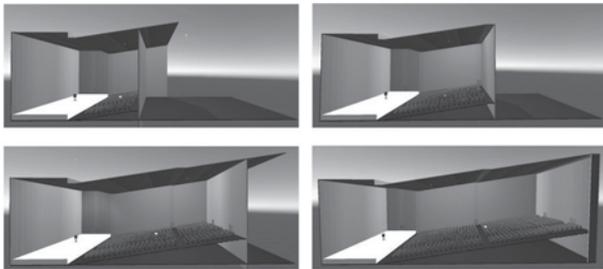


Fig. 1: 3D model of the concert hall

2-10-1

2-10-1 秋季における日本海(Japan sea)の 音波伝搬

Sound propagation of NIHONKAI (Japan sea) in autumn

○鶴ヶ谷芳昭(三陽精工), 菊池年晃(防衛大), 水谷孝一(筑波大)

- ◆日本海は外洋と水深約200mの海峡でつながっている。冬季の冷却により、密度を増した海水は日本海の底部に沈み、日本海固有水を形成する。夏季に比して、秋季は海面付近の水温上昇は小さくなり、対馬暖流の勢力も減少するが、高い垂直音速傾度(VSG)は存在する。検討ラインの南側では夏季の影響が残り、サウンドチャンネル(SC)軸深度は深い。
- ◆Fig. 1に春季から秋季の大陸側から日本サイドまでの音波伝搬を示す。a)は春季。b)は夏季。c)は秋季であり、音源深度(SD)は10mである。春季はSC軸深度が浅く、SC伝搬の傾角を有する。秋季においては夏季ほどではないが高いVSGが存在し、SDが浅いと音源を出た音波は急角度で海底方向に向かい、海底反射に伴う伝搬損失(TL)が増大し、伝搬エネルギーは急速に減衰する。深度が深くなると日本海固有水の影響を受け、浅いSC軸深度を生じる。その結果、SDが深くなると音波伝搬はSC軸を中心とした拡散場を形成する。

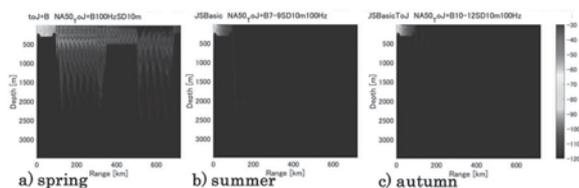


Fig. 1 Comparison of sound propagation from the continental side in spring, summer and autumn SD:10m

2-9-5

2-9-5 6DoF音響コンテンツの音空間デザインに用いる 非発音源メタデータの要求条件

Requirements of audio metadata for non-sound-source to design acoustic space of 6DoF audio contents

○杉本岳大(NHK)

- ◆イマーシブメディア用の音響技術は6DoFが前提となるため、コンテンツ制作には6DoF対応の音響メタデータが必要になる。非発音源メタデータの中でも、特に音空間デザインに関する項目の詳細な規定を目的として、コンテンツ制作者への意向調査をおこない、要求条件を抽出した。
- ◆意向調査は、コンテンツ制作における音空間デザインについての13個の設問(デザインの方針や映像との関係性、音響処理手法など)に、自由記述で回答する形式とした。
- ◆「空間の状態」「物体の状態」「状況」を記述するための要素項目を、表1-3にそれぞれ示した。表1の「空間の状態」および表2の「物体の状態」の要素項目は、物質のカテゴリに属するものと音響のカテゴリに属するものに大別した。表3の「状況」の要素項目については、時間のカテゴリに属するものと天候のカテゴリに属するものに分けて整理した。

表1 空間の状態を規定する要素項目

物質	音響
<input type="checkbox"/> 室のカテゴリ	<input type="checkbox"/> 直間比
<input type="checkbox"/> 室のサイズ	<input type="checkbox"/> 残響の質
<input type="checkbox"/> 室の形状	<input type="checkbox"/> 残響の量

表2 物体の状態を規定する要素項目

物質	音響
<input type="checkbox"/> 形状	<input type="checkbox"/> 反射率
<input type="checkbox"/> サイズ	<input type="checkbox"/> 吸音率
<input type="checkbox"/> 位置	<input type="checkbox"/> 透過率
<input type="checkbox"/> 質感(材質、密度など)	

表3 状況を規定する要素項目

時間	天候
<input type="checkbox"/> 時刻	<input type="checkbox"/> 曇量
<input type="checkbox"/> 日付	<input type="checkbox"/> 降雨量
<input type="checkbox"/> 年代	<input type="checkbox"/> 降雪量
<input type="checkbox"/> 季節	<input type="checkbox"/> 風速

2-10-2

2-10-2 新たな水中無人機研究施設 ~IMETS~の大型水槽の音響特性

Acoustic characteristics of the Large Acoustic Tank at "IMETS", a new research facility for Unmanned Underwater Vehicle

○伊藤魁, △山口智之, △嶋村英樹(防衛装備庁・IMETS)

- ◆防衛装備庁艦艇装備研究所岩国海洋環境試験評価衛星(Imwaku Maritime Environment Test & Evaluation Satellite: IMETS)に設置された大型水槽において実施した背景雑音及び音響反射特性の計測結果について述べる。
- ◆音響反射特性計測については、2種類の計測方法による計測及び結果の比較を実施した。
- ◆Fig. 1に背景雑音の計測結果を、Fig. 2に吸音材の取り付けられた、水槽壁面における音響反射特性の計測結果を示す。

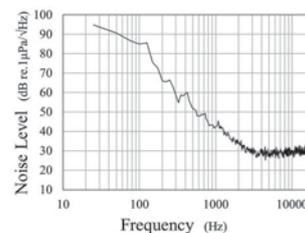


Fig. 1: Noise level of Large Acoustic Tank

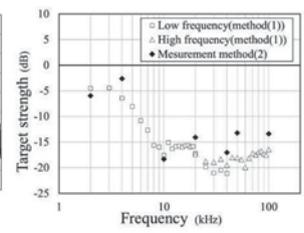


Fig. 2: Acoustic reflection characteristics of the walls of the Large Acoustic Tank with sound absorbers were attached

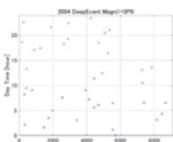
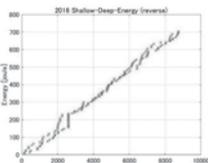
2-10-3

2-10-3 紀伊半島南東沖地震の震源構造とその発震時変化

Hypocenter structure of the Kii Peninsula Southeast-Oki Earthquake and its change in standup time.

○菊池年晃(防大), 水谷孝一(筑波大)

中部海域には地震を頻発する震源域が幾つか存在する。その中で最も活動的な紀伊半島南東沖震源域は、地域的に限定されているだけでなく深度的にも二分化されている。それらは深度 50km 付近の浅部震源域と深度 400km 付近の深部震源域とが存在する。2016 年に、これらの震源域で発生した年間累積エネルギーを上図に示す。深部のエネルギーを+で、浅部のエネルギーを+で示した。両震源域でのエネルギー推移は一致して、浅部域のエネルギーは深部域のそれより時間的な遅れがある。即ち、深部域での発震が重要である。そこで2004年に深部域で発生した地震の発生時刻(1日当たり)を発生日時(1年を通した)で下図に示した。この図の前半(0-5000)は、5時を中心とした範囲と、20時を中心とした後半とに時間的な分離が見られる。また、6000-7000の間に、地震が全く発生しない静寂期が存在する。この時間は浅部域で9月5日に発生したM7.4の初期と一致している。上図では深部域から浅部域へとエネルギーが流れているが、下図の静寂期は、浅部域から深部域へと反作用的な応力作用が働いたこと示している。



2-10-4

2-10-4 生物模倣音を用いた水中音響測位用送信信号の多重化に関する基礎的検討

Fundamental study of demultiplexing multiple signals by biomimetic pulse trains for underwater localization

○小笠原英子, 黒山喬允, 森 和義(防衛大)

- ◆海洋開発と海洋保全の両立のため、既存の海中の音環境を大きく乱さない測位信号が求められる。
- ◆ハンドウイルカのクリックスを連ねたパルス列を用いた測位信号を提案する。繰り返し回数を増やすことでメインローブのピークを大きくできるが、サイドローブも高くなる。パルス間隔に揺らぎを与えることで、サイドローブを抑制できる。
- ◆パルス間隔を変化させることで複数の信号を生成でき、**多重化しても分離可能**である。
- ◆北海道石狩沖を想定した3つのトランスポンダとの数値解析では伝搬距離が300 m程度までの受信波形から各送信点までの伝搬時間を正しく推定できた。

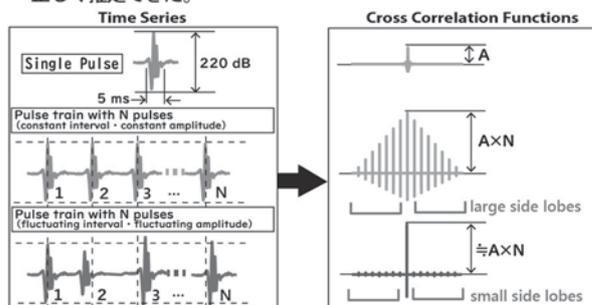


Fig. 1: Time series of pulse trains and its auto-correlation functions.

2-10-5

2-10-5 鳥島近海の地震に伴う極めて顕著な海中伝搬波

Extremely prominent underwater acoustic waves associated with earthquakes near Torishima Island

○岩瀬良一(JAMSTEC)

- ◆2023年10月9日の午前4時から6時台(JST)にかけて鳥島近海で発生した地震に伴う極めて顕著な海中伝搬波が、海底に設置されたS-net観測点のみならず、一部の陸上観測点の地震計でも観測された。
- ◆観測された地震波自体の振幅が小さく、マグニチュードや震央が決まらない一方で、伊豆・小笠原諸島並びに日本の南岸の各地において数十cmの津波を観測した特異な地震である^[1]。
- ◆一連の主要な14個のイベントのうち、05:13の地震(M5.2)の震央の位置に海中伝搬波の音源があると仮定して各観測点までの伝搬時間を計算し、観測点毎に観測波形の時間軸を伝搬時間分ずらし(観測時刻から差し引いて)伝搬距離順に波形を並べた図を作成した結果、波形の並びから、音源は震央からずれている可能性が示唆される。
- ◆特に音源に近い気象庁観測点は、S-net観測点に比べて信号が遅延しているように見えることに加え、長周期成分が卓越している。
- ◆上記長周期成分の卓越などS-netと相違している陸上観測点の波形の特徴も考慮の上、今後音源定位について詳細に検討する。
- ◆また、いずれのイベントにも観測網の北側で反射した海中伝搬波と考えられる後続波が観測されており、その成因についても検討する。

参考文献

[1]気象庁, 令和5年10月9日05時25分頃の鳥島近海の地震について, <https://www.jma.go.jp/jma/press/2310/09b/202310091100.html>.

2-10-6

2-10-6 超音波を用いたナノ薬剤投与システムの開発

Administration system of nano medicine by using ultrasound

○倉科佑太(農工大)

- ◆一般的な低分子医薬品は効果が短く、副作用が大きいため、薬効に優れた高分子医薬品に注目が集まっている。一方で、高分子医薬品は分子量が大きいことから投与が難しい。そこで本研究では、超音波を利用した新たな投与方法を提案した。超音波によるオンデマンド薬剤放出の研究として、ハイドロゲル内に音響インピーダンスの異なる振動核を含有することで薬剤放出量が向上することを示した。加えて、過飽和CO₂をリポソームの内水層に含有することで超音波への応答性が顕著に向上することを明らかにした。また、超音波による経皮薬剤投与の研究では、bFGFを高分子担体でカプセル化することで投与効率を有意に高めることに成功した。このように超音波を活用したナノ薬剤投与システムが高分子医薬品に対する有効性を示した。

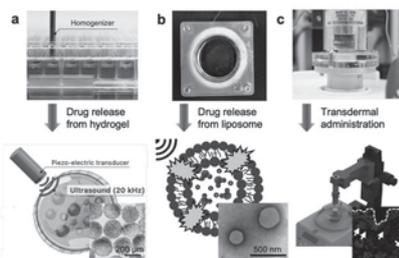


Fig. Nanodrug administration system using ultrasound. Drugs were released from the (a) hydrogel and (b) liposome. (c) Drug were transdermal administration.

2-10-7

2-10-7 超音波キャビテーションによる薬物の経皮吸収促進

Enhancing Transdermal drug permeation through ultrasonic cavitation

○杉田直広, 山本真也, △進士忠彦(東工大)

- ◆ 経皮薬物送達システムにおける物理的吸収促進手法の一つである超音波導入法では、音波照射による皮膚のダメージを最小限に抑えながら、超音波キャビテーション効果を皮膚表面に選択的に作用させることが重要となる。
- ◆ 本研究では、皮膚や粘膜などの生体表面に超音波キャビテーションを発生させることを目的として、超音波キャビテーションの発生に有効な 100 kHz 以下の低周波で駆動する超音波トランスデューサを開発している。
- ◆ 本稿では、試作機を用いてヒト皮膚構造を模倣した人工合成膜に超音波処理を行い、処理後の人工膜に対してカフェインの膜透過試験を行った結果を示す。

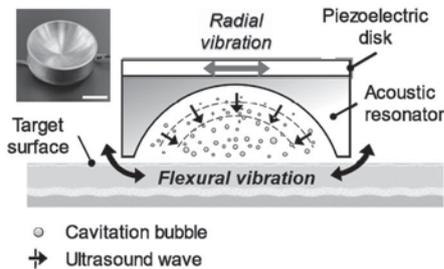


Fig.1: Schematic illustration of the proposed ultrasonic transducer generating ultrasound cavitation on the target material surface.

2-10-9

2-10-9 VHF 帯域の強力超音波照射によって発生するキャビテーションナノバブル粒径の周波数依存性

Frequency dependence of cavitation nanobubble size generated by high-intensity ultrasonic irradiation in the VHF range

☆原田裕生(同志社大), 石河睦生(桐蔭横浜大), 松川真美, 小山大介(同志社大)

- ◆ VHF 帯域の強力超音波照射によって発生するキャビテーションナノバブル粒径の周波数依存性および超音波照射回数と溶存酸素量の変化について評価した。
- ◆ KNbO₃ 圧電薄膜を用いた超音波トランスデューサを開発し、同トランスデューサを使用して VHF 帯域の強力超音波照射を行った。
- ◆ キャビテーションナノバブル粒径の測定には、動的散乱法(DLS)を用いた測定システムを使用し、高周波強力超音波照射後の測定試料にて 100-200nm 台の粒径をしたナノバブルが観測された (Fig. 1)。
- ◆ 超音波照射回数を増加させていくに従って、バブル粒径は増大していき、溶存酸素量に関しては減少していく傾向が観測された。高周波強力超音波照射によるバブルの凝集・合体が関与している可能性が高い。

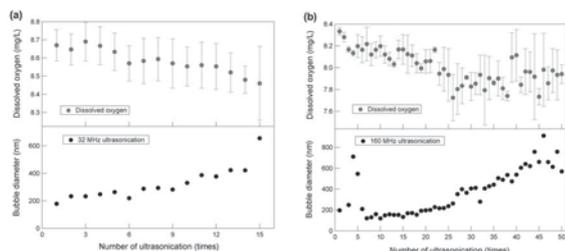


Fig.1: Relationships between the number of ultrasonication, the peak values of the cavitation nanobubble size, and the amount of dissolved oxygen change at (a) 32 MHz (b) 160 MHz.

2-10-8

2-10-8 生体吸収性 PLLA 樹脂による超音波誘発電位

Ultrasonically induced electrical potentials in a bioabsorbable PLLA film and bone

☆北嶋昇太, 児玉周大, 松川真美(同志社大)

- ◆ **ポリ-L-乳酸 (PLLA)** と低出力超音波パルス (LIPUS) を併用した**新たな骨折治療法**を実験的に検討した。
- ◆ **ウシ皮質骨試料に PLLA フィルムを接着した試料を圧電体とみなして超音波受波器**を作製し、超音波を照射した際の誘発電位を**皮質骨のみの受波器と比較した**。
- ◆ 超音波照射時において、PLLA を接着した場合と骨単体の場合の**音波波形の形状は類似していた**。
- ◆ 骨に PLLA を接着した際の平均誘発電位は骨単体の場合に比べ、**約 1.4 倍であった (Fig.1)**。
- ◆ 超音波照射時の誘発電位が骨折治療に寄与する場合、骨折部への PLLA 装着が治療の促進に役立つ可能性がある。

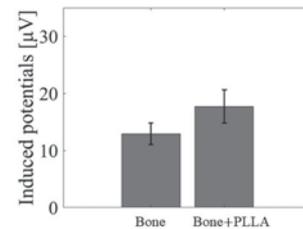


Fig.1: Average induced potentials in each area on the sensor surface when ultrasound irradiation.

2-P-1

2-P-1 3 話者以上の話者交替情報を用いた Speaker Diarization のためのデータ生成

Data generation for speaker diarization using speaker turn with more than three participants

☆市川奎吾, 上乃聖, 李晃伸(名工大)

- ◆ End-to-End Neural Diarization (EEND) モデルの学習データ生成法 Simulated Conversations (SC) 法に話者交替情報を導入する。
- ◆ 「最後に発話した話者から次の話者の確率」を実データから推定し、これを用いて複数の話者交替パターンを模したデータを生成する。
- ◆ 提案法の生成データは話者による発話区間のばらつきが顕著になり、複数話者における「話し手」と「聞き手」の音声を再現できた (Fig.1)。
- ◆ 2 話者においては話者の取り違えに関するエラーが 3.3 ポイント減少し、DER が改善した (Table 1)。3, 4 話者においては今回の実験では差はほとんど見られなかった (Table 2)。

Table 1: DERs (%) evaluated on the CALLHOME-2spks. DER consists of Missed Speech (MS), False Alarm (FA), Speaker Confusion (SC).

Dataset	MS	FA	SC	DER
SC	5.6	7.5	9.2	22.2%
Proposed	4.8	9.0	5.9	19.8%

Table 2: DERs (%) evaluated on the CALLHOME-3spks and 4spks.

Dataset	3spks	4spks
SC	37.2%	61.5%
Proposed	37.8%	60.5%

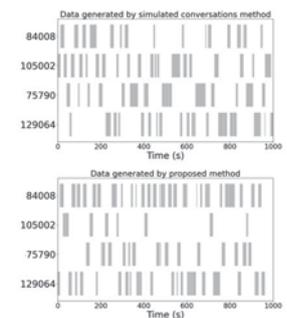


Fig. 1: Generated data

2-P-2

2-P-2

双曲空間への話者埋め込み

Hyperbolic Speaker Embeddings

©関翔悟, 宮崎晃一(サイバーエージェント)

- ◆話者埋め込み: 話者認識システムから得られる潜在表現
 - 音声合成などのアプリケーションに応用可能
 - アプリケーションでは数百次元の話者埋め込みが利用
→システムの軽量化や低遅延化においてボトルネック
- ◆低次元な空間への埋め込み手法: 双曲空間を用いたアプローチ
 - 話者認識においても低次元な話者埋め込みを実現
 - 実用されている話者埋め込みと比較すると限定的な性能
- ◆低次元でも有効な話者埋め込み手法を提案
 - 双曲空間上のデータを扱う双曲ニューラルネットワークを利用
 - 話者識別・話者認証で利用される損失関数を拡張
- ◆VoxCeleb データセットを用いた話者認証タスクで有効性を検証

Table 1 Comparison of EER [%]

Method	# of dim. D	
	64	512
Poincaré triplet [2]		
$c = 0.000$	11.59	N/A
$c = 0.003$	37.54	N/A
$c = 0.030$	10.96	N/A
$c = 0.300$	11.44	N/A
AAM-Softmax	1.88	2.08
Baseline (eqs. (2) and (6))	1.53	1.27
Euclid (eqs. (13) and (17))	1.59	1.40
Hyperbolic1 (eqs. (10) and (14))	1.23	1.21
Hyperbolic2 (eqs. (12) and (14))	1.37	1.21

2-P-4

2-P-4

深層学習を用いた 自然な音声対話システムのための 相槌表層選択モデルの構築

Construction of a Backchannel Surface Selection Model for Natural Spoken Dialog Systems Using Deep Learning

☆福永善理, 西村良太(徳島大), 太田健吾(阿南高専), 北岡教英(豊橋技科大)

- ◆音声対話システムの実装には相槌を適切に予測するモデルが必要
- ◆BERT をファインチューニングすることで 11 種類の相槌を予測
- 混同行列を用いて学習モデルの評価を行う
- 学習モデルと大規模言語モデルを比較し、相槌の自然性を調査
- モデルは、自然な相槌を選択することが可能であった

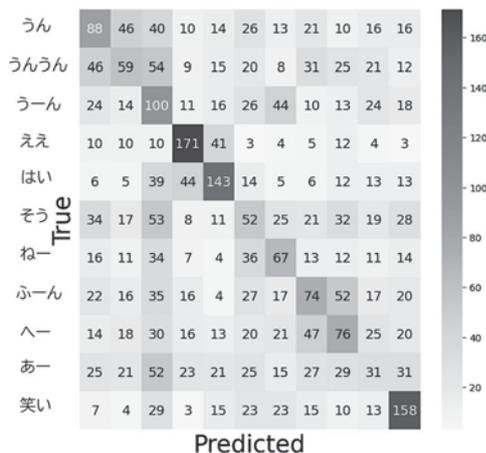


Fig.1: Confusion matrix

2-P-3

2-P-3

話者年齢埋め込みベクトルを用いた 話者年齢推定能力の評価

Evaluation of speaker age estimation based on speaker age embedding vectors.

☆藤居 謙, 西村 竜一(和歌山大)

- ◆話者年齢埋め込みベクトル (Age-vector) を用いた若年話者判別タスク (先行研究) の追加検証
 - 抽出 NN: 発話音声から Age-vector を抽出
 - 全結合 NN: 抽出した Age-vector を入力とした 3 クラス (若年者・大人 (男性)・大人 (女性)) の分類
 - 「抽出 NN+全結合 NN」と「抽出 NN 単体の出力」の比較
 - ◇ 比較的規模の小さいデータセットを用いて高い精度の分類器を構築できる高い利便性
- ◆年齢推定タスクへの展開
 - 後段の「全結合 NN」を年齢×性別の 155 クラス分類に拡張
 - 「年齢誤差 (推定年齢と実際の話者年齢の差)」を尺度とした検証
 - ◇ 男女共に若年者の年齢誤差が小さく、大人が大きくなる傾向

Table 1 年齢推定タスクの実験結果. 年齢誤差の平均と標準偏差. 緑色は年齢誤差が小さく、赤色は大きいことを示す.

	女性	男性
1歳~4歳	6.06 (3.32)	7.10 (5.69)
5歳~8歳	4.44 (3.06)	4.83 (3.22)
9歳~12歳	4.19 (2.51)	5.73 (3.45)
13歳~16歳	5.71 (3.20)	7.08 (2.46)
17歳~20歳	7.52(1.85)	5.82 (2.43)
21歳~30歳	11.85 (1.20)	9.49 (1.78)
31歳~40歳	17.69 (2.61)	12.97 (1.61)

2-P-5

2-P-5

Cocktail Machine Speech Chain: 重複あり音声を用いた 音声認識・音声合成モデルの統一的学習

Cocktail Machine Speech Chain:
Unified training of speech recognition and synthesis models using overlapped speech

☆松永裕太, 高道慎之介(東大), 上乃聖(名工大), 猿渡洋(東大)

- ◆背景:
 - 対話では、相手の発話に重複する相槌などが重要な役割を果たす。
 - 対話における重複あり音声の認識・合成技術が必要である。
- ◆提案法:
 - 重複あり音声を用いて ASR, TTS モデルを同時学習する。
 - 膨大な言語資源を用いた ASR モデルの性能向上を目指す。

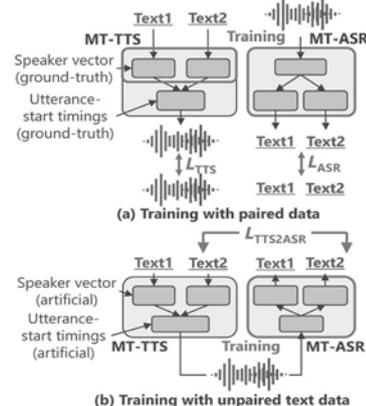


図 1: モデル構成・学習法の概要. 対データ (a), 非対テキストデータ (b) での学習.

2-P-6

2-P-6 相槌生成と話者交替のモデルを組み込んだリアルタイム音声対話システムの構築と評価

Constructing and Evaluating a Real-time Spoken Dialog System
Incorporating Models of Back-channel Responses and Turn-Taking

☆今本麟太郎, 西村良太(徳島大), 太田健吾(阿南高専), 北岡教英(豊橋技科大)

- ◆本研究で提案する音声対話システムの概略図を Fig.1 に示す。
- ◆本システムの目的は、言語特徴量に比べて解析が高速である音響特徴量のみをもとに深層学習モデルによって相槌・応答を行うことである。
- ◆応答のタイミングの決定は Wav2Vec2ForSequenceClassification をもとにした話者交替予測モデルによって行い、相槌のタイミングの決定は LSTM をもとにした相槌タイミング予測モデルによって行う。

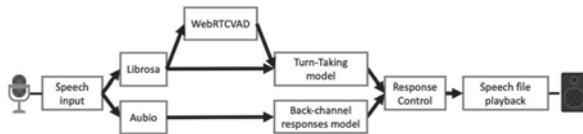


Fig.1: The overview of the proposed system

- ◆ベースラインの音声対話システムとして、音声認識によって応答のタイミングを決定するシステムを構築した。被験者が提案システムとベースラインシステムとの対話・評価を行った。結果として Table1 が得られ、相槌のタイミングが適切であるという評価が得られた。

Table1: Average of rating values for questions

Questions	Scale	Proposed	Baseline
Response timing	1:too late - 5:too fast	3	3
Back-channel timing	1:too late - 5:too fast	3.16	3.83
Back-channel Frequency	1:too few - 5:too many	2.3	4.5

2-P-8

2-P-8 マルチモーダル雑談対話データセットにおける対話相手の音声に対する聞き手の頭部動作の解析

Analysis of Listener Head Motion to Partner Speaker's Voice in Multimodal Dialogue Dataset

○三河多聞(豊橋技科大), △藤井康寿(Google Inc.), 太田健吾(阿南高専), 若林佑幸(豊橋技科大), 北岡教英(豊橋技科大)

- ◆マルチモーダルな対話において、頭部動作はコミュニケーション上の重要な機能を持つ。例えば、話し手の発話に対し、聞き手は頷きや首を傾げる動作などによって傾聴や理解度を表現することが知られている。
- ◆本研究では、雑談対話における聞き手の立場で自然な頭部動作を生成可能な対話エージェントを実現するため、実際の人間同士の雑談対話における頭部動作の映像と各話者個人の音声を含むマルチモーダル対話データセットを Zoom 上で収録した。
- ◆また、収集したデータセットを用いて自己回帰型の深層学習モデルを構築し、自然な頭部動作の生成に関する検討を行った。



Fig. 1 Result of motion generation. Upper: Ground Truth, Lower: Transformer like LSTM model.

2-P-7

2-P-7 円滑な音声対話システムのための発話終了時間の事前推定

Prior estimation of utterance end time for smooth spoken dialogue systems

☆金井良賢, 若林佑幸(豊橋技科大), 西村良太(徳島大), 北岡教英(豊橋技科大)

目的

- ◆音声対話システムで自然な会話を実現したい。
- ◆人の発話がいつ終了するのかを事前に推定することで、オーバラップを含んだ自然なタイミングでシステムの発話を可能にする。

手法

- ◆過去の発話を見ながら常に相手の発話が終了するまでの残り時間を推定する機械学習モデルを作成する。
- ◆発話の終了時間予測を入力音声長や入力特徴量を変更して 5 クラス分類で行い分類精度を確認する。

結果

- ◆対角成分の数値が大きくなっており、推定が有効に行われていることが分かる。

Table 1 Confusion matrix of the classification results (samples) when input speech length is 3 seconds and input feature is MFCC.

		Estimated value				
		0.0-0.2 sec	0.2-0.4 sec	0.4-0.6 sec	0.6-0.8 sec	0.8-1.0 sec
True value	0.0-0.2 sec	9,352	3,680	1,698	1,239	1,671
	0.2-0.4 sec	6,187	4,450	2,653	2,067	2,284
	0.4-0.6 sec	4,457	3,974	3,103	2,797	3,311
	0.6-0.8 sec	3,700	3,266	3,016	3,080	4,580
	0.8-1.0 sec	3,261	2,923	2,734	3,033	5,690

2-P-9

2-P-9 深層情報埋め込み・検出に基づくプロアクティブ型ディープフェイク音声検知

Proactive deepfake speech detection based on deep information embedding and detection

☆尹道鉉, 戸田智基(名古屋大学)

- ◆従来のディープフェイク音声の検知手法は、限定されたデータを用いて学習するため、新しい音声生成システムの結果には対応しづらい。
- ◆本研究では、音声生成システムに依存せずディープフェイク音声の検知が可能なプロアクティブ手法を提案する。本手法は、マーカーと検知器の二つの学習可能な深層ニューラルネットワークからなる。
- ◆多様な音響的歪みにも頑健なマーカーの生成・検知ができた。
- ◆音声変換等によって変形されやすいマーカーの特徴を利用し、ディープフェイク音声と自然音声を識別できるかについて検証を行った。

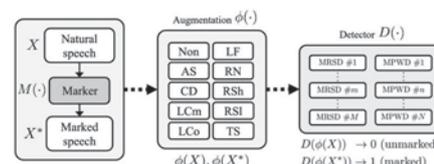


Fig. 1: Training flow of the proposed method.

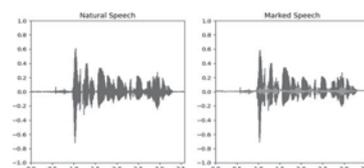


Fig. 2: Waveforms of a natural speech (left) and a marked speech (right). The orange colored shows the difference between the two.

2-P-10

2-P-10 なりすまし音声の検出に有効な音響特徴量の調査

Investigation of acoustic features for the detection of spoofed speech

☆根本結菜(日本工業大・基幹工), 大田健雄(日本工業大・基幹工)

- ◆従来の研究でなりすまし音声の検出に有効な音響特徴量は、メル周波数ケプストラム係数(MFCC)、線形周波数ケプストラム係数(LFCC)、定Q変換(CQT)が有効であると示されているが、具体的にどの周波数が有効であるか明確ではない。
- ◆本研究は音声生体認証のセキュリティ向上を目指し、本物と偽物の音声を区別するために重要な音響特徴量及びその周波数を調査した。
- ◆実験には、ASVspoof2021_DFタスクの評価用音声データを活用。MFCC(20チャンネル)、LFCC(20チャンネル)、CQT(7オクターブ分、1オクターブを12分割した84チャンネル、基本周波数(F0)の計125次元の音響特徴量を用いて、ランダムフォレストにより本物・偽物に分類し、各音響特徴量の重要度を算出した。
- ◆重要度上位10位をTable 1に示す。F0の重要度が最も高く、次にCQTの重要度が高かった。CQTの周波数としては、主に基本周波数周辺が含まれていた。

rank	features	importance[%]	rank	features	importance[%]
1	F0	3.622	6	CQT(87.307[Hz])	1.219
2	CQT(130.813[Hz])	1.370	7	CQT(164.814[Hz])	1.152
3	CQT(116.541[Hz])	1.364	8	CQT(110.000[Hz])	1.114
4	CQT(123.471[Hz])	1.296	9	CQT(103.826[Hz])	1.087
5	CQT(138.591[Hz])	1.258	10	CQT(92.499[Hz])	1.076

使用した音声データのリンクを以下に示す:
ASVspoof2021 Challenge Speech Database (URL: <https://convo.asy.nict.ac.jp/ASVspoof2021/ASVspoof2021.html>)

2-P-12

2-P-12 日本語テキストと音楽の対照学習の実験的評価

Experimental evaluation of contrastive learning between Japanese text and music

◎蓮実拓也, 小松達也, 藤田雄介, 二又航介, 橋健太郎(LINE ヤフー)

- ◆日本語テキストと音楽で対照学習をし、得られたクロスモーダル埋め込みに基づきテキストで音楽推薦をした際の性能を評価する。また、テキストエンコーダの違いが与える推薦性能への影響を調査する。
- ◆独自コーパスを用いた音楽推薦の結果から、構築された埋め込みを用いて日本語テキストによる推薦が可能であり、意味抽出能力の高いテキストエンコーダが音楽との対照学習に有効であることを示す。
- ◆日本語訳した MusicCaps を外部評価用データとした場合の推薦結果から、テキスト間のドメイン差に対する頑健性の課題を示す。

Text encoder	R@1	R@5	R@10	mAP@10	medR
BERT	17.19	38.17	46.35	25.88	14
GLuCoSE	20.27	43.27	53.23	30.07	8
Random	0.06	0.30	0.59	0.17	844

Table 1 Music recommendation results of internal corpus. N = 1687.

Training dataset	medR
Internal	137
MusicCaps	4

Table 2 Music recommendation results of MusicCaps. N = 419.

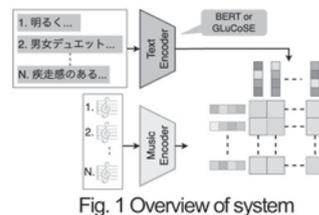


Fig. 1 Overview of system

2-P-11

2-P-11 対照学習モデルによる音声-声質表現文の埋め込み表現獲得

Latent representation for pairs of speech and speech characteristics description trained by contrastive learning

◎渡邊 亞椰, 高道 慎之介, 齋藤 佑樹, 中田 亘, 辛 徳泰, 猿渡 洋

(東大院・情報理工)

- ◆声質を表現する自由記述(以降、声質表現文)によるテキスト音声合成(text-to-speech: TTS)の声質制御需要は高い。
- ◆TTSのサブモジュールとして使用することを目的とし、本研究では声質表現文により声質を指定するモデルの構築を行う。
- ◆対照学習モデルCLAPをベースとした音声・声質表現文埋め込みモデルとして構築し、声質の特徴を予測する学習を取り入れることで効率的に声質を学習させる。
- ◆主観・客観評価によりモデルの性能と課題を評価する。

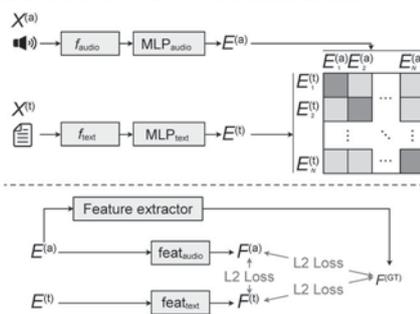


Fig. 1: CLAP overview (top) Feature prediction training overview (bottom)

2-P-12

2-P-13

2-P-12 日本語テキストと音楽の対照学習の実験的評価

2-P-13 Speech Emotion Recognition with Audio and Video Features Using Cross-Attention Transformer

☆ Jian Zhang, Nobuaki Minematsu, Daisuke Saito (Univ. Tokyo)

- ◆日本語テキストと音楽で対照学習をし、得られたクロスモーダル埋め込みに基づきテキストで音楽推薦をした際の性能を評価する。また、テキストエンコーダの違いが与える推薦性能への影響を調査する。
- ◆独自コーパスを用いた音楽推薦の結果から、構築された埋め込みを用いて日本語テキストによる推薦が可能であり、意味抽出能力の高いテキストエンコーダが音楽との対照学習に有効であることを示す。
- ◆日本語訳した MusicCaps を外部評価用データとした場合の推薦結果から、テキスト間のドメイン差に対する頑健性の課題を示す。

- ◆Multi-modal Speech Emotion Recognition (MSER) with audio and video features.
- ◆Cross-Attention Transformer for fusing two modalities and Self-Attention Transformer for capturing intra-modal dynamics.
- ◆Proposed the AV_CAT-SAT model, which effectively integrates audio and video features using CAT and SAT.
- ◆Video features are more effective in detecting happy emotions compared to audio features.

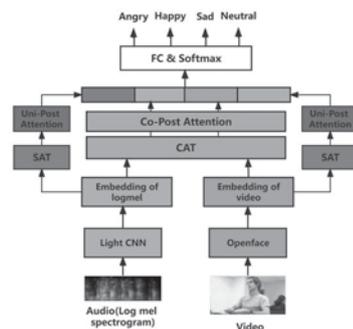


Fig. 1: Block diagram of our proposed model AV_CAT-SAT

2-P-14

2-P-14 日本語フォント文字の印象評価と感情字幕生成への応用

Impression evaluation of Japanese font characters and its application to emotional subtitle generation

☆中村史也(神戸大), 相原龍(三菱電機), 高島遼一, 滝口哲也(神戸大),
△伊谷裕介(三菱電機)

- ◆ テレビ番組や YouTube などにおけるテロップのように、話者の感情をフォントとして反映した字幕を生成することで発話の理解を補助することができると考えられる。
- ◆ 本研究では、細かい感情の差異を表現するフォント字幕を生成するため、日本語フォント文字の印象評価を実施する。評価によって得たフォントと感情の対応データを用いて感情次元からフォントを生成するモデル (Fig. 1) を提案し、連続的な感情変化に対するフォント文字の生成を試みる。
- ◆ 実験の結果、提案手法は従来手法と比較して、多様な特徴(線の太さ、丸みなど)をもつフォントを生成できることが確認できた。

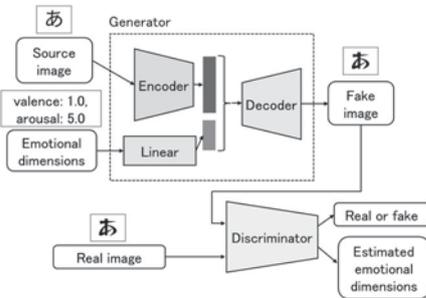


Fig.1: Model structure of the proposed method

2-P-16

2-P-16 テキスト特徴を用いた音声感情認識の多言語化に関する検討

Cross-language Adaptation of Speech Emotion Recognition Using Text Features

☆町田 裕璃奈, 山田 武志(筑波大), 石塚 賢吉, 橋本 泰一 (RevComm Inc.)

- ◆ 従来、発話内容テキストを入力とする音声感情認識手法において、ある言語で学習されたモデルを別の言語にドメイン適応することは困難であった。
- ◆ この問題を解決するためには、ドメイン適応前後で共通のテキスト埋め込み空間を用いる必要がある。
- ◆ そこで本稿では、機械翻訳を用いる手法と多言語 BERT を用いる手法を提案する。前者は機械翻訳を用いて適応前の埋め込み空間を利用できるようにし、後者は多言語で学習した埋め込み空間に統一する。
- ◆ 7つの感情音声コーパスを用いて実験を行った結果、提案手法は理想条件(学習と評価が同じドメインの場合)に迫る精度を達成し、言語の差異を含むドメイン適応を可能にした。また、多言語 BERT を用いる手法の方がより高い精度が得られることを確認した。

Table1 WA for each method (the language code is shown in parentheses.)

Method	Pretrain	Fine-tune	Test	WA (%)
Ideal Condition	JTES (JA)	-	JTES (JA)	83.97
Translation	IEMOCAP (EN)	-	JTES (JA)	57.23
M-BERT	IEMOCAP (EN)	-	JTES (JA)	65.81
Translation	IEMOCAP (EN)	JTES (JA)	JTES (JA)	65.64
M-BERT	IEMOCAP (EN)	JTES (JA)	JTES (JA)	84.58
Ideal Condition	CaFE (FR)	-	CaFE (FR)	69.29
Translation	IEMOCAP (EN)	-	CaFE (FR)	43.82
M-BERT	IEMOCAP (EN)	-	CaFE (FR)	35.93
Translation	IEMOCAP (EN)	CaFE (FR)	CaFE (FR)	66.70
M-BERT	IEMOCAP (EN)	CaFE (FR)	CaFE (FR)	68.05

2-P-15

2-P-15 大規模言語モデルを用いた自由記述型音声感情認識の検討

Free-Form Speech Emotion Recognition Using a Large Language Model.

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

- ◆ 音声に含まれる話者の感情状態を自由記述文として認識する自由記述型音声感情認識タスクの提案と、大規模言語モデルを用いた文生成モデル (Fig. 1) による実現可能性の検証を行う。
- ◆ モデル学習時には、話者性別や音量などの音声情報に関する複数種類の文生成タスクを学習させる音声インストラクション学習を行う。
- ◆ 感情カテゴリラベルから生成した感情状態の自由記述文を用いた実験の結果、主要・副次的な感情を表す自由記述文をある程度生成できること、音声インストラクション学習が有効であることが示された。

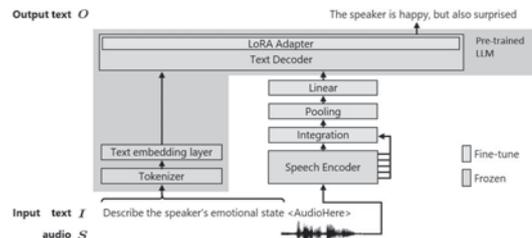


Fig.1: The structure of the text generation model, MiniGPT-Speech.

Table 1: Performances of the emotional state descriptions and the classification tasks.

model	Train tasks			Emo. state description			Major emo.		Gen		
	EmDns	MaJEm	SpCls	BLEU-1	ROUGE-L	METEOR	WA	UA	WA	UA	WA
8-class SER	-	-	-	0.409	0.685	0.765	-	-	-	-	-
MiniGPT-Speech	✓	✓	✓	0.426	0.687	0.758	57.0	31.0	-	-	-
	✓	✓	✓	0.433	0.693	0.767	55.9	30.5	98.8	84.7	82.8
							53.8	30.9	-	-	-
			(indiv)	-	-	-	-	-	99.0	87.9	84.6
				-	-	-	-	-	98.6	89.7	85.6

2-P-17

2-P-17 音声感情認識のための音響特徴とテキスト特徴のデータ拡張に関する検討

A Study on Data Augmentation of Acoustic and Text Features for Speech Emotion Recognition

☆小林 夏美, 山田 武志(筑波大), 石塚 賢吉, 橋本 泰一 (RevComm Inc.)

- ◆ 音声感情認識 (SER) は、人の発話音声を入力として、喜び、悲しみ、平静といった感情クラスを出力する技術である。
- ◆ SER における学習データは感情音声とその感情ラベルであるが、これらのデータが平静音声に比べて非常に少ないという課題がある。
- ◆ 本稿では、音声とその発話内容テキストを入力とする SER 手法を対象とし、潜在空間における音響特徴とテキスト特徴のデータ拡張手法を提案する。
- ◆ 提案手法では、音声から抽出した潜在ベクトルを拡張対象とすることにより、従来は困難であったテキスト特徴のデータ拡張を可能とする。
- ◆ 音声感情認識のベンチマークである IEMOCAP データセットを用いて実験を行い、提案手法の有効性を確認した。

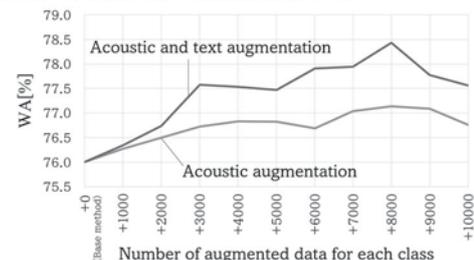


Fig.1: Experimental results

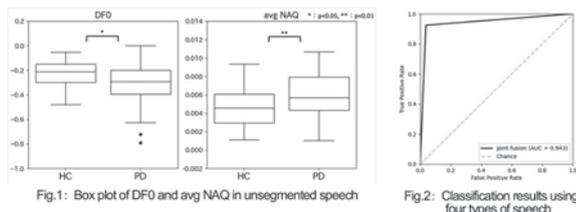
2-P-18

2-P-18 パーキンソン病患者の会話音声における音素・音節の比較分析

A comparative analysis of phonemes and syllables in conversational speech of patients with Parkinson's disease

☆丸山由華, 入部百合絵(愛県大), 北岡教英(豊技科大), 横井克典(国立長寿医療研究センター), 勝野雅央(名大)

- ◆49~75 歳(平均年齢 62.3 歳)のパーキンソン病(PD)患者, 健常者(HC)の計 106 名を対象に会話音声を収集し, 文・音節・音素単位に分割した音声を用いて PD と HC の音響的特徴量の違いを分析/分類した.
- ◆文・音節・音素単位に分割していない音声に対し, PD と HC 間で検定した結果, DF0(基本周波数の 1 階微分)で有意差が認められた. 同様に声門気流の滑らかさや急激な変化を捉える NAQ の平均値に対しても有意差が認められた. (Fig.1).
- ◆音節・音素単位の音声では, 文単位音声と比較して構音に関わる特徴量, 発声に関わる特徴量に対して, 有意差の認められた特徴量が増加した.
- ◆分割していない音声, 文単位音声, 音節単位音声, 音素単位音声の単位毎に構築した学習器を統合した joint fusion を行った結果, AUC 0.943 の精度が得られた (Fig.2).



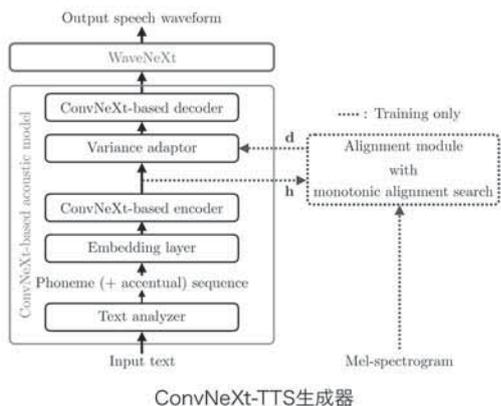
2-P-20

2-P-20 ConvNeXt型超高速 End-to-end音声合成・声質変換モデル

ConvNeXt-based very fast end-to-end neural text-to-speech and voice conversion models

○岡本拓磨¹, 大谷大和¹, 戸田智基^{2,1}, 河井恒¹
(¹情報通信研究機構, ²名古屋大学)

- ・提案法: ConvNeXt型End-to-end音声合成・声質変換モデル
 - ・Transformerを使わないTTSおよびVC音響モデルを提案
 - ・エンコーダおよびデコーダにConvNeXtを導入
 - 従来型Transformerよりも3倍の高速化および高品質化を実現
- ・ConvNeXt型ニューラルボコーダWaveNeXtを導入
 - RTF0.05の超高速生成を達成しつつ高品質な合成を実現



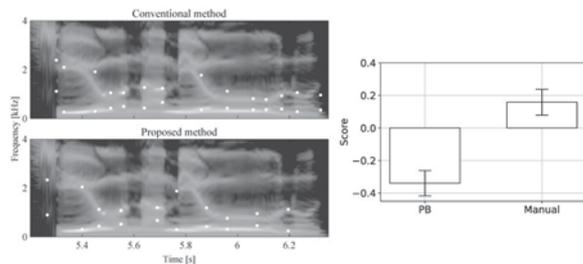
2-P-19

2-P-19 音声モーフィングにおける自動対応付けの提案と品質評価

Proposal and quality evaluation of automatically estimation of anchor points in speech morphing

☆堀部貴紀, 森勢将雅(明治大)

- ◆音声モーフィングにおける時間軸方向の対応付けでは, フォルマントが変化する時刻に対応点を打つことが望ましい.
- ◆従来手法では音素境界が用いられてきたが, フォルマントが変化する時刻とは限らず, モーフィング音声の品質も手動の場合に劣る.
- ◆本研究では, 動的時間伸縮と短時間パワーを用いた手法を提案する.
- ◆図1より, 提案手法はフォルマントが変化する時刻付近に対応点があることが確認できる.
- ◆また, 従来手法の制約である音素系列情報を必要としない.
- ◆図2で示すように, 手動 (Manual), 提案手法, 従来手法 (PB) の順に自然性が高いと判断された. 特に, 従来手法との間に有意差が認められたことから, 提案手法は対応付け手法として有効であるといえる.



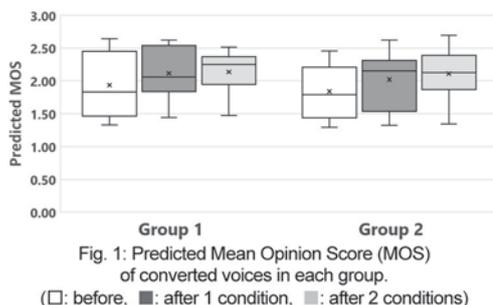
2-P-21

2-P-21 リアルタイム音声変換における聴覚フィードバックの影響に関する調査

Investigation of effects of auditory feedback in real-time voice conversion

☆丹羽希碩, 小林和弘, 戸田智基(名古屋大学)

- ◆リアルタイム音声変換システムでは, 自身の音声に加えて, 変換音声についても聴覚フィードバックを得ることができる.
- ◆目的: 変換音声の聴覚フィードバックが変換音声の品質へ与える影響および遅延時間や音量に着目した聴覚フィードバック活用法の調査
- ◆録音された変換音声の聴覚フィードバックを聴くことができる場合と, 加えて発話中にリアルタイムで変換音声の聴覚フィードバックを聴くことができる場合の計2通りの条件で発話練習実験を実施した.
- ◆被験者の変換音声を収録し, 推定平均オピニオン評点により変換音声の品質評価を行った.
- ◆推定平均オピニオン評点の結果, 発話練習による自然性の向上が見られることから, 話者が発話の仕方を調整することでリアルタイム音声変換の変換音声の品質を改善できることが分かる.



2-P-22

2-P-22 拡散確率モデルを用いたノンパラレルな Any-to-many 声質変換

Non-parallel any-to-many voice conversion using a diffusion probability model.

☆ 畠山瑠一, 奥田耕平, 中鹿亘 (電通大)

- ◆ 拡散モデルを用いた声質変換がVAEやGANを上回る精度がでている
- ◆ ノンパラレルデータの学習で、任意の話者を複数の話者へ変換する Any-to-Many の声質変換に対応しているものが少ないが、VoiceGrad は対応している
- ◆ VoiceGrad は離散時間の拡散モデルであるスコアベースモデル(SBM)を用いた声質変換である
- ◆ 本研究では VoiceGrad から派生し、連続時間化した拡散モデル(VE-VPSDE)を用い、更にエンコーダーにルールベースの拡散過程を追加したノンパラレルな Any-to-Many 声質変換を提案
- ◆ 従来手法よりもより品質の高い声質変換となった

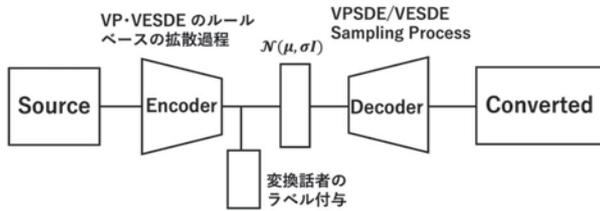


Fig.1: VC model Sampling

2-P-24

2-P-24 シャドーイングに基づく外国語音声に対する聴取崩れの主観的ラベリング

Subjective labeling of shadowing disfluency of L2 speech

☆ チェ・ジェヒョン, 張露元, 高英翔, 峯松信明, 齋藤大輔, (東大・工), 中西のりこ(神戸学院大)

- ◆ シャドーイングを用いた客観的聴取崩れラベル (従来手法) は「音声内容」を基にスコアリングし、「意味的」比較している訳ではない。
- ◆ Can't / Cannot のように音声内容が微妙に誤っても意味的に同様であることもあるが、ache / sake のように軽微な誤りが重大な意味的誤りを生じる場合もある。
- ◆ 「意味的」聴取崩れの主観的ラベリングを遂行し、以下を検討する
- ◆ 1)ラベリングする評価者の選定基準、2)客観ラベルとの合致度、3)従来の客観ラベルから聴取崩れ誤りの検出で用いた閾値の妥当性

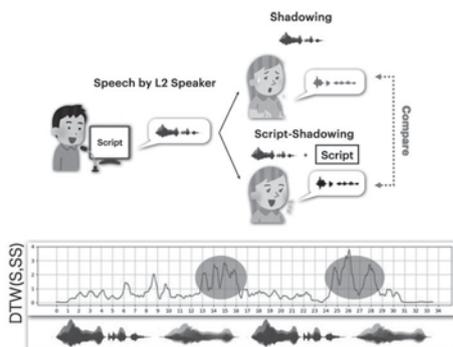


Fig.1: Measuring intelligibility of L2 speech based on shadowing technique

2-P-23

2-P-23 Non-Parallel Limited Data Emotion Voice Conversion with Variance Adapter and Non-Autoregressive Decoder

韻律予測と非自己回帰型デコーダーによる 低資源データの非並列的感情音声転換

☆ Zhanhang ZHANG (JAIST), Sakriani SAKTI (JAIST)

- ◆ Proposes a sequence-to-sequence model for Emotional Voice Conversion (EVC) tasks with limited non-parallel emotional speech data.
- ◆ Utilizes a variance adaptor to predict prosodic features such as pitch, energy, and phoneme duration to address emotion variability.
- ◆ Introduces an emotional vocoder conditioned on the emotion ID.
- ◆ Experimental results show that our proposed emotional voice conversion framework can achieve better emotional expression and inference speed than the baseline.

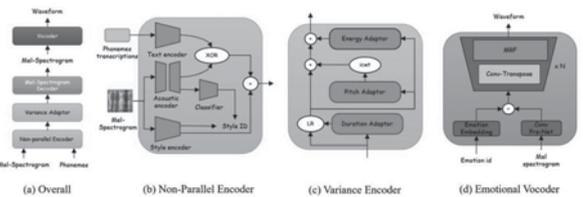


Fig 1. Proposed Model Structure

2-P-25

2-P-25 Comparing Learning Methods for Adaptation to Accented English

☆ Kentaro OGI and Ian WILSON (University of Aizu)

Despite the global demand for diverse English proficiency in universities worldwide, Japanese students have limited opportunities to use English in their daily lives, and resources for non-native English are scarce. This paper presents a Moodle course designed to expose learners to various English accents through activities like shadowing and quizzes. While past research highlights shadowing's effectiveness in single-accent adaptation, our experiment explores this with eight non-native and five native accents.

In a ten-day training experiment with 15 EFL learners, categorized into groups (A: All methods, B: Only shadowing, C: Control), we conducted a comparative analysis of pre- and post-test scores. We also investigated the correlation between training duration and score enhancement. The results showed that participants who did all training methods exhibited the most improvement, but not significantly different from those who engaged in shadowing alone. A positive correlation between training duration and score enhancement was observed, though not statistically significant. Shadowing multiple accents in a short period enhanced EFL listeners' perception of those accents. However, unlike the past study conducted with a single accent, this difference was not statistically significant.

2-P-26

2-P-26 Contexts and Biomechanics of Native English Speakers' /r/ Tongue Shapes

☆Daichi ISHII and Ian WILSON (University of Aizu)

Generally, native speakers of North American English pronounce rhotic sounds (/r/) by using two different articulations: retroflex [ɻ] and bunched [ɹ]. Previous research has tried to discover patterns of when they use retroflex or bunched, but the number of stimuli or contexts were limited, so such studies have been largely unsuccessful. In this research, we focus on the vowels and consonants before and after /r/, and try to discover if there are clear patterns. We created a list of 229 words containing possible vowel and consonant combinations [(C)rV, Vr(V), and Vr(C)]. Results showed that the consonant before /r/ has a fairly consistent effect on /r/. Retroflex /r/ occurs more often in pre-vocalic than post-vocalic position. Also, when /r/ is word-initial or follows a labial consonant, such as /p/, /b/, /m/, /f/, it may be retroflex. However, this seems to depend on the vowel after /r/. If that vowel is /ɪ/ or /i/, which are high front vowels, then the /r/ is often bunched. For English learners, it is difficult to know when to use one or the other because the acoustic differences are difficult or impossible to perceive. However, from a biomechanical perspective, it may be easier for learners to produce the type of /r/ that is used by native speakers in a given context.

2-P-28

2-P-28 歌唱音声合成における F0 の自然性向上のための Diffusion-GAN モデルの検討

Study of Diffusion-GAN Model for Improving Naturalness of F0 in Singing Voice Synthesis

☆芦田 裕飛, 中鹿 亘 (電通大)

- ◆ 歌唱音声合成における F0 の予測は非常に難しいタスクである
- ◆ 精度と速度がトレードオフの関係にある
 - アプリケーション化などの実用化に難がある
- ◆ 昨今は Diffusion モデルがよく用いられる
 - Diffusion モデルを用いた F0 予測も提案された
- ◆ Diffusion モデルは精度は高いが、速度が遅い
- ◆ TTS 分野では Diffusion-GAN モデルによる高速かつ高精度なモデルが提案された
- ◆ F0 においても同様に活用できないか、検討した

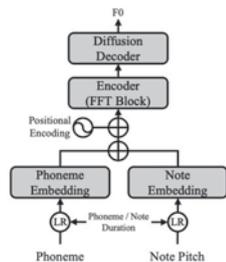


Fig. 1: Proposing architecture

2-P-27

2-P-27 音素継続長を考慮した早口講義音声の話速変換とその効果

An investigation of speaking rate conversion for fast-talking lecture speech based on phoneme duration.

☆田中 美久登, 西村 竜一 (和歌山大)

- 大学の講義を収録した講義音声を対象として、音声の音素継続長を基準とした早口音声の話速変換手法を提案
- WORLD を用いた発話の伸縮・伸長処理について 8 条件を比較
 - (a) すべての音素に対し見本音声の音素継続長に揃えた伸長
 - (b) すべての音素に対し見本音声の音素継続長に揃えた伸縮
 - (c) 母音に限り見本音声の音素継続長に揃えた伸長
 - (d) 母音に限り見本音声の音素継続長に揃えた伸縮
 - (e) 名詞・動詞の区間を伸長 (f) デクリネーションを考慮した伸長
 - (g) 音圧の変化を考慮した伸長 (h) 音高の変化を考慮した伸長
- 主観評価 (5 段階・協力者 8 名) の結果、母音に限った条件 (c, d) と条件 (e~h) から発話速度の改善傾向を確認

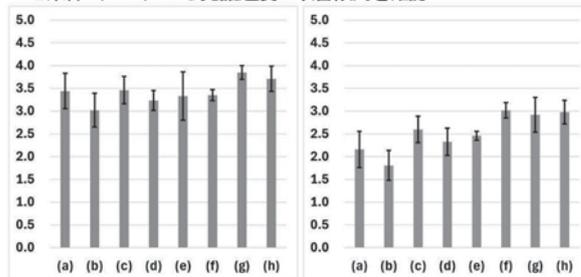


Fig. 1: 左「音声B(処理後音声)は音声A(処理前音声)と比べて内容理解に適した発話速度である」、右「音声Bは音声Aと比べて講義の受講により支障がない音質である」

2-P-29

2-P-29 歌声合成と変換を用いた未経験者音声からのオペラ歌唱音声合成

Opera-singing voice synthesis from inexperienced voice using singing voice synthesis and conversion

☆菅原 碧斗¹, 岸本 宗真², 足立 優司², 田井 清登², 高島 遼一¹, 滝口 哲也¹

(¹神戸大学, ²メック株式会社)

- ◆ 歌声合成手法 DiffSinger によってプロオペラ歌唱音声を生成し、生成した音声さらに声質変換手法 Diff-SVC によってユーザの声質へ変換することにより、**オペラ未経験ユーザの音声で任意のオペラ曲の歌唱合成を行うシステム**を検討する。
- ◆ DiffSinger を用いてオペラ歌唱を合成すると、プロのオペラ歌唱と同様に、2.2kHz~4.0kHz 帯の**中高音域のエネルギーが強調されていることが確認出来た**。
- ◆ Diff-SVC を用いてユーザの声に声質変換した際、**中高音域のエネルギーが失われている**が、中高域強調ネットワーク(M-HEN)を導入することで、プロのオペラ歌唱と同様に、**中高音域のエネルギーが強調されていることが確認出来た**。

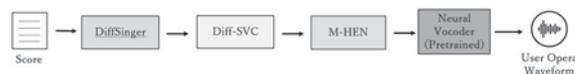


Fig. 1 Overview of the proposed the user's opera-singing synthesis system.

2-P-30

2-P-30 FaderNetworks を用いた F0 変換による歌唱技術の付与

Giving singing technique by F0 conversion using FaderNetworks

☆後藤純平, 中鹿亘(電通大)

- ◆機械学習の手法である FaderNetworks や声質変換の技術を用いて、既にある歌声に対してビブラートやしやくりなどの歌唱技術を付与する手法を検討
- ◆WORLD ボコーダーで抽出した F0 周波数を encoder/decoder と discriminator が敵対的に学習し、話者に非依存な特徴量として歌唱技術をピックアップ
- ◆複数の話者間で変換を行ったところ、再構築された F0 に比べて変換を行なった F0 は特にアタック部分により波形の揺らぎがみられた

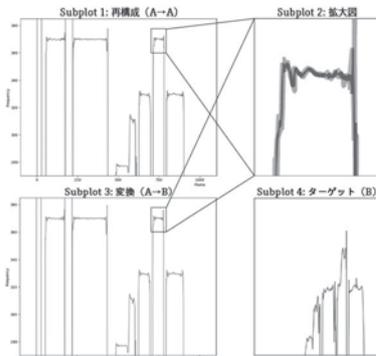


Fig.1: 再構成された歌声と変換された歌声のF0周波数の比較

2-P-32

2-P-32 統計的基本周波数推定のためのデータ拡張の検討

Data augmentation for statistical fundamental frequency estimation

◎小口 純矢, 森勢将雅(明治大)

DNNピッチ推定 (CREPE) の性能をデータ拡張で上げられるか検証しました

DNNピッチ推定 (CREPE) の性能をデータ拡張で上げられるか検証しました

DNNピッチ推定 (CREPE) の性能をデータ拡張で上げられるか検証しました

DNNピッチ推定 (CREPE) の性能をデータ拡張で上げられるか検証しました

2-P-31

2-P-31 歌唱音声の基本周波数に含まれる微細変動の増減処理によるビブラート知覚への影響の調査

A study on effect on perception of vibrato by increase and decrease processing of fine fluctuation of fundamental frequency in singing voice.

◎鈴木千文, 坂野秀樹, 旭 健作(名城大)

- ◆本研究ではこれまでビブラート音声における微細変動の聴感上の影響を調査
 - ビブラートの変動の影響が微細変動の影響よりも大きい
 - 分析対象をノンビブラート音声に変え、同様の調査を行う
- ◆以下の手順でノンビブラート音声の基本周波数の微細変動を抽出した
 - 基本周波数から算出された回帰直線の傾きである微分信号を算出
 - 微分信号に10Hzのハイパスフィルタを用いて微細変動を抽出
 - Fig.1 左より、微細変動にビブラートに類似した変動を観察することができる
- ◆微細変動に含まれる変動と基本周波数との関係性
 - 基本周波数の微分信号の振幅スペクトルを観察 (Fig.1 右)
 - 微細変動に含まれるビブラートに類似した変動は、歌手によっては基本周波数の振幅スペクトルのピークに対応した変動である可能性が考えられる

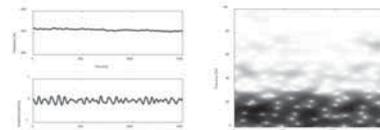


Fig.1. Fundamental frequency and fine fluctuation (Left). Spectrogram of fundamental frequency (Right).

2-P-33

2-p-33 42ch 球状マイクロホンアレイを用いた発話の空間放射特性の計測

Measurement of spatial radiation characteristics of speech using a 42ch spherical microphone array

○牧 勝弘(愛知淑徳大), 饗庭 絵里子(電気通信大), 木谷 俊介(北陸先端大), 天野 成昭(愛知淑徳大)

- ◆42 チャンネル球状マイクロホンアレイを用いて発声の同時多点計測を行い、発声の空間放射特性を調べた。
- ◆その結果、800 kHz 以下の周波数成分は発声者の前方斜め下方向、1 kHz 前後では前方斜め上と下方向の2方向、2 kHz 以上では概ね正前方方向であるが重心は下方向であることが明らかとなった。
- ◆空間放射特性における個人差は3 kHz 以上に顕著に表れた。この個人差は、周波数スペクトル(1chのデータ)からでは分かり難い、様々な声質の特徴を解明する際の音響手がかりとして有用かもしれない。

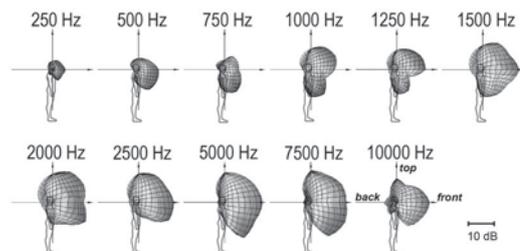


Fig. 1 Spatial radiation characteristics of vocalizations. A three-dimensional display of the average sound pressure level at each frequency of the voice when the speaker utters the phrase "Gaijin-san ha kanpeki shugi dearu." For each frequency, the lowest sound pressure level among the 42 directions is set to 0, and the sound pressure level in each direction is displayed as the difference from there.

2-P-34

2-P-34 音響特性に基づく 音声のポップアウトの程度の推定

Estimation of the degree of pop-out based on acoustic information of speech signal

○坂野秀樹(名城大), 牧勝弘(愛知淑徳大),
河原英紀(和歌山大), 天野成昭(愛知淑徳大学)

- ◆ポップアウトボイスの音響解析を行っている
- ◆複数話者による「外人さんは完璧主義である」100 音声を用いて主観評価実験 (Exp. A) を実施し、各音声にポップアウトスコアを付与
 - 長時間平均スペクトルを、ポップアウトスコアが低い音声と一致させる処理を適用した場合の Exp. B も実施
- ◆時間的同期を取り、ポップアウトスコアとの相関を調査
 - 基本周波数 (左上)、動的特徴量 (右上)、スペクトル (左下)、デルタスペクトル (右下) の場合を観察

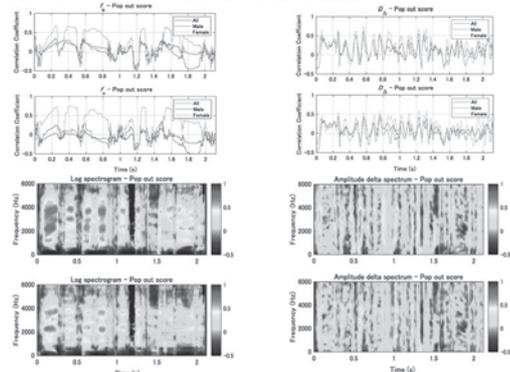


Fig.1: Time series of the correlation coefficient between the pop-out score and each acoustic feature for Exp. A (upper) and Exp. B (lower).

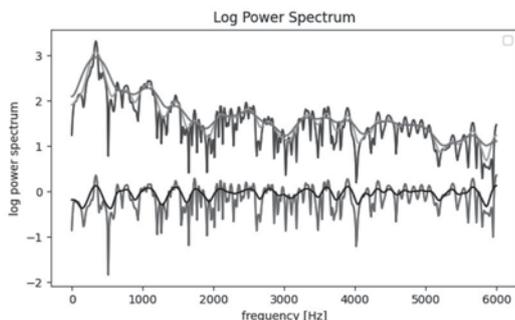
2-P-36

2-P-36 2 種のラグ窓による スペクトル平滑化を用いた F_0 推定

F_0 Estimation by Paired Spectral Smoothing Lag Windows

☆越森道貴¹, 嵯峨山茂樹^{1,2}, 中鹿亘¹(¹電通大/²東大)

音声の F_0 推定の一方法であるラグ窓法の精度と対雑音頑健性を向上した。音声の自己相関関数に幅広および幅狭な窓関数 (ラグ窓) を掛けると、スペクトルは弱くおよび強く平滑化がなされ、それらの比を逆フーリエ変換して自己相関領域に戻すと、ピークが生じる時間遅れ τ が、信号周期として推定できる。この手法により、女声などの F_0 の高い音声や、雑音混じりの音声などにおいても、 F_0 推定精度向上が期待できる。性能評価の予備実験を行った結果、 F_0 の高い音声に対して従来手法と同程度の精度を示し、また雑音に対する頑健性も示した。例えば、有声音とされている下図のようなスペクトルのフレームにおいても、提案手法で推定した F_0 は、Ground Truth 値に一致した。



2-P-35

2-P-35 TLS に基づく時変複素音声分析の検討

*A Study on TV-CAR analysis based on TLS

舟木慶一(琉球大学)

我々は解析信号に対する時変複素 AR(Time-Varying Complex AR: TV-CAR) 音声分析とその音声処理への応用の研究を行っている。近年は l_2 ノルム正規化方式の構築を行い、前回の音響学会では PLS(部分最小二乗法)に基づく方式の提案を行った。一方、全最小二乗法 (Total Least Square: TLS) と呼ばれる方式がある。これは観測行列、ベクトルに誤差が同時に存在するときの線形方程式の近似的な解法であり、1901 年から考えられていたが、1980 年代に入り、Golub と Van Loan が数値解析の観点から、この解法について詳細に解析を行い、全最小二乗法 (TLS) と名付けた。TLS はスペクトル解析、パラメータ推定、適応フィルタリング、ならびにシステム同定などの信号処理問題解決に広く応用されている。今回、TLS 基準 TV-CAR 分析を構築し、音声の F_0 推定を用いて評価した。評価の結果、前世紀の方式である TLS 基準方式は、最新の方式である l_2 ノルム正規化方式を凌駕することはないが、雑音レベルが高い、殊に Pink 雑音重畳時に、良好な結果をもたらすことを確認した。

2-P-37

2-P-37 音声認識による笑い声表現と音声合成の 検討

Toward Laughter Expression and Speech Synthesis Using Speech Recognition.

☆須藤智貴、近藤和弘(山形大)

- ◆ 背景 :
 - 機械学習による Text-to-Speech(TTS)では、合成対象音声の言語情報を文字や音素などで表現し、それを用いて合成モデルの学習などを行う。一方で、笑い声は典型的な言語的パターンや特異な吸気や呼気で生じる音によって多様で適切なテキストラベリング済みデータを作成することが困難である。
- ◆ 目的・実験内容 :
 - 英語音声認識と多言語音声認識の事前学習済みモデルを笑い声に適用して、音声の言語情報の表現を得る。それを用いて音声合成モデルを学習、合成された音声の品質について MCD で客観評価を行い比較する。
- ◆ 結果 :
 - 多言語音声認識モデルが英語音声認識に対して優れた結果が得られた。一方で音声の品質は劣化が強く感じられるものだった。

2-P-38

2-P-38 クラウドソーシングを利用した
叫び声に対する感情次元評価
Crowdsourced emotional dimension evaluation for spontaneous scream and shout

☆大石暖, △大久保港, 有本泰子 (千葉工大)

- ◆背景: 叫び声から知覚される感情を調べるためには感情評価実験が必要であるが、感情評価実験の効率化が求められる
- ◆目的: 感情評価実験のオンラインでの実施による影響を検証する
- ◆分析方法:
オフラインおよびオンラインでの叫び声に感情情報を付与する評価実験
付与された感情評価値を基に階層的クラスタリングによる分類
オフライン実験とオンライン実験の評価値の相関係数の算出と比較
- ◆結果: 叫び声の表現はオフラインでは2種類、オンラインでは4種類に分類された
2つの実験評価値に強い相関がないため、得られる評価結果が異なる可能性が示唆された

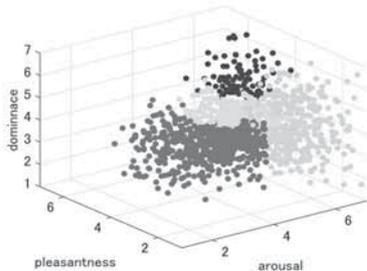


Fig. 1: Results of classification of shout and scream expressions according to the evaluation values of an online experiment.

2-P-40

2-P-40 jMARS Recorder:コーパス朗読に特化した
音声収録アプリの制作と検討
jMARS Recorder: Development and consideration
of a speech-database-focused recording application

☆俣野文義, 森勢将雅 (明治大)

- ◆目的: 初心者でも使いやすいコーパス収録用レコーダーを作る
- ◆リポジトリ: gitlab.com/f-matano44/jmars-recorder



Fig. 1 Screenshot of this application.

- ◆コーパス文の表示: 自作コーパスも読み込み可 (Fig. 1 のように表示)
- ◆リファレンス音声の再生: 発音・アクセントの確認
- ◆半自動トリミング: 発話区間のみを切り出して自動で保存 (修正も可)
- ◆設定の隠蔽: 隠しファイルにすることで初心者による誤変更を防止

2-P-39

2-P-39 クラウドソーシングを用いた大規模比較
評価のための評価ペアの組み合わせと評価数のオ
ンライン最適化

Online optimization of pair combinations and evaluation volumes for large scale preference-based subjective evaluations using crowdsourcing

◎安田裕介, 戸田智基 (名古屋大学)

合成音声の主観比較評価は評価対象が多いとペアの組み合わせが膨大になり評価できない。本稿は比較評価のペアの組み合わせと評価数をオンライン学習を用いて最適化することで、クラウドソーシングによる大規模比較評価を実現する。

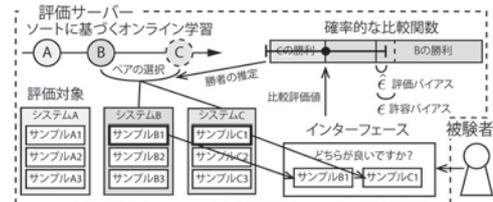


図1: 動的比較評価の概要図。

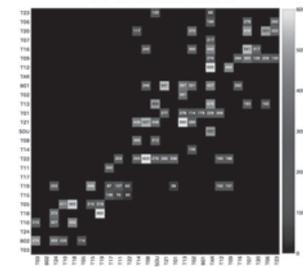


図2: 動的比較評価で評価されたペアの分布と評価数の期待。オンライン学習アルゴリズムは27システムの総組み合わせ351ペアを83ペアに削減し、評価数を不均等に割り当てても評価精度を犠牲にしなかった。