2-1-1 Gated Multi Mini-patch Extractor for pooling in Audio classification

- ☆ Bo He, Shiqi Zhang, Zheng Qiu (Waseda Univ.), Daiki Takeuchi,
 Daisuke Niizumi, Noboru Harada (NTT), Shoji Makino (Waseda Univ.)
- An insufficient pooling will lead to biased statistical estimates, it cannot capture the periodic features of audio signals, and consequently loss some key information.
- ◆To mitigate this concern, we present a simplistic method we called Gated Multi Mini-patch Extractor (GMME) for downsampling time-frames. The End-to-end Audio Transformer (EAT) model was employed as backbone network.
- Simulations indicate that the proposed method successfully retains classification features and decreasing over 98% parameters in contrast to the original Downsampling block, the overall model structure parameter size is also compressed by nearly 40%.

Table 1 Result in ESC50, K and C is respective Kernel size and Channel number. #D-Param is the parameter number of Downsample block, while #T-Param is represent Total model's parameter.

Model	Set	#D-Param(K)	#T-Param(M)	Accurancy(%)
EAT-S-AA		1961.1	5.18	91.25
(conventional)		1901.1	0.18	91.20
EAT-S-GMME (proposed)	K=32,C=8	17.4	3.23	90.80
	K=32,C=16	34.9	3.25	91.10
	K=64,C=16	67.8	3.28	91.30

Table 2 Result in Speech Commands V2. K and C is respective Kernel size and Channel number. #D-Param is the parameter number of Downsample block, while #T-Param is represent Total model's parameter. Model Set #D-Param(K) #T-Param(M) Accurancy(%)

Model	Set	#D-Param(K)	#T-Param(M)	Accurancy(%)
EAT-S-AA		493.1	1.97	97.76
(conventional)		495.1	1.94	91.10
EAT-S-GMME (proposed)	K=32,C=8	17.4	1.50	97.74
	K=32,C=16	34.9	1.51	97.71
	K=64,C=16	67.8	1.54	97.88

2 - 1 - 3

2-1-3 イベント間の共起構造を導入した隠れセミマルコフモデルに基づく音響イベント検出

Sound Event Detection Based on Hidden Semi-Markov Model Dependent on Co-occurrence Between Sound Events

☆吉永朋矢 (早稲田大/産総研), 坂東宜昭 (産総研),

井本桂右 (同志社大/産総研), △大西正輝 (産総研), △森島繁生 (早稲田大)

- ◆目的:誤り訂正性能の高い音響イベント検出 (SED) の後処理の実現
- ◆関連研究:中央値フィルタ・隠れセミマルコフモデル (HSMM)
 - ▶ 中央値フィルタは短時間の誤りをロバストに修正でき広く活用
 - ➤ HSMM はイベントの継続長を考慮でき PSDS などの指標と相性良
- ◆本研究: HSMM に基づく SED の後処理に対する拡張の検討
 - ▶ 音響シーンごとの共起構造を考慮した混合モデルを検討
 - ▶ 実時間システムを想定したオンライン推論を検討
- ◆実験: DESED データセットを用いた HSMM による後処理の評価
 - ▶ 予備実験として、音響シーン数1のHSMMで評価
 - ▶ HSMM は後処理なしの場合より検証用セットで PSDS2 が向上

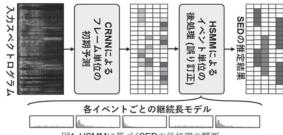


図1:HSMMに基づくSEDの後処理の概要

2-1-2

2-1-2 Time-Frequency Feature Extractor Applied for Audio Classification

☆ Haoran Xing, Shiqi Zhang (Waseda Univ.), Daiki Takeuchi,

Daisuke Niizumi, Noboru Harada (NTT), Shoji Makino (Waseda Univ.)

- Image-based CNN models have been recently used for audio classification tasks. However, two axes in spectrograms denote the time frame and the frequency bin while those in picture images do not have similar physical meanings. Besides, values in spectrograms describe the energy distribution of the audio signal while there is no such physical meaning in picture images. As a result, the image-based model may face difficulty in effectively extracting features from pure spectrograms.
- ◆ To make spectrograms effective for audio classification, we proposed a new Time-Frequency Feature Extractor (TFFE) as a front end to deliver improved spectrograms. Furthermore, we utilized EfficientNetV2 (ENV2) as the backbone of our audio classification model by integrating TEEE_ENV2
- Experimental results show that the proposed TFFE-ENV2 outperformed the original ENV2 on ESC-50 and Speech Commands datasets, proving the effectiveness of the proposed method.

Table1:Comparison of original ENV2 and the proposed TFFE-ENV2

Model Name	Parameters	Accuracy		
	Parameters	ESC-50	Speech Commands V2	
ENV2 (baseline)	8.74M	$88.63\%(\pm 0.34)$	97.74%(± 0.07)	
TFE1-ENV2 (proposed)	8.75M	$89.42\%(\pm 0.43)$	$97.78\%(\pm 0.02)$	
TFE2-ENV2 (proposed)	8.77M	89.23%(± 0.27)	97.79%(± 0.07)	
TFE4-ENV2 (proposed)	8.82M	89.67%(± 0.46)	97.85%(± 0.01)	

2 - 1 - 4

2-1-4 時間的連続性を導入した視聴覚 自己教師あり学習に基づく音響イベント検出

Audio-Visual Self-Supervised Learning With Temporal Continuity for Sound Event Detection

☆大田竹蔵(筑波大,産総研),坂東宜昭(産総研) 井本桂右(同志社大,産総研),△大西正輝(産総研)

- ◆目的:自己教師あり学習 (SSL)に基づく音響イベント検出 (SED)→ アノテーションなしの事前学習で SED の性能を改善
- ◆従来手法: ① Audio SSL ② Audio-Visual (AV) SSL
 - ▶ ①: 異なるクリップ間・異なる時間フレーム間に対する対照学習
 - → SED を含む様々なタスクに応用可能な埋め込みが学習可能
 - ▶ ②: 音と画像の対応関係に基づき対照学習 (AVC)
 - → 音響のみより情報が豊富だが、発生時刻の学習には不向き
- ◆提案手法: AVC と音のフレーム間類似性の学習を併用
 - > クリップ単位の AVC に基づき音響イベントの種類を弱学習
 - ▶ 近傍時刻の音の類似性により発生時刻を学習
- ◆実験結果: URBAN-SED を用いて音響イベント検出に転移学習
 - ➤ 従来の AVC から PSDS1 が改善することを確認

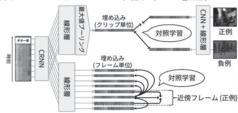


図 1 時間的連続性と視聴覚対応に基づく事前学習

2-1-5 環境音分離における入力混合音と再混合音 の誤差を考慮した損失関数の導入

Introducing A Loss Function Considering Input Mixtures and Remixtures
Errors in Environmental Sound Separation

☆下西 莞太, 福森 隆寛, 山下 洋一(立命館大·情報理工)

- ◆環境音分離:音声や楽音に限らない音の分離
 - > 環境音分析や異常音検知の精度向上に繋がる技術
- ◆分離モデル学習時,分離音と正解音での損失を計算
 - ▶ 理想的な分離が可能なら分離音を再混合させた音は入力混合音に戻る
 - ▶ 現実では分離時に発生する歪みや分離誤りによる干渉音の影響で、入力混合音と異なる可能性がある
- ◆本研究では学習時に再混合音と入力混合音の誤差を考慮した損失関数 Remix loss を提案
 - → 分離音を再混合させると入力混合音に戻るという制約を課すことで、Remix loss が正則化同様の働きをすることを期待

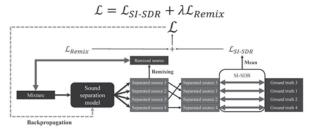


Fig.1: Overview of the proposed method

2-1-7

2-1-7 分散センサに基づくイベント分析のための

Masked self-distillation modeling

Masked self-distillation modeling for distributed sensor event analysis ②安田昌弘(NTT, 東京都立大学), 原田登, 大石康智, 中山彰, 齊藤翔一郎 (NTT), 小野順貴 (東京都立大学)

【分散センサに基づくイベント分析】

- 分散センサによる様々な視点やモダリティの観測を組み合わせることで広範囲にわたって人の行動やその周辺環境を分析
- 本研究では特に分散マイクおよび分散カメラに注目

【課題】

- 分散センサの捉える情報は状況に応じて変化しており、どのセンサの情報が 有用で、それらがどう関連するかをモデル化することが必要
- 実応用上の制約から、センサの欠損に対しても頑健であることが望ましい 【提案法】
- 分散センサの観測をランダムにマスクし、他のセンサ観測に基づいて補完 することを通じてセンサ間の関係のモデル化を行う Masked self-distillation modeling (MELD) を提案

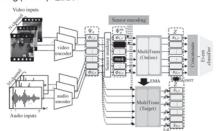


Fig. 1. Masked self-distillation modeling (MELD)

2 - 1 - 6

2-1-6 複数仮説トラッキングに基づく 音響イベント定位・検出

Sound Event Localization and Detection Based on MHT 〇坂東 宜昭 (産総研), 升山義紀 (産総研/都立大) 井本桂右 (産総研/同志社大), 佐々木洋子 (産総研)

- ◆目的: 容易に構築できる音響イベント定位・検出 (SELD) の実現
- ◆従来法: End-to-End 型の深層学習に基づく SELD システム
 - ▶ 多チャネル信号とその空間的アノテーションデータから学習
 - ▶ 方向と種類のアノテーションに膨大なコスト

◆提案法:構築が容易なサブシステムを確率的に統合して構築

- ➤ SELD を音響イベント検出 (SED) と音源定位 (SSL) に分解
- ▶ 各サブタスクの推論結果を複数仮説トラッキング (MHT) で統合
- ▶ 学習はSED サブシステムのみ → モノラル信号のみで学習可能

◆実験結果: STARSS22 検証セットで性能評価

- ➤ SED システムは AudioSet (4,500 時間) のモノラル混合音で学習
- ➤ DCASE 2023 Challenge の E2E 型のベースラインと同程度の性能

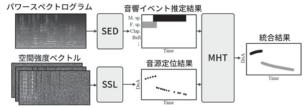


図1. SEDとSSLを後処理で統合するMHT-SELDの概要

2-1-8

2-1-8 効果音生成におけるクラス条件付き 潜在拡散モデルと FAD に基づく後処理

Foley Sound Synthesis with a Class-conditioned Latent Diffusion Model and FAD-based post-filtering

◎シャイブラー ロビン, 蓮実 拓也, 藤田 雄介, 小松 達也, 山本 龍一, 橘 健太郎 (LINE)

- ◆本研究では、効果音生成のためのシステムを提案する。DCASE Challenge 2023 Task 7 のルールに従って、7種類の音を生成する。その種類は犬の鳴き声、足音、銃声、キーボード、移動する車両、雨、くしゃみ・咳である。
- ◆提案のシステムは潜在拡散モデルに基づいてメルスペクトグラムを生成する. 最終的に、HiFi-GAN ボコーダを用いて時間波形に変換する.
- ◆音品質を向上させるために、フレシェ音声距離 (FAD; Fréchet audio distance) に基づいた後処理を提案する.
- ◆比較実験により、Task 7 のベースラインシステムと text-to-audio モデル Tango と比べて、提案法の方が FAD の値が低いサンプルを生成することを示した.

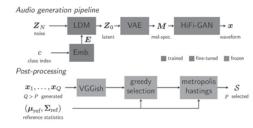


Fig.1:Top: Overview of the audio generation pipeline. Bottom: FAD-based post-processing pipeline.

16チャネル円形スピーカアレイを用いた ポータブルマルチスポット再生システムの 2-1-9 改良

Improving portable multiple sound spot generation system with a compact circular array of 16 loudspeakers

○岡本拓磨 (NICT),

△大山慎二, △上野克司, △岡部司, △谷健太郎 (北日本音響) △須藤美優希, △桑原万苗, △疋田啓太 (NICT)

Q:「以前聞いたマルチスポット再生システム、分離性能はよ かったのですが音圧と音質がイマイチなのですが、...」 NICT&北日本音響: 「それならこれで!!」







従来型 最大出力小&音質イマイチ

Chinese Korean



4言語マルチスポット再生

改良型 最大出力大&音質良好

発表会場にて実機デモあり!!

2-1-11

2-1-11 放射モードに基づく近傍再生用 スピーカアレイ:音響インテンシティの分析

Radiation mode-based loudspeaker arrays for confining sounds in near field: Analysis on acoustic intensities

〇貝塚勉(工学院・先進工)

- ◆周囲への音漏れの生じにくい、オープンイヤー型スピーカ (スピーカ アレイ) の実現を目指した研究。
- ◆放射モードの理論に基づいて各スピーカの振幅・極性を設定し、近距 離場を強調する。
- ◆提案のスピーカアレイを駆動したときの音響インテンシティを計算 機シミュレーションによって分析したところ、アクティブ成分に対す るリアクティブ成分の比率が大幅に高められることが示され、機序の 一端が明らかになった。

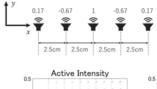


Fig.1: Loudspeaker array

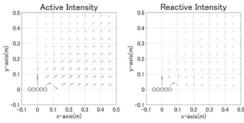


Fig.2: Acoustic intensities generated by loudspeaker array at 2000 Hz.

2-1-10

2-1-10 低域指向性・局所増音を実現する 3ch 音場制御技術の開発

Development of sound field control technology

〇江波戸明彦, 蛭間貴博, 西村修(東芝)

- ◆生活環境において情報伝達手段の一つとして音の活用が進む一方で、 音は騒音にもなり、音の混在や聴き疲れなどの課題も生み出し、パー ソナル空間だけに音が伝わる技術が望まれる。
- ◆本研究では特定方向に局所増音を実現するスピーカ再生制御技術を 開発した。音響パワ一最小化と増音化を両立する制御則の提案により、 スピーカ3個で急峻な増音勾配、低域指向性が実現する(Fig.1)。
- ◆スピーカ間隔で決まる適用可能な帯域音源を対象に増音勾配 12dB を 実証し、広帯域音声信号でも効果を確認した(Fig.2)。
- ◆本技術は東芝デジタルソリューション(株)の音響ソフトウェア Soundimension®音場制御に活用されている。

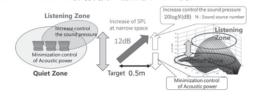


Fig.1 Basic principle



Fig.2 Experimental result 1/3 Octave Band at Point A~D

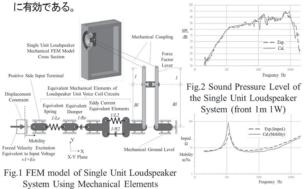
2-1-12

2-1-12 電気回路等価機械要素法の提案と それを用いたスピーカシステムの 電気機械音響強連成 FEM 解析

Electro-Mechanical Acoustic Coupled Analysis of Loudspeaker System Using Mechanical Elements Equivalent to Electric Circuits

○藤本圭祐, △宮本泰憲, △小井土哲平, △大嶋宏暢(パナソニック コネクト)

- ◆ボイスコイル等の電気回路を、等価な機械要素に置き換えて機械系 FEM モデルに組み込む「電気回路等価機械要素法」を提案した。
- ◆この手法を用いて単発及び複数ユニットのスピーカ音圧及びインピ 一ダンス計算を実施し、計算結果が無響室測定結果とよく一致するこ とを確認した。
- ◆本手法は、スピーカシステム音圧計算の計算手順簡素化と複数ユニッ ト相互影響及びネットワーク回路の影響を考慮した計算の精度向上



Equivalent to Electric Circuits

Fig.3 Impedance (Mobility) of

the Single Unit Loudspeaker System

2-1-13 カーネル法に基づく散乱体を含む領域内の 入射音場推定 —MM アルゴリズムによる ハイパーパラメータ最適化—

Kernel interpolation of incident sound field in region including scattering objects

—Optimization of hyperparameters by EM algorithm—

☆山野孝太(東大院), 小山翔一(NII), △中田雅貴, 猿渡洋(東大院)

- ◆従来の音場推定手法では、対象領域内部が自由音場であることを仮定 していたため、散乱体が存在する場合に著しく精度が低下する可能性 がある。
- ◆入射音場と散乱音場をそれぞれカーネル関数と球波動関数展開の線 形和によって分離して表現することで、領域内に未知の散乱体を含む 場合の入射音場推定手法を提案する。
- ◆推定式に含まれるハイパーパラメータは、EMアルゴリズムによって 音源や散乱体についての事前情報を用いずに決定する。
- ◆3次元自由空間における数値実験の結果,特に400Hz以下の低周波数 領域において,提案手法によって推定精度が向上することが示された。

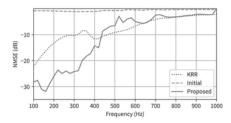


Fig.1: Normalized mean square error of incident field

2-1-15

2-1-15 Towards Optimized Crosstalk Cancellation in Bone Conduction: An In-Depth Exploration of FxLMS and Lateralization

- Our study presents a strategy to overcome "crosstalk," an issue where sound designated for one cochlea unexpectedly reaches the opposite one through bone conduction (BC).
- ◆By leveraging the FxLMS algorithm, we initially optimized crosstalk cancellation at an accelerometer on the left temporal bone, as depicted in **Fig. 1**. Realizing this original setup may not optimally cancel the crosstalk at the cochlea, we employed lateralization to refine the phase and level of the crosstalk compensation filter in the frequency domain, shifting the cancellation target from the accelerometer to the cochlea.
- Our results show varied success in crosstalk cancellation across subjects and test frequencies, demonstrating the potential of a combined FxLMS and lateralization approach to enhance performance while acknowledging individual differences.

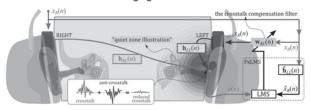


Fig. 1 Unilateral Crosstalk cancellation process using the FxLMS algorithm.

2-1-14

2-1-14 Optimization of a VAE-based personalized HRTF prediction algorithm and subjective hearing evaluation

☆CHANG Yuan, LI Xinpei, OIKAWA Yasuhiro (Waseda Univ.)

- Based on R. Miccini and S. Spagnol's research, we refined and optimized the application of a personalized HRTF prediction system, the algorithm utilizes the listener's pinna edge and anthropometric data to predict the personalized HRTF.
- We have created spatial, multi-angle test audio, involving participants in subjective hearing localization experiments to evaluate the optimization level of the algorithm, which proved that the original system and the improved system had an optimized effect on the accuracy of spatial audio localization.

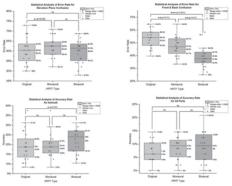


Fig.1: The statistic on optimization effects of the original system and the improved system.

2-1-16

2-1-16 遠隔音響データ収録システムの開発

Development of remote measurement system for acoustic data.

☆堀野由輔, 及川靖広(早大理工), 生谷正一(小松製作所)

- ◆工場における騒音の改善を図る研究において、大量の騒音データが必要であるが、現在主流である騒音計によるデータ収集は高価である、 スペースを必要とするという点が難点である。
- ◆Wi-Fi を内蔵する ESP32 というマイクロコンピュータと MEMS マイ クロホンを用いて、安価かつスペースを必要としない、長時間の録音 に適したシステムを提案する(Fig.1)。
- ◆提案システムでは、MEMS マイクロホンで測定した音響信号を ESP32 から UDP 通信にてホスト PC に送信。ホスト PC では Unity アプリケーションで送受信の制御を行う。
- ◆提案システムの指向性, 周波数特性を測定した。
- ◆工場におけるトランスミッションの検査騒音の測定実験により,実際 に長時間のデータ収録が可能であることが確認できた(Fig.2)。

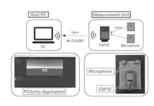


Fig.1: System overview

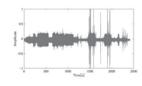


Fig.2: Noise of transmission test

2-1-17 複数観測信号レベル比に基づき音源方向を 8カテゴリーに分類する手法の精度について

Accuracy of classifying a radiation direction of sound source into eight categories based on multi-channel level ratio.

> ☆津田龍星(阪産大院), 福田光太郎, 粟谷朋留(阪産大), 中山雅人, 高橋徹(阪産大院)

目的 格子状配置したマイクロホン内部にある音源の放射方向を推定 ・格子内部にあるスピーカーの放射方向を推定

手段 スピーカーの空間放射特性とマイクロホンで観測される 1/3 オク ターブバンドレベル比から方向を推定

・相対音圧レベルの偏差の比を用いる手法 →レベル変動に頑健

評価 雑音条件 4 種類で放射方向推定の正解精度 (=8 方向中から正解方 向を推定する頻度) を比較

結論 中心周波数 2000 Hz の時、雑音条件による精度のばらつきが少な く方向推定が可能

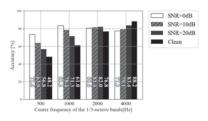


Fig: Accuracy of estimating speaker direction based on 1/3 octave band level

2-2-1

2-2-1 音響インピーダンスを指標とした 生体組織の三次元音響特性評価

Evaluation of three-dimensional acoustic properties of biological tissues using acoustic impedance as an index

☆橋本明楽(千葉大·院融合), 平田慎之介, 吉田憲司(千葉大·CFME), 伊藤一陽(東京農工大・工学研), 丸山紀史(順天堂大・医), 山口匡(千葉大·CFME)

- ◆中心周波数 250 MHz の ZnO 振動子を搭載した超音波顕微鏡を用い て, 自己回帰モデルを適用した信号分離により筋組織の三次元音響イ ンピーダンスを算出し、内部構造と音響的性質の関係性を評価した
- ◆超高周波超音波計測で生じる特定の周波数成分の低減などを考慮し て、元信号および使用帯域を制限した信号を入力信号として評価を行 い、結果を比較した.
- ◆信号分離前後で筋組織の音響インピーダンスが大きく特徴的に示さ れる深度が異なり、適切な周波数帯域で波形分離を行うことで生体組 織表面以外の内部の特性も評価可能と考えられる (Fig. 1).

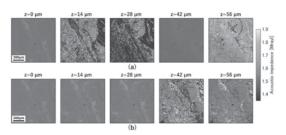


Fig. 1 Acoustic impedance maps for each depth before signal separation (a) and in the -20 dB frequency domain after signal separation (b).

2-1-18

2-1-18 複数の球体障害物で構成された 音響レンズの平面波の拡散効果の評価

Evaluation of the diffusion effect of plane waves in an acoustic lens composed of multiple spherical obstacles

☆増川英寿. 高橋義典(工学院大)

- ◆音響ホーンから出力される平面波拡散を目的として複数の球体で構 成された音響レンズを設計した(Fig. 1).
- ◆レンズを配置したときとしなかったときの FDTD シミュレーション を実施した(Fig. 2)
- ◆レンズを配置したときとしなかったときの音圧差を計算したところ、 所望する角度である 30 度以上でパワが上昇していた(Fig. 3).

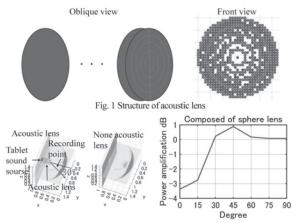


Fig. 2 FDTD Simulation of wave front with and without acoustic lens

Fig. 3 Power difference with and without acoustic lens

2-2-2

2-2-2 後方散乱係数解析における 特異値分解フィルタの設計条件の検討

Investigation of design conditions of SVD filters in BSC analysis ☆鈴木舜也, 大村眞朗, 長岡 亮, 長谷川英之(富山大·工)

- ◆ 血流の超音波高速イメージングにおいて、特異値分解 (SVD) によるクラッタ フィルタで血流エコーを可視化できるが、血球サイズ推定のための後方散乱係 数(BSC)解析においてフィルタ設計の適応的手法が必要である。
- ◆ 本報告では SVD フィルタの設計条件 (クラッタ対血流成分の上限閾値) が血流 像の画質性能および BSC 解析に与える影響を確認した.
- ◆ 使用データは、クラッタ信号を含む流路ファントムに生理食塩水で伸展したブ タ赤血球を定常循環させた際の超音波 RF データである.
- ◆BSC 解析における参照・解析データについて、SVD フィルタの上限閾値を変 化させた際のコントラスト (クラッタ対血流部分) が Fig. 1 である.
- ◆ 参照データの上限閾値を-41 dB で固定し、解析データの上限閾値を変化させ た際の BSC の傾きが Fig. 2 であり、コントラストが最大のとき、レイリー散乱 の傾向を示す周波数の4乗に近づいた。 適切な閾値条件のフィルタリングによ って赤血球単体の特性を示す理論値の傾向を捉えられたと考えられる

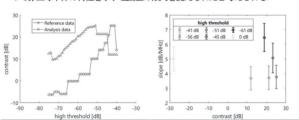


Fig. 1 Contrast of reference and analysis data

Fig. 2 Comparison between BSC slope and contrast at each high threshold of SVD filte

2-2-3

2-2-3 大型動物の脚骨皮質骨の超音波物性

Ultrasonic wave properties in the cortical bone of large animal's leg
☆児玉周大, 津幡大聖, 松川真美(同志社大)

- ◆大型動物への定量的超音波法 (QUS) の応用を目指し、基礎研究としてサラブレッド第三中足骨骨幹部皮質骨の超音波物性評価を行った。
- ◆骨幹部皮質骨を前(A), 外前(AL), 外(L), 外後(PL), 後(P), 内後(PM), 内(M), 内前(AM) の8つの立方体に加工し、それぞれ表面の微細構造を電子顕微鏡により観測した。また, <u>Axial, Radial, Tangential 方向</u>から超音波を透過させ<u>音速と減衰</u>を測定した。
- ◆減衰は密度と相関がみられた(Fig.1(a))。減衰が高い部分は血管と思われる太い管が多いため密度が低い。このような微細構造の違いが減衰に影響したと考えられる。
- ◆Radial, Tangential 方向では音速がほぼ一定であるのにもかかわらず、 減衰値が大きく変化した (Fig.1(b))。これは <u>haverse 管や空孔の異方</u> 性が影響したと考えられる。

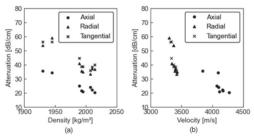


Fig.1:(a) Density—Attenuation (6MHz), and (b) Sound velocity—Attenuation(6MHz).

2-2-5

2-2-5 超音波を用いた肝臓内音速推定における 血管の走行角度の影響に関する基礎検討

Examination of effect of rotation angle of measured cross-section from short-axis plane of blood vessels on speed-of-sound estimation in liver

〇森 翔平, △小野田啓司, 荒川元孝, 金井 浩(東北大)

- ◆血管からの高SN 比信号を利用した肝臓内音速推定法を検討している。
- ◆血管短軸に対するプローブの回転角が音速推定に与える影響を検証 した。ファントム内のワイヤーA, B, C, D からの受信信号を用いて、 プローブとワイヤー間の平均音速をプローブ回転角ごとに推定した。
- ◆回転角が大きくなると正の音速推定誤差が生じた。さらに、中心から 離れた素子で受信信号の振幅が小さくなることが明らかになった。
- ◆中心素子に対する端の素子での受信信号のパワー減少量と音速推定 誤差の間に一定の傾向がみられた(Fig. 1)。
- ◆血管短軸に対する実際のプローブ回転角が未知の in vivo 計測においても、実測可能な受信信号の素子ごとのパワーに基づき、音速誤推定が生じる条件を判別できる見通しを得た。

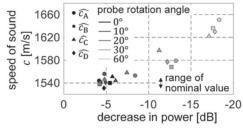


Fig.1: Relationship among probe rotation angle, estimated speed of sound, and decrease in power at edge elements.

2-2-4

2-2-4 高周波リニアアレイの送受信特性と振幅包 終特性の評価精度の関係性の検証

Verification of the relationship between the accuracy of evaluation of the transmit/receive and amplitude envelope statistics of high-frequency linear arrays

☆比賀大成(千葉大・院融合工),△Jeffrey Ketter I ing,△Jonathan Mamou

(Weill Cornell Medicine·Radiology), 平田慎之介,吉田憲司,山口匡(千葉大·CFME)

- ◆特性の異なる複数種の高周波リニアアレイプローブで二種の散乱体が混在する脂肪肝ファントムを観察し、Double-Nakagami (DN) Model を用いて振幅包絡特性を評価し、散乱体構造と評価パラメータとの関係性及び音場特性が評価精度に与える影響について検討した。
- ◆高周波・高分解能プローブで、正常肝ファントム(Fig. 1(i))に比して、 脂肪肝ファントム(Fig. 1(ii))では、脂肪滴を模した強散乱体からの 散乱強度に相当するパラメータが評価領域全体で高値を示し、弱散乱 体(肝臓)と強散乱体(脂肪)が弁別評価された。
- ◆相対的に低分解能のプローブに比して、高分解能プローブでは強散乱体を単体評価する精度が高く、弱散乱体と強散乱体の検出精度がpoint spread function (PSF)に依存することが確認された。

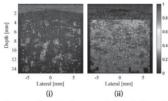


Fig. 1: Results of evaluation of echo signal intensity from fat component in normal liver phantom and fatty liver phantom by DN Model

2-2-6

2-2-6 夏季における日本海(Japan sea)の 音波伝搬

Sound propagation of NIHONKAI (Japan sea) in summer 〇鶴ヶ谷芳昭(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 200 m の海峡でつながっており、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、冷水塊である日本 海固有水を形成する. 夏季、対馬暖流の勢力が増大し、検討ラインの 南側では水中音速は増大し、サウンドチャンネル軸深度は深くなる。 また、海面付近の水温が上昇し、高い垂直水温傾度(VTG)を生ずる.
- ◆Fig. 1 に春季と夏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)は夏季であり、音源深度は10 mである。夏季は音源深度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が増大し、伝搬エネルギーは急速に減衰する。また、能日本サイドからの伝搬に関しても、高いVTGが存在することによって大陸棚内の伝搬損失が増大する。大陸棚を出た音波は大陸棚を下り、収束帯伝搬に移行するが、大和堆での伝搬損失は大である.

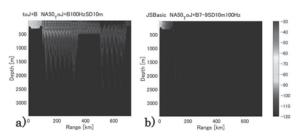


Fig. 1 Comparison of sound propagation from the continental side in spring and summer a) in spring b) in summer SD;10m

2-2-7

2-2-7 潮汐が存在する水中音響通信路における インパルス応答の計測

Measurement of Impulse Response in an Underwater Acoustic Channel with Tidal Effects.

☆齋藤大雅(筑波大・エシス), 海老原格, 若槻尚斗, △前田祐佳(筑波大・シス情系)

- ◆ 水中音響通信路の環境は、気象(水温・波高)や潮汐(水深)によ ってダイナミックに変化する、そのため、水中に適した無線 通信や測位システムを確立するためには、通信路の環境を理 解することが不可欠である。
- ◆ 大潮の日に、同一フレームに固定した送波器と受波器を、静 岡県鍋田湾の防波堤壁面から吊り下げ、水中音響通進路のイ ンパルス応答を測定した(Fig. 1).
- ◆ 16 時間にわたり実験を実施したところ、アルミフレームを伝 わる固体振動、水中音響通信路の直達波、海面反射波が計測 された(Fig. 2). 特に海面反射波の到達時刻は潮位の影響を 受けており、また、その広がりは、海面の粗さに影響を受け ていると考えられる.

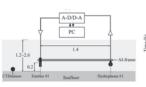


Fig. Experimental environment.

Fig. 2: Impulse response obtained from experiments.

2-2-9

2-2-9 MHz 強力超音波により生じる 液面水柱高さと音圧の関係

Relationship between height of water column and sound pressure under illumination of high intensity MHz ultrasound

☆安田直生, 和田有司, 中村健太郎(東工大)

- ◆数 MHz かつ MPa オーダーの水中超音波の音圧を、水面にできる水 柱(Fig. 1)の高さから求める方法について検討した。
- ◆1.6 MHz 水中超音波により隆起した水柱にかかる重力と水柱端部で の音響放射力の釣り合いから音圧を求めた。
- ◆この測定方法によって求めた音圧を、反射型光ファイバ音圧センサで 測定した音圧と比較した。
- ◆ どちらの手法を用いた場合でも、測定範囲内 (0.8~1.0 MPa) でほぼ 一致した音圧値が得られた(Fig. 2)。
- ◆以上より、1MPa 程度の音圧を水柱高さから求められることがわかっ t-





1.6 MHz ultrasonic waves.

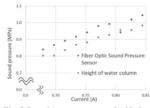


Fig. 2 Sound pressure measured with the two methods vs. current to the transducer.

2-2-8

2-2-8 中部海域の震源構造とその地域的変化

Hypocenter structures in the Central sea area and their regional changes

〇菊池年晃(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。 これらの震源域で発生する地震の活動を エネルギー推移で調べた結果、それらの年 間累積エネルギーはその海域の地震活動 を表わしていることが分かった。この方法



を応用して近隣の震源域、即ち東海地方南方はるか沖(上図)及び駿河湾 南方沖(下図)の年間累積エネルギーを求めた。東海地方南方はるか沖の 年間累積エネルギーは、赤線で示した浅部震源域エネルギーが青線で示

の差は次期地震のエネルギーとして蓄積さ れている。そしてその蓄積エネルギーは年 と共に増大しているので監視が必要である。 一方、駿河湾南方沖の年間累積エネルギー

した深部震源域エネルギーより大きい。そ



は、2009年に発生したM6.5の本震とその余震のエネルギーを示してい る。それは近年、減少中で、2020年には赤線で示す深部震源域のエネル ギーとほぼ拮抗している。

2-2-10

2-2-10 エルミート補間型微分マトリクスを導入した 放物型方程式法による海洋内の3次元音 波伝搬解析

Analysis of 3-D sound wave propagation in the ocean by parabolic equation method introducing Hermite-interpolated differential matrix.

☆川崎 圭一朗, 鮫島 俊哉(九大·芸工)

- ◆本研究では波動理論に基づき、波動方程式を前方一方向の進行波を表 す微分方程式に近似し遠方界解析を行う放物型方程式法を用いた。
- ◆放物型方程式法で得た微分方程式に対してスペクトル法を適用した。 スペクトル法は少ない自由度で高精度の解が得られる手法である。
- ◆スペクトル法の中でも、エルミート補間を用いて微分マトリクスを構 成する手法を用いた。エルミート補間型微分マトリクスを用いること で、領域端の境界条件を簡単に計算に組み込むことができる。この手 法によって海面や海底の境界面の条件を柔軟に取り入れた計算を低 負荷で行うことが可能となると期待される。
- ◆解析の流れを Fig. 1 に示す

Governing Equation $\frac{\partial^2 p}{\partial r^2} + \frac{1}{r} \frac{\partial p}{\partial r} + \rho \frac{\partial}{\partial z} \left(\frac{1}{\rho} \frac{\partial p}{\partial z} \right) + k_0^2 n^2 p = 0$ Slowly Envelope Approximation Factorize and Neglect the $p(r,z) = \psi(r,z)$ effect of receding waves Parabolic Equation $\frac{y}{x} = jk_0(-1 + \sqrt{1 + X})\psi$ $\frac{1}{k_0^2} \rho \frac{\partial}{\partial z} \left(\frac{1}{\rho} \frac{\partial p}{\partial z} \right)$ Padé Approximation t Equation to be solved Analyze by spectral method $\frac{\partial \psi}{\partial r} = jk_0 \frac{X/2}{1 + X/4} \psi$

Fig. 1 Analysis Flow

2-2-11

2-2-11 日本海溝沿い海底で観測された小笠原・ 北マリアナからの海中伝搬波の特徴

Characteristics of underwater acoustic waves from the Ogasawara and Northern Mariana Islands observed on the seafloor along the Japan Trench 〇岩瀬良一(JAMSTEC)

- ◆日本海溝沿いの150 観測点からなる海底ケーブル型観測網S-net の海底地震計により観測されている、観測網の南方から深海サウンドチャネル (Sound Fixing and Ranging, SOFAR チャネル) を伝搬してきたと考えられる海中伝搬波の解析を行った。
- ◆当該伝搬波のうち、2022 年8月以降断続的に繰り返し観測されている顕著な信号については、筆者の先行研究門において、小笠原諸島南端付近の海溝斜面上に音源があると推定している。一方、Wake 島沖のハイドロフォン・アレー観測から、当該信号が北マリアナの Ahyi海底火山の活動に伴う信号であるとの指摘がある四.
- ◆そこで今回 S-net の観測波形を解析した結果, 音源が Ahyi 海底火山 にあるとして矛盾がないことを確認した. 文献1]の推定位置は誤差の 範囲と考えられる.
- ◆また、小笠原諸島の海徳海山の火山活動に伴う微弱な海中伝搬波を、 2022 年9月1日 12:20 頃を中心に検出した。
- ◆海徳海山の火山活動は海面の変色水などが中心で、2021 年 8 月の福徳岡ノ場のような顕著な噴火活動は観察されていないが、同程度の火山活動でも海底地震計観測網で検出可能なことを示している。

参考文献

- [1] 岩瀬, 海洋音響学会 2023 年度講演論文集, 81-82, 2023.
- [2] Tanaka et al., JpGU2023, SVC34-P01, 2023.

2-3-2

2-3-2

周りに配慮をしながらイマーシブメディアに 没入するための音響デバイスとその技術

Acoustic equipment and technology to playback immersive media with minimal sound leakage to surroundings.

〇鎌土 記良

(日本電信電話株式会社 NTT コンピュータ&データサイエンス研究所)

- ◆5G 通信・ネットワークの普及により、イマーシブなコンテンツを場 所を選ばず楽しめる世界が実現しつつある
- ◆イヤホンやヘッドホンなどの音漏れしにくいデバイスは耳への高い 圧迫感を伴い、長時間装着時の耳への身体的負担により没入感を損な う大きな原因となる
- ◆ また、 周りの音が聞こえないことから移動時の利用においては事故や 周囲への迷惑行為の誘発などが社会課題ともなっている
- ◆オープンイヤー型のイヤホン・ラウドスピーカなどでは、近隣への配 慮として「音漏れ」を解決する必要がある
- ◆本講演では、これらの課題をターゲットとした音響デバイスの最新技術動向を紹介する



2 - 3 - 1

2-3-1 オブジェクトベース音響を用いた 放送サービスにおける 制約条件のある視聴環境への最適化

Optimization for constrained listening environments in broadcasting services using object-based audio

◎久保弘樹(NHK)

- ◆現在放送導入が検討されているオブジェクトベース音響は視聴者の好み や聴取環境に合わせて番組音声のカスタマイズが可能である。放送で想定 される様々な制約(以下、【隅付き括弧】)と、その制約の下でオーディオ 再生の最適化に資するオブジェクトベース音響の機能(以下、下線)、お よび関連する NHK の研究開発について紹介する。
- ◆ 【スピーカ配置に関する制約】家庭で制作環境と同じフォーマット(スピーカ数・配置)で再生できない場合に、音声オブジェクトの3次元パンニングとチャンネルベース音響のダウンミックスを組みあわせて、番組全体で高い臨場感を保持したフォーマット変換ができる。
- ◆ 【ダイアログの聞き取りに関する制約】コンテンツの臨場感を損なわずに ダイアログを聞き取りやすく調整したい場合に、聴取環境ごとにダイアロ グと背景音のレベルパランスを変更したり、ダイアログの再生位置を変更 もしくはセカンドデバイスに出力したりできる。
- ◆【伝送路に関する制約】複数の番組構成の中で音声オブジェクトの組み合わせを変えて、データ量に制限のある伝送路においても多様な副音声を可能とする。伝送路間あるいは番組構成間でラウドネス値のノーマライズができる。

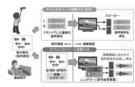


Fig.1: Overview of object-based audio

2-3-3

2-3-3 Optimal Source Distribution 原理に基づく バイノーラル制御の優位性

Binaural synthesis over loudspeakers based on the optimal source distribution principle

〇矢入幹記, 星野嗣人(鹿島技研), △武内隆(OPSODIS Limited, ISVR)

スピーカを用いたバイノーラル合成により3次元音場再生を実現する際の留意点を整理するとともに、バイノーラル合成の一手法として、 Optimal Source Distribution の原理を適用することの利点を考察した。

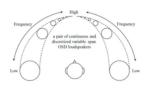


Fig.1:Conceptual monopole sources to satisfy the principle of the optimal source distribution

2 - 3 - 4

2-3-4 単純な利得と遅延操作による 2 チャンネルの Optimal Source Distribution 原理に基づくバイノーラル合成

Binaural synthesis adjusting a simple gain and delay based on the two-channel optimal source distribution

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

- ◆スピーカによるバイノーラル合成手法の一つである Optimal Source Distribution (OPSODIS)原理は、方位角が周波数の関数である音源対という概念を利用する。
- ◆この原理をより単純化し、離散化 OPSODIS 原理のスピーカを用い、 利得と遅延操作といった簡易な信号操作のみでクロストークキャン セルを実現する手法を提案する。
- ◆提案モデルの特徴は、再生伝達関数を平坦化せずクロストークキャンセルのみを実現することで、逆システムが消費するダイナミックレンジを抑制できる点である。

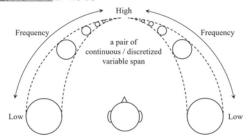


Fig.1:Conceptual continuous and discretized sound sources to satisfy the two-channel OPSODIS principle

2-4-2

2-4-2 楽音の周波数特性を考慮した 仮想残響エンベロープ指数と楽曲聴取時に 感じるリラックス感との関係

Possibility of relaxation level measurement during music listening using reverberation envelope that considering the frequency characteristics

☆笠島菜那, 脇田由実(大阪工大), 入交英雄((株)WOWOW)

◆これまで我々は、残響の聴こえに関する楽曲の音響信号特徴量の指標として従来から提案されているエンベローブ指数 (E値) を用いて、楽曲の E値とリラックス感との関係について調べてきた。そして、 E値は、楽音を聴取した際に感じるリラックス感の指標の一つになる可能性が示唆されたことを報告した。今回はさらに、周波数軸上でのパワー分布が広い場合の残響の減衰量を考慮するために、周波数軸方向標準偏差を S値と定義し、 E値と S値との積値を指標としてリラックス感との関係を調べた。その結果。 E値よりも ES値の方が、より両者の関係が強くなる結果となった。

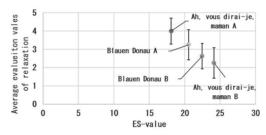


Fig.1:Relationship between the ES-value of a musical piece and the degree of relaxation.

2 - 4 - 1

2-4-1 同期タッピング課題において 音楽経験による成績の差を生む要因 一時間知覚と非同期誤差の訂正に着目して―

Factors Contributing to Differences in Performance by Musical Experience in a Synchronized Tapping Task: Focusing on Time Perception and Correction of Asynchronous Errors

☆横谷有紀(法政大院・人文科学研)

【目的】 ① 空白の刺激音間隔の時間知覚能力と ② 同期タッピングの誤差の訂正能力を、楽器演奏経験の有無で比較する

【方法】楽器演奏経験の有無(3年以上/なし;参加者間)

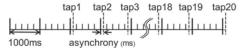


Fig.1 Stimulus sound sequence used in the subdivision condition.

The horizontal axis represents time, the long vertical solid line represents the stimulus sound to be synchronized to, the short solid line represents the subdivision stimulus sound, and the long dotted line represents the timing of the participant's tap.

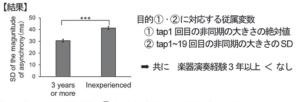


Fig.2 Result of ANOVA for @ the SD of the magnitude of asynchrony. Error bars indicate standard errors. ***p<.001

【考察】楽器演奏経験者は非経験者よりも優れた非同期の訂正能力を持つ

2-4-3

2-4-3 音楽聴取時に生起する鳥肌感の測定と評価 --部分的な音量操作による影響---

Measurement and Evaluation of Pleasurable Musical Chills
—Effect of Partial Sound Volume Control—

〇横山彩乃, 岩城護(新潟大院)

- ◆音楽聴取時に生起する鳥肌感 (Pleasurable Musical Chills) は、人間 の行動を促進する効果が期待できる。鳥肌感は、楽曲中の音が大きな 部分で生起しやすいことが報告されている。
- ◆本研究の目的は、選択的かつ段階的に音量を増幅することによって、 鳥肌感の促進効果に対する音量の影響を調査することである。
- ◆無加工の聴取楽曲自身とその中で鳥肌感が報告された部分の音量を +3dBと+6dB増幅したものとを聴取し、鳥肌感生起時にボタンを押 し続けてもらった。皮膚電気反応 (GSR) と心拍数 (HR) を計測し、 楽曲聴取後に感想アンケートに回答してもらった。
- ◆無加工の場合に比べて+3 dB 増幅した場合の方が鳥肌感の促進が見られた。一方、この傾向は個人個人の鳥肌感生起の要因に影響されている可能性も示された。

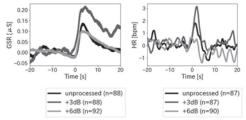


Fig.1: Synchronous averaging of GSR and heart rate (Synchronization point 0 s was the time when the button was pressed.)

2-4-4 フルートの演奏支援システムのための 音響分析: 演奏音の自動分類の検討

Acoustic Analysis for Flute Practice System: Towards Automatic Sound Classification

☆大下沙偉(日大院・総基科研), 大澤智恵(武庫川女大・音楽), 北原鉄朗(日大・文理)

- ◆目的: フルートの適切/不適切な吹き方の分類のための音響特徴分析
- ◆使用データ:3名のフルート専攻の音楽学部生が次の指示に従って演奏 楽譜:5種類,演奏法:[普通][口大きめ][口小さめ] など計7種類
- ◆ 主観評価: 上記の演奏者に評価を依頼
- ◆ 抽出する音響特徴:振幅基本周波数の時間変化の平均, はじめ/中間 1 秒の倍音数, 音高の正確さ (v) など計26 個
- ◆ vrの分布を Fig. 1 に示す。 左から[普通] [口を大きめ] [口を小さめ] [楽器を手前に傾けて] [楽器を奥に傾けて] [タンギングを強めに] [タンギングを弱めに] の順に配置。 演奏者ごとに色分けを行った。

楽器を傾けた際に値が変化しやすいことがわかる.

- ◆ 音響特徴量の分布だけでなく、主観評価の分布や主観評価と音響特徴量の関係をグラフ化し、分析を行った。
- ◆一部予測と異なる結果が得られたため、データをさらに収集し分析を継続する。

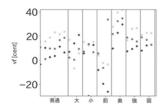


Fig.1: Stability of pitch (Horizontal: Sound ID, Vertical: Feature Value(Vi))

2 - 4 - 6

2-4-6 シングルリード楽器における 管の振動特性の解析

Analysis of tube vibration characteristics in single reed instruments

☆塩野菜々子, 高橋義典(工学院大)

- ◆リード楽器は、直接リードの振動が管に伝わることから、管の振動に 起因する音響出力がエアリード楽器に比較して大きくなる。
- ◆熱溶解積層方式の3Dプリンタを用いて作成した音響管と、既製楽器のシャリュモーを使用し、管の振動加速度を計測した波形を円筒関数の合成として、空間フーリエ変換することでシングルリード楽器の管の振動特性の分析を試みた。
- ◆振動周波数において、いずれも4次モードまでの結果では音色に影響 すると考えられる音響出力は確認されなかった。



Fig.1:Acoustic tubes created by 3D printer

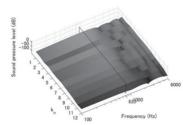


Fig.2: Estimation of sound power of the source due to vibration of acoustic tubes created by 3D printer (t = 5 mm)

2-4-5

2-4-5 口笛の重音奏法の声道模型を用いた検討

A Consideration of multiphonic whistling by using the vocal tract model

☆川原田美月, 森 幹男(福井大院)

- ◆2音が同時発生(重音が発生)する5通りのプレートの組合せの口笛 声道模型(Type I~V)を用いて、2音の基本周波数と口唇部における 流速との関係を実験的に調べた。
- ◆流速の重複する区間において Pucker whistle (口唇をすぼめて吹鳴する一般的な口笛) 同士の重音が発生することが確認できた。
- ◆Pucker whistle と Hollow whistle (Pucker whistle より基本周波数が低く口唇の開きが大きい口笛)の重音も発生することが確認できたことから、新たな重音奏法の可能性が示唆される結果となった。

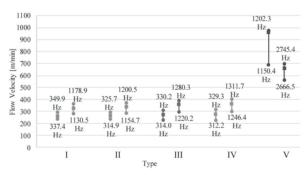


Fig.1 Relationship between the fundamental frequency of generated sound and flow velocity

2-4-7

2-4-7 ピックのダイナミクスを考慮した撥弦物理 モデルと箏爪への応用

Physical model of plucked strings involving dynamics of a pick and its application to *koto* plectrum

☆塩屋仁菜(九大·芸工), 鮫島俊哉(九大·芸工)

- ◆撥弦振動について、擦弦楽器で見られるような stick-slip 運動、および 打弦楽器で見られるようなバチと弦の衝突の両者を導入した撥弦物 理モデルを提案する。
- ◆弦は、両端固定の理想弦として物理モデル化し、固有関数展開法による理論解析解で表現する。ピックは、シェル振動場として物理モデル化し、シェル要素を用いた有限要素法によって解析する。
- ◆弦とピックの物理モデルは、その間の衝突特性および摩擦特性を媒介 として連成する。
- ◆構築した撥弦物理モデルを箏に応用する。計算例として箏の弦と箏爪をモデル化し、ピックの形状による計算結果の比較および計算結果と現実の動作との比較を行う。

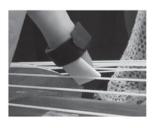


Fig.1: Koto plectrum.

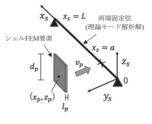


Fig.2: Geometrical arrangement of the pick and string.

講演取消

2-4-10

2-4-10 ドラムを用いた三線胴のモデルによる 心が膜の振動に与える影響の評価

Evaluation of the effect of *shin* on membrane vibration by a model of a sanshin drum body using a drum.

〇小島知也, 西宮康治朗, △長秀雄(青山学院大)

- ◆三線(サンシン)は沖縄の伝統ある民族楽器の1つである.
- ◆心の存在および形状が三線胴の両面膜の連成振動に与える影響を評価することを目的として、打撃試験により両面膜の振動を計測した。その結果、228 Hz のモードでは心の影響は見られなく、768 Hz のモードにおいて心の挿入で両面膜が同位相から逆位相に変化し、減衰も大きくなることが分かった。
- ◆Fig. 1. に示す 768 Hz のモードでは、心なしの状態で同位相だったモードが、心を挿入することによって逆位相になったことから、胴内部の空気が圧縮膨張の単極子のモードとして振動し、その結果、減衰が大きくなった可能性がある。今後は有限要素法解析によって胴内部の音圧を評価し、音圧と減衰の関連性を評価する。

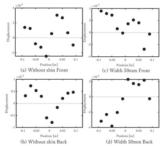


Fig.1: Mode shape of the membranes at 768 Hz

2-4-9

2-4-9 ヴァイオリンの有限要素モデル化と 3D プリンタを用いたヴァイオリンの比較研究

Comparative Study of Violin Finite Element Modeling and 3D Printer Violin

〇黒沢良夫(帝京大・理工)

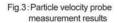
◆ヴァイオリンの表板・裏板等1つ1つの部品の3D データを取得し FE モデルを作成した。また、同形状で3D プリンターを用いてヴァ イオリンを作成した。実験と計算による振動解析結果の比較や、粒子 速度プローブによる実験結果を紹介する。

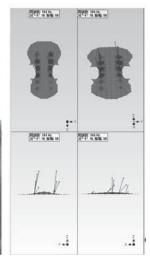


Fig.1:FE model for violin



Fig.2: Violins made of wood and clear resin





2-4-11

2-4-11

ピエゾピックアップを用いた スネアドラム打面の振動の収録

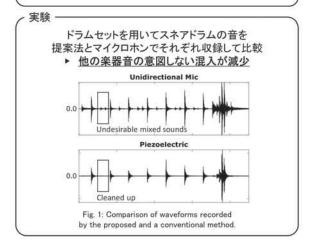
Vibration recording of snare drum head using piezo pickup ☆安藤元暉, 山田宏樹, 矢田部浩平, 倉科佑太(農工大)

従来法

ドラムセットは各楽器の音をマイクロホンで収録 ▶ 他の楽器音が混入してしまう

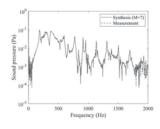
提案法

ピエゾ素子を用いて打面の振動から収録
▶ 意図しない混入音を減らす



2-4-12 モード重ね合わせ法による 太鼓の放射音合成

- ◆ 自由度の高い音の調整に対応できる音響合成手法の構築を目指し、膜 からの放射音に関して太鼓のモード解析に基づき、モードの重ね合わ せの原理を応用した音響合成法を提案する。
- ◆ 音の調整の一つとして太鼓に対して演奏者によって行われる、膜の特定の位置に質量付加などを行い、境界条件を変更することで特定のモードを励起されづらくするミュートと呼ばれる操作を模擬する。
- ◆ 得られた音響合成結果は実際に測定した音とよく一致しており、高精度な音響合成手法が構築できたと考えられる。また、選択したモードのみを使用して音を合成することが可能であり、特定のモードを励起されづらくするミュートと呼ばれる操作を模擬できた。



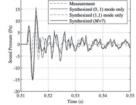


Fig.1: Result of Sound synthesis

Fig.2:Synthesized and measured sound pressure time histories

2-4-14

2-4-14 不均一性を有する軸対称形状体鳴楽器の 数値解析へのフーリエ級数展開の適用

Application of Fourier series expansion to numerical analysis of axisymmetrically-shaped idiophones with inhomogeneity.

☆小口雄也, 鮫島俊哉(九大・芸工)

- ◆軸対称形状体鳴楽器の数値解析に用いられる, フーリエ 級数展開を利用した半解析的なアプローチを, 物性値や 形状が不均一である場合に拡張する手法を提案する。
- ◆応用例として薄肉円筒シェルでモデル化された梵鐘を 取り上げ、その厚さが周期的に変化する場合に、特定の 部分音で唸りが生じることを示す。



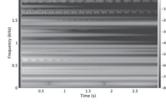
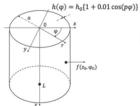
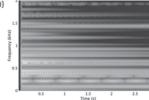


Fig. 1 Example of temple bell





 $Fig.\,2\,Coordinates\,of\,thin\,cylindrical\,shell.$

Fig. 3 Spectrogram of temple bell, calculated by the proposed method. (Top: p = 4, bottom: p = 6.)

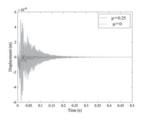
2-4-13

2-4-13 スネアドラムのコイルばね振動における 膜との摩擦の影響について

The effect of friction with membrane on helical spring vibration of snare drums

☆永田茉莉咲(東大院), △齋藤彰(明治大), 坂本慎一(東大生研)

- ◆スネアドラムは打撃膜と呼ばれる上面の膜と共鳴膜と呼ばれる下面 の膜を持つ両面太鼓であり、共鳴膜に接するように帯状に並ぶ弦が張 られた発音機構が複雑な楽器である。
- ◆弦がコイルばね状のものであることから、弦の振動を長手方向の伸びを含めた2次元で解析し、膜との接触の影響に着目し実験と数値シミュレーションを行った。
- ◆膜と弦の接触音を測定し、コイルばねを使用した場合の方が音圧が大きく、周波数成分が細かく数多く存在することを確認した.
- ◆長手方向の伸びと共に摩擦を考慮することで、摩擦力が周期外力のように長手方向の振動を励起し、変位が大きくなることを確認した。また、長手方向の振動では周波数成分が細かく数多く存在し、振動が複雑化することを確認した。



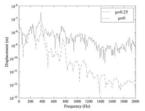


Fig.1: Time histories of the longitudinal displacement of the

Fig.2:Result of the FFT of the longitudinal displacement of the string

2-4-15

2-4-15 ドラムの張力分布推定を目的とした 張力判定アルゴリズムと均一性評価

Tension determination algorithm and uniformity evaluation for estimating drum tension distribution.

◎橋本涼汰(早大理工), 矢田部浩平(農工大), 及川靖広(早大理工)

- ◆(1,1)モードの湾曲から小張力領域を表示するアルゴリズムの提案
 - 1. モードを円環領域に切り抜き、極小値を2か所検出
 - 2. 中央から極小値へ向かう2つのベクトルを加算し正規化
 - 3. 半径の 0.6 倍程度の長さに調整し、Fig. 1 の赤円として表示
- ◆1 チャンネル録音信号から張力均一性を評価
 - Fig. 2: 白のボルトを締め、黒のボルトの強さに近づける
 - ▶ 白のボルトを締めると、評価値が増加
 - ▶ 均一な条件でも増加 → 張力均一性と相関無し
 - 6 つのボルトの平均的な強度と対応する

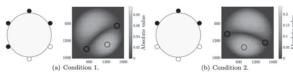


Fig.1 The low-tension areas marked by the tension determination algorithm.

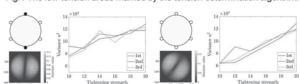


Fig.2 Evaluation index u² for tightening strength of the white points.

2-4-16 ピアノ連弾時におけるヒト演奏者のテンポ・リズ ム制御:自動伴奏システムへの組み込み

☆平賀佑哉, 饗庭絵里子(電通大)

- ◆ヒトどうしのピアノ連弾のような演奏者の演奏特徴を反映する自動 伴奏システムはないため、ピアノ連弾時の Secondo のピアニストの 演奏テンポについて解析を行い自動伴奏システムへの組み込みを目
- ◆様々なテンポ変化をする ritardando 指示部付近の提示録音を用いて. ピアノ連弾を模した演奏実験をピアニストに対して行った
- ◆図1より、n=25 までの様々な提示録音のテンポ変化に関わらず、ピ アニストの演奏 IOI の変化は類似した
- ◆ このことから Secondo の ritardando 部の演奏 IOI の変化は一定であ る可能性が示唆されたため、さまざまなフィッティングを検討した

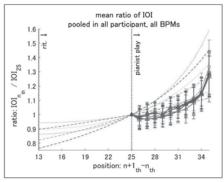


Fig.1: Mean IOI ratio in HANON (added rit.) (-: performance, --: recording)

2-4-18

2-4-18 物理モデルシミュレーションを用いた ピアノのタッチと音質の検討

Study of piano touch and tone quality using physical model simulation ☆佐藤菜緒, 鮫島俊哉(九大·芸工)

- ◆グランドピアノのハンマーシャンク, ハンマーヘッド, ピアノ3本弦, ブリッジを含めた物理モデルを構築し、FEM、FDM による数値解析 を行う。
- ◆構築した物理モデルを、打弦直前のハンマーヘッドの速度を揃える条

件のもと、pressed touch と struck touchの2種類のタッチで加振する。

◆放射音の数値計算結果から、二段 減衰の緩やかな減衰(余韻)に比べ 急速な減衰(打音)にタッチによる 違いが生じることが示された。

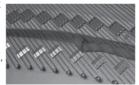


Fig.1: Multiple piano strings at the bridge.

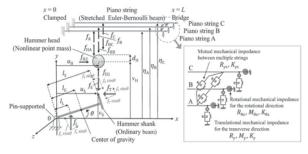


Fig.2: Physical model of multiple piano strings, hammer head, and hammer shank; coupling configuration of them.

2-4-17

2-4-17 残響除去 DNN に基づく特徴量を用いた ピアノ演奏におけるサステインペダルの タイミング推定

Piano sustain-pedal detection using feature value based on dereverberation DNN ☆浦田大輔(早大理工), 赤石夏輝(農工大), 及川靖広(早大理工)

◆目的

> ピアノ演奏でのサステインペダルのタイミング推定の精度向上

◆提案手法

- ▶ 残響除去 DNN 用いて新たな入力特徴量を生成し、ペダル推定 DNN と統合する機構
- ▶ サステインペダル操作時, 音が持続するという特徴を反映

◆結果

- ▶ 残響除去した入力特徴量を付加することによる精度向上
- ▶ ネットワークの学習構造の検討が必要

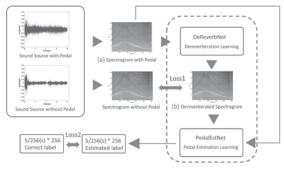


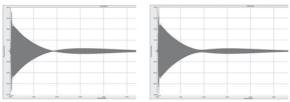
Fig.1: Overview of the proposed method

2-4-19

2-4-19 ハンマースティフネスのピアノ音への影 響に関する検討(第3報) -実験及びその分析-

Examination of the effect of hammer stiffness on the piano sound III -Experiments and analyses-

- ☆落谷颯 (神工大), △八代哲幣 (神工大), △倉田尚彦 ((株) クオルピア), 西口 磯春 (神工大)
- ◆報告者の1人は、「叩いた弦と響板のエネルギー伝達の効率が最適に なるような硬さにハンマーを調整すること」 として整音を捉えてい
- ◆前報では、有限要素法 (FEM) およびモード重ね合わせ法による弦モ デルの数値解析の結果について報告した.
- ◆本報告では、3 種類のニードリングと低次の部分音の音量の時間経過 の関係を実験的に明らかにすることにより、倉田の提案する整音のク ライテリアの検証を行った.



(middle)

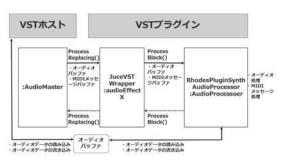
Fig.1: Time variation of 2nd overtone Fig.2: Time variation of 2nd overtone (hard)

2-4-20 振動体と電磁ピックアップを有する電気ピア ノ音源に関する研究 V ―実用的な実装の検討―

On Physical modeling of the E-piano with vibrating bodies and electromagnetic pickups V -Practical implementations-

☆川内詩土, 西口磯春(神工大)

- ◆振動体と電磁ピックアップを有する電気ピアノ(商品名: ローズピア ノ)を対象とする物理モデルの検討を行っている.
- ◆これまでに、作成した物理モデルをアプリケーションとして実装した.
- ◆本研究では、このアプリケーションを VST プラグインとして JUCE (Jules' Utility Class Extensions) と呼ばれるフレームワークにより実装した。
- ◆また、報告者の研究室で開発している新型クラヴィコードの物理モデルの実装に向けてアフタータッチによるピッチベンドの実装を行った。



Flowchart of VST Plugin

2-5-2

2-5-2 (招待講演) 視覚的要因が音の大きさに及 ぼす影響—色・映像を中心とした考察—

Influence of visual factors on perceived loudness: Focusing on color and video

〇金基弘(駿河台大)

- ◆普段、我々は身の回りに存在する音を聴覚のみでなく、視覚(物や風景)とともに音を知覚している場合が多い。そのため、音のデザインを考える際に、視覚と聴覚の相互作用に伴う取り組みも必要である。
- ◆本講演では、マルチモーダルな音のデザインに焦点を当て、音の大きさやうるささの知覚における色や映像の影響を中心に解説を行う。
- ◆音の大きさや音量は、色や映像といった非音響的要因からも影響 を受ける。さらに、色や映像の影響によるラウドネス知覚や音量 設定に明白な性差が存在しており、女性の方が影響を受けやすい。
- ◆以上のことは、誰もが利用しやすい音のデザインの基礎知見として、製品音などのデザインやマーケティングに役立つ。

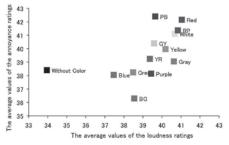


Fig. 1: The average values of the geometric means for the perceived loudness and annoyance in each color condition. "Without Color (i.e., pink noise only)" indicates the control condition.

2 - 5 - 1

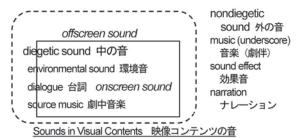
2-5-1 映像コンテンツにおけるサウンドデザイン ―音がストーリーを語り、リアリティを生み出 し、感動を増幅する―

Sound Design in Visual Contents: Sound Tells Story, Creates Reality, and Gives Stronger Impression

〇岩宮眞一郎(日大)

◆音は、時には映像に寄り添い、時には映像とずれながら、時には映像 と対峙しつつ、私たちの心をゆさぶる

→映像コンテンツは(((音)))があってこそ作品として成立する



- ◆音楽:登場人物の心情を表現、場面の状況の表現、ストーリーを作る、 劇中音楽の利用も効果的(音楽と映像の対位法)
- ◆効果音: ムードの醸成、場面の強調、ショック、リアリティの生成、 句読点的効果、注意喚起、笑いやかわいさの創出
- ◆環境音:空間の情報を表現(画面に映っていない音も含め),効果音 と同様の機能を持つ環境音も(演出的環境音)

<参考文献> 岩宮眞一郎, 図解入門 よくわかる最新映像サウンドデザインの基本, 秀和システム, 2023.

2 - 5 - 3

2-5-3 (招待講演)仮想把持における 触感再現のための視聴覚情報提示

Audiovisual information presentation for grasping feeling of a virtual soft object with a bare hand

〇佐藤美恵(宇都宮大・工)

本研究では、素手による仮想把持に着目し、実際には何も掴んでいない状況下でも、視聴覚情報の変化から仮想物体の把持感や柔らかさ 感をユーザに与えられるかを検討した。

その結果、視覚情報のみでも、仮想物体の柔らかさに合わせてユーザの手形状を補正することで、ユーザは仮想物体の柔らかさの違いを感じられることが確認できた。また、特に仮想把持においては、仮想物体を掴んだ際や放した際に聴覚情報を提示することで仮想物体の把持感が向上し、仮想物体の変形限界を知らせる聴覚情報を提示することで仮想物体の柔らかさ感が向上することが示された。

以上より、より効果的に仮想把持の触感を再現するためには、視覚 情報のみではなく、ユーザの動作のきっかけとなるよう聴覚情報を提 示することが有用である。

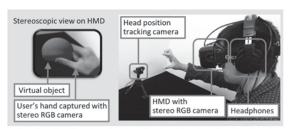


Fig.1: Our proposed AR system

2 - 5 - 4

2-5-4 スイッチの感触および操作音から得られる 操作感覚とオノマトペの関係性

Relationship between onomatopoeia and operation feeling based on tactile sensation and operation sound of switches

○宮入徹(都産技研)

- ◆スイッチの操作感覚は、クルクルやカチカチなどのオノマトペ(擬音 語、擬態語)によって表現されることが多い。
- ◆本研究では、オノマトペを活用した定量的な製品設計手法の提案を目的として、ロータリースイッチの操作感覚を構成する物理量と、その操作感覚を表すオノマトペの対応について検討した。
- ◆感触と操作音を構成する物理量と自由回答から得られたオノマトペ の同時布置による対応分析の結果から、操作感覚を表すオノマトペに おいて、物理量の大きさに応じた音象徴性を示した(Fig.1)。

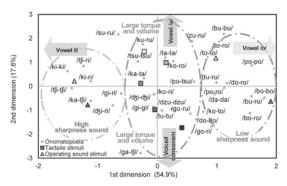


Fig. 1 Correspondence analysis results

2 - 5 - 6

2-5-6 視覚情報と聴覚情報の角度差が運転時の 情報探索に与える影響

Effect of angular difference between visual and auditory information on information searching during driving

☆鮎川 栞 (九大・芸工)、中 貴一 (東大・生研)、山内勝也 (九大・芸工)

- ◆本研究では、自動車運転中の視覚情報探索において視覚情報と聴覚情報の配置について検討を行った。
- ◆本研究ではドライビングシミュレータを用いた運転操作や視覚刺激のドアミラー位置への配置など、運転環境を再現して実験を行った。
- ◆結果として、視覚情報と聴覚情報が30°~40°離れていても視覚情報 探索への精度や反応時間に影響を及ぼさないことがわかった。一方で、 ドライバーの精神的作業負担については聴覚情報の位置が原因で妨 害感をもたらす可能性があることがわかった。
- ◆参考に、実験1のシステム図をFig.1に掲載する。

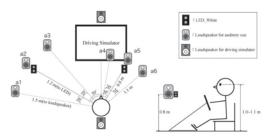


Fig.1: Experimental set-up (Exp. 1)

2 - 5 - 5

2-5-5 聴覚情報による自動車運転中の 視覚情報探索支援

Assistance of visual information search using auditory information while driving

◎中貴一(東大·生研), 山内勝也(九大·芸工)

- ◆聴覚刺激により視覚の空間的注意を誘導することで、視覚刺激に対する反応を促進することができる。これを Crossmodal spatial cuing effect といい、実験心理学、人間工学の両研究分野で長年議論されてきた。
- ◆本発表では自動車運転中の視覚情報確認において、聴覚情報を視覚情報方向から提示することの効果について議論する。本発表の要点は以下のとおりである。
 - 自動車運転中の視覚情報確認において、視認するべき方向から聴覚情報を提示することにより、素早い情報確認、安定した運転作業が可能となり、また主観的作業負担も軽減する。
 - 視聴覚情報を表示するディスプレイの配置に完全な空間的一致がなくても、視聴覚間のクロスモーダルな空間的対応づけは得られる。視聴覚情報間に±40°ほどの誤差があっても、情報確認支援の効果も期待できる。
 - 聴覚情報が後方から提示されても視覚情報と左右領域が等しければ、同じくクロスモーダルな空間的対応づけは起こりうる。身体後方空間や、身体近傍空間を含めた Crossmodal spatial cuing effect については、学術的課題が多く残り、マルチモーダルユーザーインターフェースデザインの観点からも今後も継続的な議論が必要である。

2 - 5 - 7

2-5-7 ワークプレイスのバイオフィリックデザイン に適した自然音の種類と提示レベルの検討

Study on the types and presentation levels of natural sounds suitable for biophilic design in workplace

〇吉田香純, 辻村壮平(茨城大·工),

北川冬弥, △佐土島哲次, △坂井剛(パナソニック HD)

近年、企業では健康経営が促進され、ウェルビーイングの観点から、バイオフィリックデザイン(以下 BD)が注目されている。BD に関する研究は視覚情報に焦点をあてたものが多く、聴覚情報まではまだ十分に検討されていない。そこで本研究では、音環境を含めた BD の基礎的な知見を得るために、オフィスワーカーを対象とした聴感評価実験を実施し、音源の効果を整理した。因子分析の結果3つの因子(作業性、内発的動機、空間の雰囲気)が抽出され、波音は高い評価であった。鳥の音源は空間の雰囲気を高める可能性がある。また、音源を想起できることが評価に大きく影響することが示唆された。

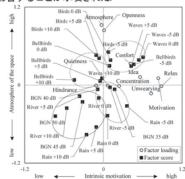


Fig.1:Factor scores and Factor loading for Factor 2 and Factor 3.

2-5-8

2-5-8 集合住宅団地を取り巻く音環境の改善手法 の現状と課題

Current and tasks of sound fields about housing apartment ☆金井樹, △田上健一(九大・芸工)

- ◆高度経済成長期(1955-1972)に住宅の量的供給が行われ大規模集合住宅団地が多く建設された。建て替え・更新の時期が迫っていることに加え近年では質的供給を行う必要があり、環境基準を整理し、次世代の音環境について考察する。
- ◆JIS A1419 に記載される聴感に近い D_wがあり、そのスペクトル調整 項 C_jの算出に A 特性音圧レベル差 X_{jj}を求める必要がある。JIS に指 定されたスペクトル特性では実施的な遮音性能を満足させることが できないため、状況に応じたスペクトル特性を算出する必要がある。
- ◆JIS A1419に指定される空気音遮断性能を満たす集合住宅においても 騒音問題は減少するものの暗騒音とのレベル差が1 dB 程度の新たな 騒音の問題が顕在化している。今後ハード的な技術研究をベースとし ながらもソフト的な研究が今後不可欠になり、従来の遮音性能だけで は解決が難しい。遮音に加え Amenities of Sounds の概念を取り入れ ながら音に対して取捨選択を行い、地域らしさなどの音を残すことで すべての音を取り除くのではなく環境との共生にシフトしていく可 能性がある。
- ◆集合住宅団地においてはその周辺環境の計画が重要であり、視覚情報 も聴覚につながることから今後視覚と聴覚に起因する騒音のマネジ メントの可能性について研究を行っていく必要がある。

2 - 5 - 10

2-5-10 「音空間デザイン」の提案 —提案の背景及びデザインプロセスの検討— Proposal of "Sound space design"

- Background of the proposal and study of the design process-

☆宮本岬、石川あゆみ(岐阜高専)

- ◆様々な音の環境デザインの考え方や課題を踏まえ、音を用いた空間デザイン—「音空間デザイン」を提案する。
- ◆音空間デザインのプロセスを検討し、その計画概要・計画の実現手 法・完成した音空間デザインの評価の具体的な内容といった過程をと りまとめる。
- ◆音を空間構成要素の1つとして捉え、音によって空間をデザインする ことを「音空間デザイン」と定義する。音空間デザインの対象となる 場としては、不特定多数の人が利用する公共空間などが想定され、音 空間デザインにおける主なデザイン手法として、空間に調和した音の 選定などを挙げる。音空間デザインにより期待される効果として、そ の空間らしさを音によって引き出す空間の特色化が考えられる。
- ◆Fig.1 に音空間デザインのプロセスをまとめる。

1. 事前調查	2. テーマ決定と 音空間設計図の作成	3. 音の作成	4. スピーカーの 選択と配置	5. 実践と評価
基本情報の調査・ ・地理的条件 ・空間用途・ ・名用実施 など ・名談音・信号・ 権議者 ・その地や機動な音 ・その地や機動な音 ・との地としたした。など ・したしたした。などを ・インバルスな答測定で発酵 ・対したなどを把きませる。	・テーマに合う音や音楽、 リズム、テーマから連想	トラックに編集して	きる指向性、出力などの スピーカーの配置検討 スピーカーの配置検討 ・選択したスピーカーの 適切な設置場所を検討し、 音空間設計図に書き加える	実数 ・現地でスピーカーの尼恵、 明さ、再生音量を最終調 型したうえで、最終調 サインを実践する 評価 デアンスケート音短問 デアンスケート音短問 デアンスケート音の デアンスケートでは の。 評価をもとに、改奏 で、で、で、で、で、で、で、で、で、で、で、で、で、で、で、で、で、で、で、

Fig.1 Design process for sound space design

2-5-9

2-5-9 パノラマ写真により 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

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

◎石川あゆみ(岐阜高専)

- ◆筆者は、人が建築空間の見た目から予想したその空間の響きを"予想 残響時間"と定義し、建築空間の視覚情報と響きの関連性を検討して きた。
- ◆本報告では、VR ゴーグルを用いて建築空間のパノラマ写真を被験者 に提示した場合の視覚印象および予想残響時間をもとに、建築空間の 視覚情報に調和する残響時間を提案する。
- ◆今回対象とした建築空間の視覚情報の要素のうち、予想残響時間との 相関係数が比較的高く、かつ予想残響時間に対する決定係数が最も高 かったのは、室容積m³であった。

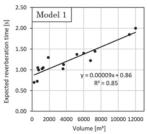


Fig.1: Relation between expected reverberation time and volume

2-5-11

2-5-11 仮想空間による実空間の 再現性に関する研究 -3D モデルで表現された仮想空間の体験に 対する主観評価実験-

Research on the Reproducibility of Real Space by Virtual Space -Subjective evaluation experiment on the experience of virtual space represented by a 3D model-

☆木村健伸, 石川あゆみ(岐阜高専)

- ◆仮想空間にインタラクティブな要素を追加したうえで、実空間を体験する前後の仮想空間の視聴覚印象を測定し、その印象差を分析することで、再現性が高く、かつ違和感のないVR体験を提供するための要素を明らかにすることを目的とする。
- ◆実験で使用した教室の3DモデルをFig.1に示す。
- ◆教室の実空間の体験前後における教室の仮想空間に対する視聴覚印象の差を基に、実空間の再現性が高く、かつ違和感のないVR体験を提供するためには、段差の高さ調整、マテリアルに実空間と同様の素材の画像データの使用、自然光による影の表現とレンダリング、音源位置と距離減衰処理で提示音量を変化させるといった要素が関与していると考えられる。



Fig. 1 3D model of classroom

2-5-12

2-5-12

仮想空間における

音の距離減衰の印象に関する研究 一仮想空間の視覚情報に調和する距離減 衰と残響時間の検証-

Study on the impression of attenuation in distance of sound in virtual space —Verification of attenuation in distance and reverberation time in harmony with visual information in virtual space—

☆風岡翔太, 石川あゆみ(岐阜高専)

- ◆本研究では、建築空間の仮想空間における視聴覚印象の関連性を明ら かにするため、人が建築空間の視覚情報(見た目)から予想した響き を「予想残響時間」と定義し、これを測る被験者実験を行ってきた。
- ◆本報では、Fig. 1 に示す公演等に用いられる大空間を想定して作成した 3D モデルを仮想空間として HMD で体験する場合に、仮想空間内で聴取する音声および音楽に調和する距離減衰と残響時間の組み合わせを明らかにするために行った主観評価実験の結果を報告する。
- ◆コンサートホールの仮想空間で音声・音楽を提示する場合について、 音声・音楽いずれも、仮想空間の容積と吸音率に応じた残響時間と実 空間の距離減衰の組み合わせが視覚情報に調和すると評価された。



Fig.1:3D model of the concert hall

2-7-2

2-7-2 シールド工事に伴う発生振動に関する調査

Investigation on vibration caused by shield tunneling machine 〇田中ひかり、 △竹中計行、 △磯部将吾、 △斉藤功輔、 △山田紀之(大成建設)

- ◆シールド工事において生じた衝撃性の振動を対象として、地盤上およびシールドマシン内の振動、ジャッキに加わる力(歪)を測定した。
- ◆ 測定の結果、振動時にはシールドマシン全体がほぼ同時にトンネル軸 方向の前方に動くことがわかった。
- ◆また、ジャッキの圧縮力が増加するにつれて振動が発生し、その後、 圧縮力が減少した。マシン胴体と地盤間の摩擦力が大きくなると圧縮 力が増加し、圧縮力が摩擦力よりも大きくなると摩擦が切れ、シール ドマシンが勢いよく前方に動くことが衝撃振動の原因と考えられる。
- ◆さらに、ジャッキ付近の胴体に円周上に配置した振動の測定結果より、 衝撃振動が生じる瞬間には、いずれかの位置でトリガーとなる振動が 発生し、それに続いて他の位置が振動していた。また、トリガーが生 じた位置と対向する付近で振動が大きくなる傾向がみられた。

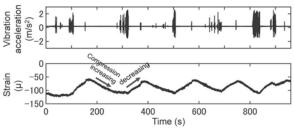


Fig.1: Vibration acceleration near jack and jack strain

2 - 7 - 1

2-7-1 Event-based Visual Microphone: Non-contact Vibration Sensing using Event Camera イベントベースドビジュアルマイクロフォン: イベントカメラを用いた非接触の振動計測

○丹羽遼吾, 伏見龍樹, 山本健太(筑波大学), △落合陽一(筑波大学, ピクシーダストテクノロジーズ株式会社)

非接触測定装置はレーザードップラービブロメータ (LDV) や高速カメラなどが、様々な振動を測定するために開発されてきた. LDV は高価であり、高速カメラは解像度とサンプリング周波数の間で大きなトレードオフがある. そこで、私たちは明るさの変化だけを記録するイベントカメラを用いて振動を測定する非接触でシンプルかつ低価格な手法を提案した. この研究では、イベントカメラを使用して聞こえる音の再構築を実証した. この研究は、非常に明るい照明を必要とせず、コスト効率の良い方法で人間の声などの可聴音を再構築できる振動を測定し、再構築する課題に取り組んだ. この方法は、科学者やエンジニアが振動や音を測定し再構築するためのコスト効率の良い非接触手法を開発するための新たな道を提供する.

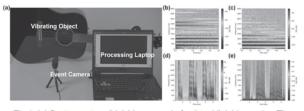


Fig.1:(a) System setup, (b) (c) is a sound of guitar, (d) (e) is a voice. The middle column is the sound recorded by the microphone, and the right column is the sound restored by the proposed method.

2-7-3

2-7-3 ドローンを用いた音響測定に関する検討 :バルーンおよび遮音板の付加による飛行騒音低減

A study on acoustic measurement with UAVs : Flight noise reduction using balloons and sound insulation panels 〇小泉 穂高, 松岡 明彦, 小林 正明, △本木 章平(戸田建設)

◆ ドローンにマイクを搭載し音響測定に活用することを目的とし、ドローン発生音の低減方法について検討した。ドローンにバルーンを付加することで、発生騒音が低減されることが分かった。付加した浮力の大きさによって、騒音低減効果は変化した。ドローンとマイクの間に遮音板を設けることで、さらに発生騒音の影響を低減することが可能であった。

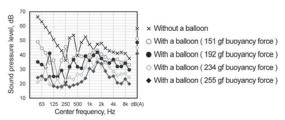


Figure 1: Flight noise of the UAV with a balloon.

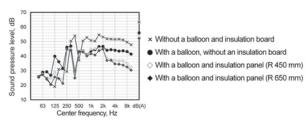


Figure 2: Flight noise of the UAV with a balloon and insulation panel.

2-7-4

2-7-4 自律移動型ロボットの音響計測その2 - 床面材質の違いが 騒音特性に与える影響 -

Sound measurements of autonomous robots. Part-2 -Effect of different floor materials on noise characteristics -

☆増渕優太(神奈川工科大), 中島康貴(リオン)廣江正明(小林理研),
△脇田敏裕上田麻理(神奈川工科大)

- ◆本研究では、移動型ロボットの床面材質の違いによって与える影響について騒音測定を行った。
- ◆床面との干渉や距離減衰のため、マイクロホン位置により最大騒音レベルはばらついているが、床面毎に音圧が違う傾向が現れている。

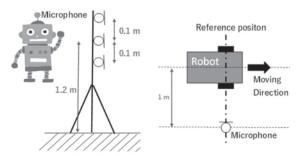


Fig.1 Configuration of measurements

2-7-6

2-7-6 State Detection of Factory Transmission Based on Acoustic Deep Learning Method

☆ Li Xinpei, Di Hao, Yasuhiro Oikawa (Waseda University), Shoichi Kiya (Komatsu)

◆Sound is appealing in automatic and proactive detection of transmission state as it can provide information about the internals of the transmission than visual inspection. Therefore, we use transmission recording sound to classify the working state and design a system. Table 1 summarizes the results of our system. Among the four model, AST has the highest AC of 87.51% and F1 of 86.51%. For the audio-based classification task in this paper, employing pre-trained model outperforms the conventional CNN models.

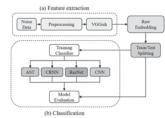


Fig.1: Overview of the classification system.

Table.1: Test results of the classification system.

Model	AC	F1
CNN (Baseline)	53.13%	54.26%
CRNN	65.63%	65.01%
AST	87.51%	86.51%
ResNet18	86.58%	85.88%

2-7-5

2-7-5 加速度及び角速度に基づくロバストな 扉開閉動作の状態推定

Prediction of door state and movement pattern during opening/closing using acceleration and angular velocity data

◎岡本怜奈(IC)、テオリン・アクセル、有光哲彦(フィート)

- ◆高齢者およびその家族にとって負担の少ない見守り装置が必要とされているが、 扉開閉時において動作の初期状態の推定が課題であった。
- ◆本研究では、扉開閉時の振動波形に着目し、ルールベース、機械学習 手法およびニューラルネットワークを用いることで、動作主や扉の種 類に依存しない開閉の状態推定を行う。
- ◆オフィス、ロッカーおよび住宅扉に設置した加速度・角速度センサから4種類の開閉を計測する。扉開閉方向の加速度と回転方向角速度の 2軸 (Fig. 1) をルールベース (RB)、K近傍法 (KN)、シェイプレット変換 (ST)、CNN、FCN、LSTMFCN により分類し、結果を比較した。
- ◆F値はFCNによる推定精度が最も高く 0.9 以上であった。また、R BとKNの手法を組み合わせることで精度を改善でき、扉の種類によらず同様の傾向が見られた(Fig. 2)。

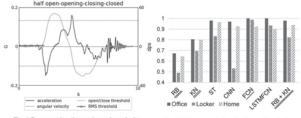


Fig. 1 Door acceleration and angular velocity

Fig. 2 F1-scores of classifications of each door

2-7-7

2-7-7 身の回り VHF 音 -VHF 音の発生源調査-

VHF sounds around us – VHF sound sources survey

☆向井基貴, 青木大成(神奈川工科大), 廣江正明(小林理研),

中村健太郎(東京工業大), 上田麻理(神奈川工科大)

- ◆ 先行研究で公共交通機関や太陽光発電施設等、家電用品などから出ている屋内外の VHF 音発生を調査してきた。
- ◆ 本研究では、在来鉄道および自動車コーナーセンサから出ているVHF音の 周波数特性とその発生源を調査した。
- ◆Fig.1 は在来鉄道沿線で電車の通過時に収録した走行音のサウンドスペクトログラムである。通過中、15~35 kHz の周波数帯域で、かつ連続的な約60dBの音圧レベルのVHF音が含まれていた。

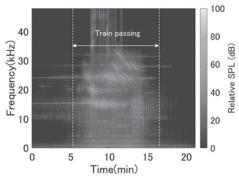


Fig.1 The spectrogram of sound radiated from a passing train of conventional railway.

2-7-8

2-7-8 ダミーヘッド(HATS)を用いた VHF 領域の 耳介周りの音響計測

-正中面と水平面の HRTF の比較-*

春澤恒輝(神奈川工科大),中村健太郎(東工大),上田麻理(神奈川工科大)

- ◆先行研究では詳細な水平面の HRTF を算出し、頭部・耳介形状が HRTF に与える影響を明らかにした。
- ◆本研究では正中面の詳細な HRTF を求め、その特徴を考察すると共に、 水平面の HRTF との比較・考察を行った。
- ◆TSP 信号を用いて、HATS のあり・なしにおけるインパルス計測をし、 両者の差から HRTF を求めた.
- ◆水平面,正中面のいずれのHRTFにも角度の、φに依存しない外耳道 長さに起因した共鳴周波数と推察されるピークが存在した。
- ◆正中面の HRTF より, 耳介の上下非対称形状に起因した反射や干渉の 違いから周波数特性の仰角依存性が生じている可能性が示唆された.

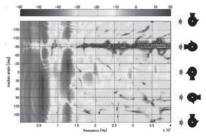


Fig.1 Contour maps of HRTF (SAMURAI HATS) in median planes.

2-8-2

2-8-2 波動音響解析と 3D プリンタを用いた マイクロスリットによる音響メタサーフェス 吸音体パネルの開発

Development of microslit acoustic metasurface absorbers using wavebased acoustics simulation and 3D printer

○奥園 健(神戸大院·工)

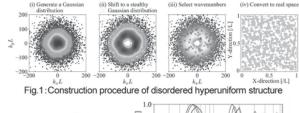
- ◆ 各種音響メタサーフェス (AMS) 吸音体の中でも、マイクロスリットによる微細穿孔を活用する AMS 吸音体に着目する。
- ◆ 低域の室内音響調整のための理論予測が可能、かつ、吸音 特性の調整が容易な AMS 吸音パネルを理論インピーダンス モデル並びに効果的な吸音体配置とあわせて提案する。
- ◆ まず、等価流体モデルに基づく理論インピーダンスモデル を提示し、熱粘性音響シミュレーションを用いてその妥当性 を示す。
- ◆ 次に、3D プリンタを用いて AMS 吸音体のプロトタイプを 試作し、音響管法による測定値が理論・熱粘性音響シミュレー ションと概ね一致することを示す。
- ◆ 最後に、提案する AMS 吸音体が低域の室内音響調整に効果的であることを波動音響解析を用いて示す。

2-8-1

2-8-1 有限要素法解析を用いた無秩序超一様 (hyperuniform)な微細周期構造の 吸音特性に関する検討

A fundamental study on sound absorption characteristics of disordered hyperuniform microperiodic structures using the finite element analysis 〇菅原彬子、△堂寺知成(近畿大)

- ◆無秩序な超一様(hyperuniform)系は、液体のようなランダムで等方 的な性質と、結晶のような長距離秩序を両立した特異な状態をなす。
- ◆単位格子の周期性を満たす多数のランダム波からなる無秩序超一様な微細周期構造を提案し、数値解析・実験により吸音率を算出した。
- ◆ステルス性の導入やランダム波の波数分布の調整により、有用な吸音 材となる可能性が示唆された。



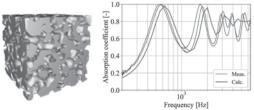


Fig.2: An example of stealthy hyperuniform models and sound absorption coefficient

2-8-3

2-8-3 ネックを延長したマイクロスリットパネル共鳴器 を埋め込んだメタポーラス吸音体に関する

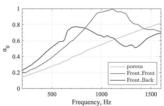
実験的研究

Experimental study on metaporous sound absorbers embedded neckextended microslit resonators.

☆藤井倫太郎, 奥園健(神戸大院・工学研)

- 多孔質材における吸音の低域化と構造の薄さの両立の限界を克服する吸音体として、メタポーラス(MP)吸音体に着目し、その構造について実験的に検討を行った.
- 同一厚さの多孔質材と比較して、より低域で高い吸音を得る MP 吸音体の実現のために、スリットネックを延長したマイクロスリット共鳴器(MSR)を試作し、その吸音特性を音響管法により検証した。
- 試作した MSR を多孔質材に埋め込んだ MP 吸音体を提案 し、多孔質材に埋め込む MSR のレイアウトの違いや、ネッ クの延長が吸音特性に与える影響について明らかにした。
- 本検討で効果的であった 2 種のレイアウト (Front_Front, Front_Back)の吸音パネルのプロトタイプを 3D プリンタを

用いて試作し、同一厚さの多孔質材(porous)と比べて低域でより高い吸音が得られることを確認した(右図)。



2-8-4

2-8-4 ヘルムホルツ共鳴器の格子状配列による 音響メタサーフェスの広帯域吸音

Broadband sound absorption by acoustic metasurface of planar array of small Helmholtz resonators

〇中西伸介(広島工業大)

◆ヘルムホルツ共鳴器を格子状に配列した音響メタサーフェスの吸音 率に Peng et. al.による定式化を適用し、前報の実測値と合わせて考察 した結果、吸音特性の振る舞いがおおよそ一致した (Fig.1).

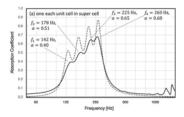


Fig.1 Measured and calculated absorption coefficient of acoustic metasurface.

◆互いに異なる共鳴周波数を有する 7 つの共鳴器を組み合わせて吸音 特性の広帯域化を検討した結果、それぞれの共鳴器の個数を調整する ことで平坦な吸音特性が実現できることが示唆された(Fig.2).

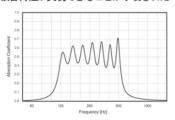


Fig. 2 Calculated absorption coefficient of broadband acoustic metasurface.

2 - 8 - 6

2-8-6 3D たて編構造による静音化向けたて編吸 音材の開発 - 保育施設での発育サポート 用テント型音響製品への応用事例-

Development of the Acoustic Absorption Materials using 3D Warp Knitting Structures for Noise Reduction —Application to Tent-type Acoustic Products to Support Child Development—

○ 寛 瑞恵(福井県工業技術センター)

- ◆本研究では、仮撚り加工糸のちぢれ形態を利用して、特殊な立体編地 構造により広帯域な吸音性を有する「たて編吸音材」を開発した。
- ◆さらに、たて編吸音材を利用し、保育施設向け幼児の発育サポート用 テント型音響製品「カルム」を開発した。吸音試験の結果、250~2500 Hz を含む子どもの高い声を適度に静音化できる吸音性が実現できた。 ※「たて編吸音材」は、福井県工業技術センターと福井経編県業株式会社の共同開発製品です。

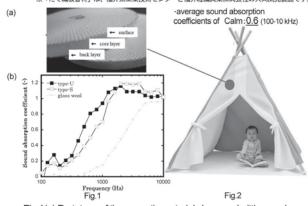


Fig.1(a) Prototype of the acoustic materials by warp knitting, and (b) sound absorption coefficients of 2 types, type-U and type-S. Fig.2 Image of "Calm" (produced by JAKUETS Inc., FUKUI)

2-8-5

2-8-5 スリットを有する立方格子構造からなる 微細周期構造の吸音特性に与える影響 に関する基礎的検討

Study micro-periodic structure consisting of simple cubic unit cells lattice with slits on sound absorption characteristics

☆城修平(近畿大院), 菅原彬子, 平栗靖浩(近畿大)

- ◆多孔質吸音材や共鳴器型吸音材にはそれぞれ長所と短所がある。そこで、既存の吸音材の長所を活かし、短所を補う新たな吸音材の開発ができるのではないかと考えた。
- ◆本研究では上記の新たな吸音材の開発を「音響メタマテリアル」により実現し、優れた吸音特性を有する微細周期構造を提案することを目的とする。
- ◆本研究では、スリットを有する微細周期構造を提案し(Fig. 1)、その吸音特性を有限要素法により数値解析を行った。その結果、スリットの数や開口率の変化により、低周波数帯での吸音率の向上につながる可能性が示唆された(Fig. 2)。

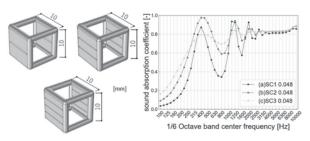


Fig.1:Study model

Fig.2: Numerical analysis with varying number of slits

2 - 8 - 7

2-8-7 ボロノイ分割を利用した拡散体の音響拡散特性に関する研究

Study on acoustic diffusion characteristics of amorphous diffusers using by Voronoi tesselations

☆濱谷晃希(近畿大学院), 菅原彬子, 平栗靖浩(近畿大学)

- ◆背景:周期を持たない形状を音響拡散体に利用することで、広帯域で 無指向性に近い拡散特性が得られる可能性がある。
- ◆目的: ボロノイ分割を利用して非晶質な形状の音響拡散体を作成し、 その音響拡散特性について検討する。
- ◆内容:有限要素法を用いた数値解析より、散乱係数を算出した。 非晶質な形状のうち、斜面を持つ試料(3D)と持たない試料(2D) を比較すると(Fig.1)、3D では630 Hz 以上の帯域で高い散乱係 数を示した。ここから、試料の斜面が散乱特性に寄与している と考えられる。

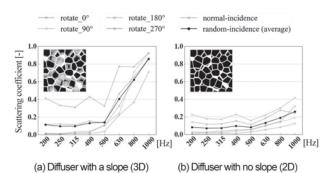


Fig.1:Scattering coefficient calculated by numerical analysis.

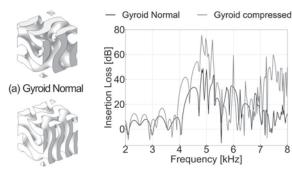
2-8-8

2-8-8 有限要素法を用いたジャイロイドの 遮音特性に関する基礎的検討

Fundamental study on sound insulation characteristic of gyroid using finite element method

☆矢野拓実(近畿大院), 菅原彬子, 平栗靖浩(近畿大)

- ◆蝶の翅の構造に存在するジャイロイドは、バンドギャップ構造として 特定の波長を散乱、反射させる働きがある。これを音波に応用できる と考えた。
- ◆ジャイロイドの幾何パラメータを変化させたモデルを作成し(Fig. 1(a),(b))、遮音特性を検討、評価を行う。
- ◆有限要素法を用いた数値解析により(Fig.1(c))、ジャイロイドが特定の 周波数に対してバンドギャップを形成することがわかった。また、ジャイロイドの層を増やすことで遮音性能の向上を確認でき、それを圧縮することで同じ厚みの状態で挿入損失の増加ができた。



(b) Gyroid Compressed

(c) Analysis result of Gyroid

Fig. 1 Gyroid model and numerical analysis results

2 - 8 - 10

2-8-10 二本のカーディオイドマイクロホンと音響管による 垂直入射吸音率の測定

〇星 和磨, 羽入 敏樹 (日大・短大)

音響管内でカーディオイドマイクロホン 2 本を組み合わせ、垂直入射吸音率を測定する方法を紹介する。吸音率の測定に際し、予めカーディオイドマイクロホンで得られる信号の音圧と粒子速度それぞれの較正値 α と β を得ておく必要がある。本報告では、音響管を使って α と β を計測し、さらに材の垂直入射吸音率を測定できるか検証した。

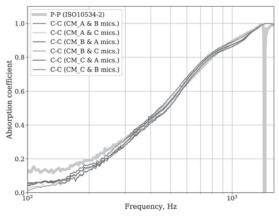


Fig. 1 Measurement results of 50mm-thick glasswool using three cardioid microphones (combined two cardioid mics. from three ones).

2-8-9

2-8-9 共鳴器を内蔵した吸音壁の 吸音特性に関する検討

-縮尺および実大模型を用いた検討-

A study on sound absorption characteristics of the sound absorbing wall with built-in resonators

- : experiments with scale model and full-scale model.
 - ◎神谷優, 財満健史, 黒木拓, 大脇雅直(熊谷組)
- ◆工場の設備音等の対策として、水や粉塵の影響を受けにくい吸音機構が必要とされる。本報では共鳴器を仕上げ壁に内蔵することで水や粉塵の影響を防ぎ、共鳴器と壁の板振動による吸音を同時に利用する吸音壁の吸音特性について、縮尺および実大模型を用いて検討した。
- ◆両模型にて模擬壁端部に開口を設けた場合に共鳴器の共鳴周波数よりもやや低い周波数帯域で吸音率が上昇した。一方で、共鳴器による 吸音帯域や吸音率の値には両模型で差が見られた。
- ◆実大模型にて開口幅が広いほど共鳴器の効果が発揮される傾向が見られた。また、開口が広がっていくと共鳴器の共鳴周波数帯域により近い帯域でも吸音率が上昇する傾向が見られた。

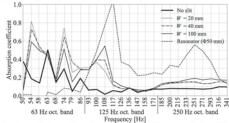


Fig. 1: Sound absorption characteristics of the sound absorbing wall (Full-scale model)

W: Opening width, Φ : Diameter of resonator aperture

2-8-11

2-8-11 多重共鳴器構造における 貫通孔形状が吸音特性に与える影響

Effect of central hole shape on sound absorption characteristics in multiple resonator array

〇靍羽琢元(竹中技研)

- ◆多重共鳴器構造において共鳴器の配列順序、すなわち貫通孔の形状が 吸音特性に影響を与える可能性が考えられる。その影響を概括的に把 握するため、解析的手法を用いて代表的な貫通孔形状の吸音特性を調 査した結果を報告する。
- ◆異なる貫通孔形状をもつ多重共鳴器構造の吸音率を計算した結果、低 周波数の吸音率を高めるためには貫通孔の径を均一にすることが有 効であり、高周波数の吸音率を高めるためには終端に近づくにしたが い径が減少する貫通孔形状とすることが有効であると示唆された。

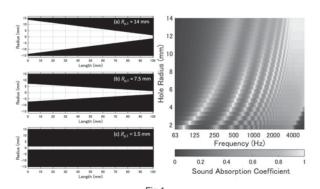


Fig.1

Left: Multiple resonator arrays with various central hole shape

Right: Calculated sound absorption coefficients of multiple resonator arrays

2-8-12

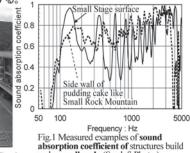
2-8-12 細砕石積層体の音響反射吸音特性の観測

-ロームシアター京都に仮設された 「石ころの庭」の構築物を例として-

Observation of sound reflection and absorption characteristics on structures build-up by crushed small gravel - Example of "Small Rock Garden" temporally build in ROHM Theater KYOTO

- ○岩瀬昭雄(新潟大学)、△岩瀬諒子(京都大学院·岩瀬諒子設計事務所)、 △遠藤郁(岩瀬諒子設計事務所)
- ◆2022 年秋にロームシアター京都の中庭に細砕石を積層して小ステー ジや観覧席、ピラッミド型等のモニュメント群を配したイベント空間 「石ころの庭」が実現、構築物の吸音率等の音響特性を現場計測した。
- ◆ステージ等の積層水平面の計測結果の例を Fig.1 に示すが、吸音率の 高いピークとディップを繰り返す周波数特性が特徴的である。斜壁面 の場合は吸音率がより高く落込みディップは深くない。音響管法によ る積層砕石の音響計測も実施、砕石層内の音響伝搬定数や特性インピ 一ダンスも得られて、任意に積層構成の音響特性が推定可能となった。
- ◆素材として硬い砕石も積層構造では閉開管共鳴に類似の多次周波数 列で高吸音特性を発揮する。





up by small rock. (See left Photo.)

Photo. 1 View of Small Rock Garden.

2 - 9 - 2

2-9-2 CFVC: 制御可能な音声変換のための 条件付きフィルタリング

CFVC: Conditional Filtering for Controllable Voice Conversion 〇田中宏, 金子卓弘, 亀岡弘和, 関翔吾(NTT)

近年、より豊かなコミュニケーションを実現する技術として、音声 の各特徴量を所望の特徴量へと変換する音声変換が注目されており. 発話速度や基本周波数 (F0) などの高レベルの属性の効率的な制御方 法も活発な研究分野である. 例えば、声質を保持したまま発話速度の みを制御するためには、話者ベクトルを発話速度に関する成分とその 他の成分に分解することが不可欠である.

しかし、このような分解された話者表現を、アノテーションデータ がないあるいは極小の状態で学習させることは難題であった. 本稿で は、事前の統計量計算やアノテーションを用いずに、話者の声質を保 ちながら話速などのハイレベルな属性を制御する音声変換手法を提 案する.



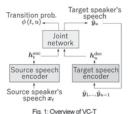
2 - 9 - 1

2-9-1 VC-T: ニューラルトランスデューサー に基づくストリーミング声質変換

VC-T: Streaming voice conversion based on neural transducer ◎金川裕紀, 森谷崇史, 井島勇祐 (NTT)

- ◆Sequence-to-sequence (seq2seq)に基づくパラレル声質変換(VC)
- ソース・ターゲット話者の事前アラインメントなしに直接学習可能
 - → DTW の精度に依らず、話速も含めてモデル化できる

 の
- 注意行列が単調にならない場合にアラインメントが破綻
- ◆ (**提案法**)ニューラルトランスデューサーに基づく VC (VC-T)
- アラインメントの破綻とストリーミング動作に優れるニューラル トランスデューサー(RNNT)を VC に初めて組み込む
- 音声合成用に提案された生成用 RNNT アルゴリズムに適合 するよう、VC 用のアライメントを設計



"M2F", "F2M" denote Female2-to-Female1,
Male1-to-Female1, and Female1-to-Male1 scenario,
respectively. The scores written in bold signify the
column-wise best.

Table 1: Averaged character error rate (CER), "F2F",

III Cu loca	_	1,00		w.v.a.	
GROUDTRUTH	12.9	14.1	14.5	14.0	
RESYN	13.2	14.8	15.0	14.5	
BNE-S2SMoL-VC [Liu+'21]	25.6	23.8	26.8	25.4	
CONVS2S-VC (OFFLINE) [Kameoka+ "20]	16.9	21.3	24.4	21.4	
CONVS2S-VC (STREAM)	16.9	27.4	28.5	25.2	
VC-T(OFFLINE)	14.3	16.7	17.9	16.6	
VC-T(STREAM)	14.4	18.5	19.4	17.9	4

従来法で見られた言語情報の著しい損失なしに 提案法はストリーミング動作を達成!

2 - 9 - 3

2-9-3 出力遅延と時間伸縮変換を考慮した リアルタイム声質変換の検討

Real-time voice conversion in consideration of output delay and time warping transformation

☆鈴木耀, 藤本崇人, 高木信二, 橋本佳, 南角吉彦, 徳田恵一(名工大)

- ◆Sequence-to-sequence 声質変換モデル
 - ▶ 異なる長さの元話者・目標話者であっても扱うことができ、声質 や話速、韻律を同時変換可能
 - ▶ 入力系列全体を必要とし、入力音声の発話が終了するまで変換音 声を生成できないため、リアルタイム声質変換は不可能
- ◆提案手法の概要
 - リアルタイム声質変換に向け、すでに入力された音声のみを参照 するアテンション機構に基づく声質変換モデルを使用
 - ⇒ 元話者の発話完了を待たずに変換が行えるが、元話者の話速が 目標話者よりも遅い場合,変換時にエラーが発生 (Fig. 1(a))
 - ▶ 目標話者の話者性を保ちながら、無音を伸縮させて元話者との時 間的因果関係が取れた音声データを音声合成システムにより生成 し、 擬似目標話者データとして学習に使用 (Fig. 1 (b))
 - ⇒ 時間伸縮変換を考慮したリアルタイム声質変換を実現

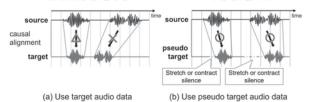


Fig.1: Overview of proposed method.

2 - 9 - 4

2-9-4 ASR および F0 一貫性ロスに基づく 正則化ランジュバン動力学音声変換

Regularized Langevin Dynamics for Voice Conversion Based on ASR and F0 Consistency Loss

☆中村 泰貴 (東大院・情報理工), 亀岡弘和、金子卓弘、田中宏、関翔悟(日本電信電話)

- ◆ VoiceGrad はスコアベースのノンパラレル声質変換
 - > ランジュバン動力学による入力音声特徴量の変換
 - ▶ 変換前後で発音や抑揚が保持される保証がないため、歪んだ変換 音声を生成してしまう欠点が存在
- ◆ 本研究では、VoiceGrad において変換前後で入力音声に含まれる 発音や抑揚を保持できる正則化項を 2 つ提案
- ◆ 提案する正則化項の導入により、通常の VoiceGrad と比較し、 自然性、明瞭性、話者再現度、抑揚再現度が高い音声を合成可能

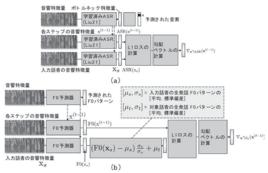


Fig. 1:提案する正則化項の概要図. (a) 学習済みASR [Liu21] を用いた ASR 一貫性ロス. (b) 学習済み F_0 予測器を用いた F_0 一貫性ロス.

2 - 9 - 6

2-9-6 End-to-End 音声要約のための転移学習に 基づく外部言語モデルの活用

Leveraging External Language Models via Transfer Learning for End-to-End Speech Summarization

◎松浦孝平, 芦原孝典, 森谷崇史, 田中智大, 叶高朋, 小川厚徳, マーク・デルクロア(NTT)

- ◆音声の要約を生成する音声要約技術が近年注目を集めている。
- ◆音声から要約を直接生成する End-to-End 型モデルは動作の軽量さ や音声認識誤りへの頑健性から有望であるが、学習に必要な大量の音 声と要約文のペアデータの収集は困難である.
- ◆そこで、本研究では大規模な教師なしテキストで事前学習された外部 言語モデルを、転移学習によって End-to-End 音声要約モデルへ組み 込むことで、要約精度を改善する方法を提案する. (Fig. 1)
- ◆How2 データセットにおける実験において、Table 1 に示すように提案手法 (P-1) はデータ拡張 (B-2) と Cascade モデル (C-1) を上回る要約精度を達成した。

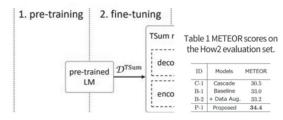


Fig. 1 Our proposed method composed of three stages: pre-training, fine-tuning, and transfer learning.

2 - 9 - 5

省メモリな音声・テキストエンコーダー

を用いた End-to-end 音声要約の提案

speech summarization of long spoken document: Improving memory efficiency of speech/text encoders.

◎叶高朋, 小川厚徳, マーク・デルクロア(NTT), 渡部晋治(カーネギーメロン大学)

- ◆End-to-end 音声要約では、長い音声を一度にエンコードするため大量 の計算資源を必要とする。そのためハードウェアの制約により、数分 に及ぶ音声を要約することが困難である。
- ◆本研究では、省メモリなネットワークを活用し、音声とテキスト2種類のエンコーダーを持つ End-to-end 音声要約モデルを提案する.
- ◆従来法は 100 秒までの音声しか要約できないが、提案法は、10 分の 音声を要約することが可能であり、音声全体の文脈情報を考慮するこ とで高精度な音声要約を実現した。 Summary

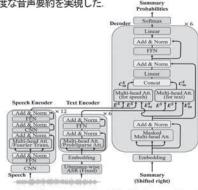


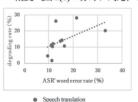
Fig.1: Overview of the proposed model of dual speech/text encoders model. The Fourier Trans. and ProbSparce Att. are memory-efficient

2-9-7

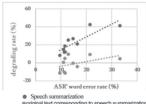
2-9-7 TED 講演音声の音声認識性能と音声翻訳 および重要文抽出要約の性能の関係

〇中川聖一, 足立十一郎, 桜井陽生, 坂野晴彦, 山本一公 (中部大)

- ◆英語の講義や講演音声を自動音声認識し、要約して日本語で字幕を表示し、理解を支援するシステムの開発を行ってきた。
- ◆以前の研究では、TOEIC 750 点程度の学生でも英語講義の正しい書き 起こし率は約60%程度、翻訳精度はBLEU 値でテキスト入力で15程 度、音声入力で8程度であった(言語処理学会、第23回大会、2017.3)。
- ◆TED 英語講演の音声認識率とテキスト入力および音声入力による英 日機械翻訳、重要文抽出要約を行い、これらの性能の関係を分析した。
- ◆音声入力にる音声翻訳性能の低下は音声単語認識誤り率(WER)が 10%程度で、BLEU の低下率は10%程度であった。WER 5%程度なら、 翻訳性能の低下率は5%程度に抑えられると考えられる。
- ◆重要文抽出要約では WER が 10~20%程度なら、重要文の抽出箇所の 精度 (図 1(b)の赤印に対応) に影響はないことがわかった。







eoriginal text corresponding to speech summarization

(b) WER and summarization

Fig.1:Relationship between ASR' WER and degradation rate of speech translation/summarization' performance compared with text translation/summarization' performance

2-9-8

2-9-8 Teacher-Forcing により歌詞を与えた際の Attention の崩れに着目した替え歌検知

Parody Detection Focusing on Attention Deviation with Teacher-Forced Lyrics

☆有質智輝、樋口陽祐 (早大), 早坂一寿, 岡本直紀 (第一興商), 小林哲則, 小川哲司 (早大)

- ◆研究目的: 替え歌の自動検知システムを構築する
- ◆提案手法: Attention の崩れに着目した替え歌検知
 - ▶ 歌唱音声に対し正解の歌詞をTeacher-Forcingで与えることが可能
 - ➤ 歌唱内に替え歌が含まれる場合には、歌詞に対応する音声信号が 存在しないために Attention が崩れることを利用して替え歌検知
- ◆実験: 替え歌を含む1フレーズ歌唱音声を用いた検知実験
 - ➤ 正しい歌唱では右肩下がりに Attention がかかり、 替え歌箇所では Attention が崩れる (Fig. 1)
 - ▶ 従来手法と比較して検知性能の向上を確認 (AUC 0.963 vs 0.875)
 - ▶ 替え歌箇所が少ない場合にも高い検知性能を示した

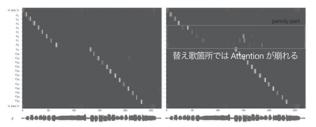


Fig.1: Visualisation of attention weight of singing voice with parody.

2 - 9 - 10

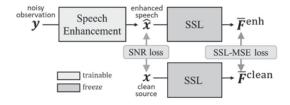
2-9-10 後段タスクに非依存な音声強調手法の 検討

General Speech Enhancement with Self-Supervised Representation Domain Loss

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

◆概要

- ・音声の教師なし事前学習(SSL)の後段タスクの耐雑音性向上
- ・シングルチャンネル音声強調を SSL のフロントエンドとして使用
- ・音声強調と SSL のミスマッチを改善する SSL-MSE Loss を提案



◆評価実験

· SSL-MSE 損失により後段タスクの性能を改善

タスク	ASR	PR	ASV	IC	SF
評価指標	WER↓	PER↓	EER↓	Acc↑	F1 ↑
clean	5.6	4.4	4.5	98.8	90.3
noisy	17.1	16	10.9	67.4	82.4
+SE frontend	14.6	9.9	8.6	84.8	82.4
+SSL-MSE	13.1	8.8	8.0	87.6	83.5

2 - 9 - 9

SpeechGLUE:音声向け自己教師あり学習は どの程度言語知識を捉えるか

SpeechGLUE: How Well Can Self-Supervised Speech Models Capture Linguistic Knowledge?

◎芦原孝典、森谷崇史、松浦孝平、田中智大、井島勇祐、浅見太一、 デルクロアマーク、本間幸徳(日本電信電話株式会社)

- ◆自己教師あり学習 (SSL) によって多様な音声タスクの精度が改善.
- ◆最近では音声言語理解タスクの精度も改善していることから、音声 SSL モデル自体が一定の言語知識を捉えている可能性が示唆される.
- ◆ そこで本稿では、NLP 分野のベンチマークである GLUE の音声版と して SpeechGLUE を構築し (Figure 1), 音声 SSL モデルのプロービ ングを通じて、どの程度言語知識を捉えているのかを分析した。
- ◆SpeechGLUE の構築には単一話者のTTS を利用し、様々な音響的変動を抑制することで、音声SSL の言語知識だけに着目した。
- ◆実験から、音声 SSL モデルは、テキスト SSL モデルよりは劣るもの の、ベースラインよりも高い精度を達成した (Table 1). この結果から、音声 SSL モデルは音声データだけからある程度の一般的な言語 知識を獲得していることが示唆された。

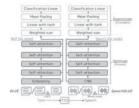


Table 1: Summary of averaged score on the development set of GLUE and SpeechGLUE.

Upstream	Avg. w/o WNLI
FBANK	45.9
WavLM	<u>68.1</u>
BERT	74.0

Figure 1: Schematic diagrams of the evaluation system for GLUE and SpeechGLUE.

2-9-11

2-9-11 深層学習で獲得される音声シンボルは 自然言語シンボルと同様に

Zipf 則に従うか?

Do learned speech symbols follow the Zipf's law as well as natural language symbols?

前田 紘希, ○高道 慎之介, 朴 浚鎔, 猿渡 洋 (東京大学)

解答)

特定条件で Zipf 則ではなく冪乗則に従う

2-9-12

2-9-12 声門の閉鎖期の音声信号による IAIF に基づく声門流推定

IAIF based glottal flow estimation using the speech signal in the glottal closed phase

☆河内太一, モクタリ パーハム, 森川大輔(富山県立大)

- ◆Iterative Adaptive Inverse Filtering (IAIF) に基づく声門流推定は声門 の開放期の影響によりスペクトル包絡のモデリングが正確ではない。
- ◆ そこで、IAIF と IOP-IAIF について、 声門の閉鎖期 (CP: Closed Phase) のみを分析区間として用いる IAIF-CP と IOP-IAIF-CP を提案し、 推定 精度を比較した。
- ◆IAIF と IAIF-CP, IOP-IAIF と IOP-IAIF-CP の推定誤差はそれぞれ2.0%, 2.6%小さくなった。

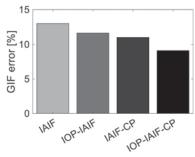


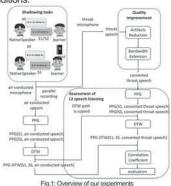
Fig.1: glottal flow estimation error by conventional and proposed methods

2-9-14

2-9-14 Integration of Throat Microphone Recording and Bandwidth Extension for Robust Assessment of L2 Speech

O Yu Xu, Haitong Sun, Yingxiang Gao, Minematsu Nobuaki, Daisuke Saito (The University of Tokyo, Electrical Engineering and Information Systems)

- A novel approach that applies the throat microphone to L2 robust listening speech assessment in a crowded environment is proposed.
- Quality of throat speech is improved by artifacts reduction and the pretrained BWE model.
- Performance of the proposed approach is evaluated in the task of assessing L2 shadowing speech.
- Experiment results indicate our proposed approach improves the L2 speech assessment performance of throat speech in some special conditions.



2-9-13

2-9-13 MRI 正中矢状断面画像からの F0 推定の可能性について

F0 Estimation from Midsagittal Plane MRI Images

◎大谷祐人,澤田隼,大村英史,桂田浩一(東京理科大)

我々は real-time MRI (rtMRI)で正中矢状断面を収録した調音運動データを入力とする音声合成法を提案してきた。この手法では rtMRI 動画像からメルスペクトログラムを音声フレームごとに推定し、Foを含めて高い精度で音声を復元している。音声フレームと対応する MRI 画像 4 フレームからの Fo 推定が可能であることは、MRI 画像内に必要な情報が含まれていることを示唆しているが、これはソース・フィルタモデルに直感的に反する。そこで、1 フレームの MRI 画像のみからの Fo 推定が可能であるかを検証し、その推定根拠を Grad-CAM を用いた CNN の可視化により分析する。

ATR503 文 rtMRI データベースを用いて実験を行った結果、MRI 画像から推定した Foの RMSE は22.7 Hz と良好であり、単一の MRI 画像からの Fo 推定が可能であることを確認した。 Grad-CAM を用いた可視化では 110-120 Hz 付近を境に着目箇所が変化することを確認し、その前後で Fo.上昇/下降と連動する運動が観察された。 Fo.上昇時は喉頭が前方・上方に引かれていた。 Fo 下降時は舌が引き込まれ、喉頭が後方・上方に引かれていた。 このことから、 Fo.上昇/下降に影響する運動が正中矢状断面を捉えた MRI 画像に表れることを確認した。

今後はF₀上昇/下降時に観察された運動が声帯やF₀に及ぼす影響を検討する予定である。また、他の話者や言語についても分析を行い、一般性を検証する予定である。

2-9-15

2-9-15 A comparative study of symbol-based PPG and frame-based PPG used for utterance comparison in CALL

CALL を目的とした発話比較における シンボル PPG とフレーム PPG の比較

☆Haitong Sun, Yingxiang Gao, Daisuke Saito, Nobuaki Minematsu(UTokyo), Takehiko Makino (Chuo University), Nakanishi Noriko (Kobe Gakuin University)

- ◆Frame-based PPG-DTW generated by Kaldi showed good but time-consuming performance in comparison-based L2 speech assessment.
- ◆In this study, we employed a universal phone recognizer Allosaurus to generate symbol-level PPG. Symbol-based PPG-DTW (Allo-PPG-DTW) were compared with the symbol-level results of DTW calculated by frame-based PPG generated by Kaldi on utterance pairs as figure 1.
- Our experiments showed that in some cases although the correlation between the results of symbol-based PPG-DTW and frame-based PPG-DTW is not very high, symbol-based PPG-DTW can detect most of the peaks in frame-based PPG-DTW.

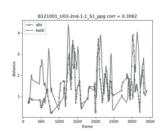


Fig.1: Allo-PPG-DTW and Kaldi-PPG-DTW on an example S-SS utterance pair

2-10-1 【招待講演】声質制御性の高い 音声分析合成のこれまでとこれから

Past and future of speech analysis-synthesis systems with flexible controllability of voice quality

〇坂野秀樹 (名城大)

- ◆音声分析合成の応用について、入力音声の声質を何らかの形で制御して合成する手法を中心に紹介する
 - ▶ 短時間位相の群遅延による制御
 - ▶ 音声モーフィング
 - > 声道断面積関数による声質制御
 - ▶ 動的特徴の強調による音声強調
- ◆現在は、DNN 音声合成やニューラルボコーダーなど、機械学習ベース の手法の一部に高品質音声分析合成が組み込まれることも多くなっ ている
 - ➤ ニューラルボコーダーにおいては、WaveNet などの波形生成方式を 組み合わせることによって、元となる音声分析合成より高い品質 の音声が合成されている点が興味深い
 - ▶ 間接的に短時間位相情報が復元されている効果も含まれていると 考えられ、個人的にはこの点を追究していきたい

2-10-3

2-10-3 声帯の厚さと喉頭位置を独立に変えた 歌声の音響分析と聴覚印象

Acoustic Analysis and Perceptual Impression of Singing Voices with Independently Manipulated Vocal Fold Thickness and Laryngeal Position

〇高野佐代子*、長塚全**、土田義郎*

Sayoko TAKANO*, Zen Nagatsuka** and Yoshio TSUCHIDA*

(*Kanazawa Institute of Technology, **Zen Voice Factory)

- ◆ 歌声に関して代表的な音源の特徴として、地声と裏声があげられている。一般に、声が低い地声の時は喉頭位置は低く、声が高い裏声の時は喉頭位置が高い傾向がある。一方で、プロは同じ声の高さであっても、声帯の厚みと喉頭を独立に調整できるという。
- ◆ 音響分析の結果、地声・裏声の違い、喉頭の位置によってそれぞれ 音響パラメータの特徴が表現されていた。
- ◆ また SD 法による聴取実験の結果、一般的な声帯の厚みと喉頭の位置を自然に連動させた場合は軽柔因子に関する印象の違いとなる。一方で、プロによる意識的な声帯の厚さと喉頭の位置のコントロールにより、鮮明因子の違いとなることが明らかになった(Fig.1)。

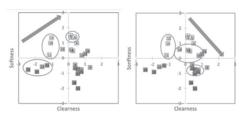


Figure 1. A part of the poster for presentation contest.

2-10-2

(招待講演)聴覚におけるフィルタ分析と 音声了解度予測への応用

Filter analysis in hearing and application to speech intelligibility prediction

O山本克彦(サイバーエージェント AI Lab)

- ◆ヒトの聴覚における情報処理
 - ▶ 周波数情報:基底膜の進行波で対応した場所の有毛細胞が振動
 - > 時間情報: 聴神経の発火確率の時間変化
 - → それぞれの情報処理は『フィルタ分析』として表される
- ◆聴覚モデルベースの音声了解度の客観評価指標
 - ▶ 周波数情報:『聴覚フィルタバンク』で狭帯域信号に分解
 - ▶ 時間情報:振幅包絡を抽出して『変調フィルタバンク』で分析
- ◆本講演では、客観評価指標の最新動向や課題についても解説
 - > The Clarity Project:補聴器向けの音声強調・予測コンペを開催
 - ▶ 大規模な評価データを用いた機械学習ベースの手法が上位を独占



Figure 1: "A dancing cell which has a single hair as a cute, kawaii and funny illustration" generated by Stable Diffusion 2.1 Demo https://hugqingface.co/spaces/stabilityai/stable-diffusion

2-10-4

2-10-4

声道形状が声帯振動へ及ぼす影響

- ◆音源フィルタ理論は、通常の話声などでは声帯音源と声道フィルタの 相互作用は小さく、無視できるとした理論である。
- ◆本稿では、声帯振動の2質量モデルと1次元の時間領域差分法に基づ く声道モデルを用いて、比較的相互作用が小さい場合に声道形状が声 帯振動へ与える影響を検討した。
- ◆声道形状の変化は声道伝達関数を変化させるだけでなく、声道からの 反射波を変化させ、声帯音源波形に影響を与える。音声スペクトルは この両者の影響を反映する。
- ◆そのため、声道形状や伝達関数の変化が小さくても、反射波の影響で 音源スペクトルが変化して音声スペクトルが大きく変化する場合が あると考えられる。

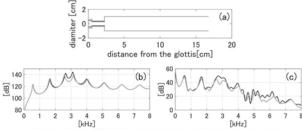


Fig.1:灰色から黒色に声道形状を変化させたときの伝達関数と音声スペクトル (a) 声道形状, (b) 伝達関数, (c) 音声スペクトル

2-10-5

声帯模型の吹鳴実験に関する一考察

A Study on Sounding a Functional Vocal Cords Model 〇青木 直史(北大)

模型教材の活用は五感に対する訴求力に優れており、初学者の興味を引きつけるアプローチとして教育的効果が高い。ひとつの事例として、本研究では、声帯の模型教材として、機能に焦点をあてた実体モデルの作成について検討してきた。本稿では、吹鳴実験を行うことで明らかになった声帯模型の音響的特徴について報告する。



2枚のゴム膜を重ね合わせた声帯模型



デモ動画: https://youtu.be/WBn11IkaXnA

2 - 10 - 7

2-10-7 単純声道モデルにおける仮声帯および 披裂喉頭蓋の狭窄が発声に及ぼす影響

Effects of false and anyepiglottic folds' constrictions on voice production in a simplified vocal tract

〇吉永司(大阪大), △Zhaoyan Zhang(UCLA)

- ◆これまで声帯上部の仮声帯が声帯の振動に与える影響について調べられてきているが、披裂傾頭蓋による狭窄との違いについてはあまり調べられてこなかった。
- ◆本研究では、仮声帯及び披裂喉頭蓋による向きの異なる狭窄を Fig.1 のように単純化した声道内に設定し、気流シミュレーションを行うことにより、声帯による気流と音の発生が狭窄のサイズによりどのように変化するのかを調べた。
- ◆仮声帯および披裂喉頭蓋の狭窄を狭めることにより、声帯の基本周波数や音量に影響があり、仮声帯と披裂喉頭蓋の狭窄で影響が異なることが確認された。

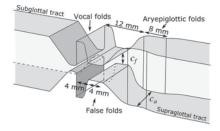


Fig.1:Simplified vocal tract geometry with false and aryepiglottic folds constriction.

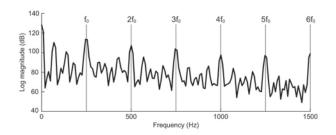
2 - 10 - 6

2-10-6 発声における仮声帯振動の影響に関する数値流体 解析

Numerical fluid analysis on the effect of false-vocal fold oscillation during phonation

〇鏑木時彦, 加藤日花里, 李庸學 (九州大)

- 有声音の音源は、声帯弾性体と呼気流の相互作用による声帯音源である。声帯上部にある仮声帯は、通常の話声では両側に開かれた状態にあるが、喉歌などの歌唱発声では声帯とともに振動し、音声のスペクトルにサブハーモニクスを生じる。
- 本研究では、格子ボルツマン法により、声帯と仮声帯を含む喉頭内流路を通過する2次元の非定常・非圧縮性の流れの数値解析を行なった。 声帯と仮声帯を正弦的に振動させて周期性の自励振動を模擬し、それぞれの出口における流速の時間変化を求めた。
- その結果、仮声帯振動は、声門出口からその下流域における流れに影響することが示された。流速波形のスペクトルには側帯波に相当する 周波数成分が観察され、仮声帯振動は文字通りに流れを振幅変調する 働きを有していると解釈される。



2-10-8

2-10-8 高速度ディジタル撮像を用いた ボーカルフライ声区における声帯振動の分析

Analysis of vocal fold vibration in the vocal fry register

utilizing high-speed digital imaging

☆加藤日花里 李庸學 鏑木時彦 若宮幸平 (九州大)

- ◆ボーカルフライ声区における声帯振動メカニズムを明らかにするため、 高速度ディジタル撮像(HSDI)によって低音域ボーカルフライと高音域ボ ーカルフライを発声中の喉頭を観察し、声門面積波形、声門開放率、ト ポグラフを算出して分析を行った。
- ◆低音域ボーカルフライでは、振動部位は被験者によって異なっていたが、全ての被験者で長い閉鎖期とパルスのように鋭い立ち上がりの開放期からなる声門面積波形が観察された。
- ◆高音域ボーカルフライでは非周期的な声帯振動が観察された。
- ◆低音域、高音域共に声門開放率は地声よりも小さかった。
- ◆ 声門上の絞扼を伴わずに発声されたボーカルフライは、単一パルス状 の声門面積波形を持つ発声となることが示唆された。

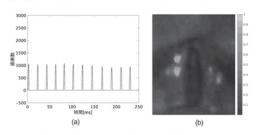


Fig.1:(a) Glottal area waveform and (b) Topogram. Expiratory vocal fry at low frequency. Subject:M1

2-10-9

チューブを利用した音声障害の治療

Voice Therapy using tubes for voice disorders

〇城本修(広島県公立大学法人)

- ◆ これまで、言語聴覚士は、音声障害患者に対する音声訓練として、音源・フィルター理論に基づき、呼吸器、発声器、共鳴器の各器官に個別に訓練を行ってきた。
- ◆ たとえば、腹式呼吸訓練、喉頭の過緊張を緩和する喉頭マッサージ、 咽頭を開くあくび・ため息法などである。
- ◆ しかしながら、各器官別の訓練が音声障害つまり嗄声の改善にはあまり有効でないことも経験的に知っていた。
- ◆ 実際に、最近の報告では、単一の訓練の効果はないとメタアナリシスでも報告されている。
- ◆近年、音声障害の治療は、音源とフィルターの相互作用を利用したチューブ発声を適応するようになった。
- ◆チューブ発声とは、内径が3mm 以下のストローを使った発声法で、 開口端を狭めて発声することによって、口腔内圧の上昇により声門上 の喉頭前庭が狭められ、いわゆる通りの良い響く声になると去れている。
- ◆本発表では、具体的なチューブ発声の練習方法とその意義について詳細に述べる。

2-10-11

2-10-11

ST のための音響学の教科書を考える

Making a textbook of acoustics for speech therapists.

〇竹内京子(順天堂大),中村健太郎(東工大),

峯松信明(東大),世木秀明(千葉工大)

- ◆ことばの障害のリハビリを行う言語聴覚士の養成校では、音響学 (聴覚心理学を含む) が必修科目である。音響学では、「音源フィ ルタ理論」が学習項目である。
- ◆「音源フィルタ理論」を学習するにあたり、声道の共鳴にあたる 「音響管の共鳴」についてまず理解する必要がある。
- ◆言語聴覚士の音響学で扱われる「音響管の共鳴」は声道だけでは ない。今回,養成校の教科書を作成するにあたり、「音響管の共鳴」 が複数回扱われ、それぞれ違った現象を表している。
- ◆本発表では、これらの科目の中でどのように扱われているのかを 紹介し、どのように教えるべきかについても考える。
- ◆音源フィルタ理論における母音の声道特性を考えるためには、音響管の共鳴の理解から何ステップも隔たりがある。この隔たりを 少しでも埋める必要がある。
- ◆養成校の音響学のように明確な目標があり、多少難しくても実際 に使える知識として必要な場合、そのための教え方、カリキュラ ムが求められている。
- ◆「音響管の共鳴」を教える際に、何をどのように扱い、どのような 流れで学ぶかについて改めて考えるのが今後の課題である。

2-10-10

2-10-10 音声障害患者を対象とした発声訓練支援システム「スマートチューブ」の評価

Evaluation of "Smart Tube," a voice training support system for patients with voice disorders

〇川村 直子(姫路獨協大), 北村 達也(甲南大), △前川 圭子(神戸市中央市民病院)

- ◆先行研究で開発した「スマートチューブ」は、安価な加速度センサ、マイコン、LED等で構成され(Fig. 1)、チューブ発声中にはっきりと自覚することが望ましいとされる口唇部振動(感覚)を LED の発光で視覚化しフィードバックする(Fig. 2)。
- ◆本研究では、このスマートチューブを音声障害患者のリハビリ臨床 場面に導入し、アンケート等を実施してその有用性を検討した。
- ◆音声障害患者によるアンケート評価から、スマートチューブの視覚化機能は、発声の正否の認知に寄与し、訓練に対する安心感や自信に作用することが示唆された。

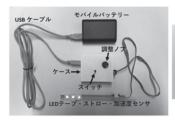




Fig.1:Smart Tube system

Fig.2:Smart Tube trial run

2-10-12

2-10-12

音源とフィルタの相互作用を 声道模型を使って示す試み

Demonstrating interactions of the source with the vocal tract using vocal-tract models

〇荒井隆行(上智大·理工)

- ◆生成面:音源とフィルタの相互作用
 - → 声道模型を用いて物理現象として示すことが可能
- ◆例として、リード式音源と声道模型
- ◆その他, チューブ発声法のための断面積可変式のチューブの開発



Figure: Tube with variable cross-sectional area for the semi-occluded vocal tract exercises.

◆知覚面:音源とフィルタの相互作用を広義に捉えることも可能

◆例1: 開鼻声 vs. 嗄声

◆例2:母音の質(異なる基本周波数と声道長の組み合わせ等)

◆その他:音声のダイナミックスについて

2-10-13 音声治療の継続性向上のための ゲームアプリの開発

Development of a game application to improve continuity of voice therapy ☆村井武人, 北村達也(甲南大), 川村直子(姫路獨協大)

研究の背景と目的

- ◆ 本システムはチューブ発声を用いた音声治療の研究である。
- ◆ 音声治療における訓練は毎日が同じ内容の繰り返しであるため、患者 の訓練に対するモチベーション低下が問題となっており、訓練を安定 して継続できる方策が求められている
- ◆ そこで、本研究ではゲーミフィケーションに注目し、患者がモチベ ーションを維持しながら日々の訓練を行えるシステムの開発を行う。

システムの構成

- ◆ 川村・北村が開発した IoT クラウドシステムにゲーミフィケーション を導入する. (Fig.1)
- ◆ Unity で作成したスロットゲームをチューブ発声と BLE 通信で連携 することで、ゲーミフィケーションの導入を実現している
- ◆ 患者はチューブ発声を行うことで、スロットゲームを遊ぶことができ るシステムとなっている



Fig.1: Overview of the proposed voice rehabilitation support system.

2 - 10 - 15

2 - 10 - 15音響管を用いた代用発声における 音源位置の変化がフォルマントに及ぼす影響

Effects of changes in source location on formants in substitutional speech using acoustic tubes

☆小野寺悦希 (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)



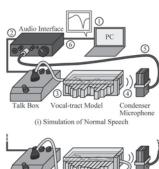
声帯を用いない代用発声法として、 音響管をくわえて音源を声道に伝達する 発声法に注目した。

目的

音響管をくわえる発声において、声道内 の音源位置が変化した際の、音韻の変化 を計測する。

原理

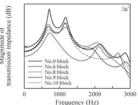
音韻の特徴は、発声の第1第2フォルマント周波数が支配的 声道を模したモデルを用意し、音源位置を変化させて録音 第1第2フォルマント周波数の変化を計測・考察した。



(ii) Simulation of Substitutional Speech

結果

モデルにおいて、音源位置(声 帯側から何番目のブロック位 置か)の変化によって母音の 第1第2フォルマント周波数 が変化した。



2-10-14

2-10-14 チューブ発声法を用いた訓練による効果 の持続性

Sustainability of effects of tube-phonation exercises ☆井上智瑛, 竹本浩典(千葉工大)

- ◆チューブ発声法は、ストローなどをくわえて発声を繰り返し行う音声 治療法の1つで、嗄声の治療や声学の訓練などに広く用いられている。
- ◆前報では、20 代日本人成人男女 6 名の実験参加者が 9 カ月間にわた りチューブ発声法の訓練を行った結果、基本周波数の最高値が有意に 上昇することを示した。
- ◆本研究では、上記の実験参加者が訓練を停止してから約1年5カ月後 の基本周波数の最高値を計測した。
- ◆その結果,基本周波数の最高値は訓練直後に比べて低下したが、依然 として訓練前より有意に高い状態であり、現在も効果が持続している ことが明らかになった。

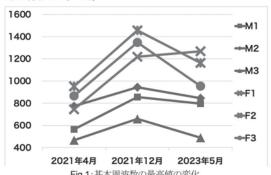


Fig.1:基本周波数の最高値の変化

2 - 10 - 16

2-10-16 鼻副鼻腔悪性腫瘍手術の術前・術後の 鼻腔形態とその音響変化の検討

Examination of morphological and acoustic changes before and after surgery for malignant tumors of the nasal and paranasal sinuses.

☆福島侑希, 竹本浩典(千葉工大),

△鴻信義, △武田鉄平, △今川記恵, △小黒亮史, △大村和宏(慈恵医大)

- ◆鼻副鼻腔の小規模な手術による音声変化は、模擬手術や音響シミュレ ションによりある程度の精度で予想できる。しかし、悪性腫瘍の手 術など、切除部位が大きい手術についてはまだ検討されていない。
- ◆本研究では、悪性腫瘍を切除する手術の術前・術後の鼻副鼻腔の形態 と音響変化を検討した。特に術後では、鼻副鼻腔形態は単純化される ものの、音響伝搬経路は複雑化する可能性があるため、音声スペクト ルと伝達関数を比較してシミュレーションの妥当性を吟味した。
- ◆手術により鼻副鼻腔の容積は最大 60.2%増加した。また、音声スペク 🖯 トルと形状から計算した伝達関数 は術前・術後とも概形はよく一致し 手術によって音響経路が複雑化し てもシミュレーションの精度に問 題がないことを確認した。そして、 術前に約1kHz 以下に見られた副鼻 腔に由来するディップ(O)は、術 後に高域(約1~2 kHz)に移動した。

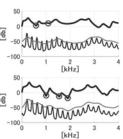


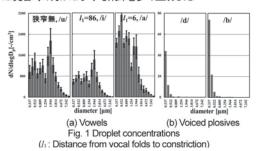
Fig.1: Transfer function (thick line), speech spectrum (thin line), and liftered spectrum (gray line) before (upper) and after (lower) surgery

2-10- リード式人工声帯と口腔モデルを用いた 17 調音様式が飛沫生成に及ぼす影響の調査

Investigation of the effects of articulation style on droplet generation using reed-type artificial vocal folds and vocal tract models

☆南智哉(豊橋技科大),吉永司(大阪大学),荒井隆行(上智大), 飯田明由,(豊橋技科大)

- ◆呼吸器系のエアロゾルの発生源として、粘液ブリッジが声帯振動により破裂して生成する喉頭エアロゾルがあると考えられているが、それが調音様式によってどう影響されるのかは明らかでない。
- ◆母音と有性子音を再現できる3音響管モデルおよびBMW-DNモデルを用いて、口腔形状と声帯振動が飛沫生成量に与える影響を調べる。
- ◆母音発生では声帯部、/b/および/d/では口腔に人工唾液を付着させ、空気を流入させた時に口腔から生成される飛沫量をパーティクルカウンターによって計測する。
- ◆3音響管モデルによる飛沫計測の結果では、声道内の狭窄が声帯に近いほど飛沫量が増加した。BMW-DN モデルによる飛沫計測の結果では、/b/発音時の方が/d/よりも飛沫を多く生成した。



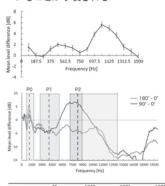
2-11-1

2-11-1 頭部伝達関数の 4kHz 以下の帯域の 前後方向知覚の手掛かり

A cue for front-back sound image localization in the head-related transfer function below 4kHz

◎中村風香(千葉工大·院), 飯田一博(千葉工大·先進工)

- ◆118 耳介の正中面の HRTF を分析し、1kHz 付近で後方は正面よりも 卓越することを示した(Fig. 1). この卓越周波数帯域は Blauert の後方 の方向決定帯域と一致した(Fig. 2).
- ◆この卓越周波数帯域に相当するピーク(PO)を新たに追加した HRTF モデルを構築した. 音像定位実験の結果, 従来よりも後方の音像定位 精度が向上し, 実測 HRTF と有意差のない精度が得られた(Table 1).
- ◆以上より、1kHz の卓越周波数帯域が後方知覚の手掛かりに寄与して いることが示唆される.



N1N2P1P2

118 ears. Bars denote the 95% significance range. Fig. 2 Relationship between

Fig. 1 Mean amplitude level

difference of the HRTFs in the rear-front direction for

Fig. 2 Relationship between the directional bands for the rear (orange), front (green), and above (blue) directions and the P0, P1, and P2 (broken lines).

5.1 2.0 Table. 1 Mean vertical angle error [deg.].

Measured 4.3

2-10-18

2-10-18

A Possible Prosthetic Device for Speech Sound Disorders

〇荒井隆行(上智大・理工)

- ◆発話に障害がある場合に対応した補綴器具の可能性を検討した.
- ◆Fig. 1 にその外観を示す.





Fig. 1: Overview of proposed device: (a) schematic illustration and (b) photograph.

- ◆口腔内に挿入するフレームはアクリル製.実験では舌にこんにゃくを 使用.
- ◆器具には、口唇より外側に下唇プレートがアルミフレーム内を上下に 動くように設置されている.
- ◆提案する器具を使用し、母音や子音、いくつかのフレーズを発話し、 その分析を行った。

2-11-2

2-11-2

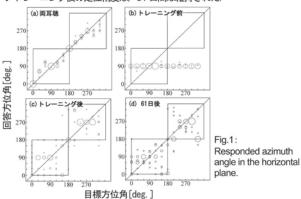
ー側性難聴者のための 水平面音源定位トレーニング

- 擬似一側聾状態でのトレーニング効果の検証 -

Horizontal sound source localization training for people with unilateral hearing loss - Validation of the effects of the training with normal-hearing listeners under simulated unilateral hearing loss -

☆稲葉春香(千葉工大·院), 飯田一博(千葉工大·先進工)

- ◆本研究では、擬似一側聾状態を模した健聴者に、HRTFとヘッドホンを用いて、水平面全周を目標方向とする音源定位トレーニングを実施し、音源定位テストでその効果を検証した。その結果、以下のことが示された。
- ◆ 両耳聴と比較して擬似一側聾状態では、平均側方角誤差、左右誤判定率、 前後誤判定率が増加した。
- ◆トレーニングにより全ての被験者で平均側方角誤差と左右誤判定率が改善した。しかし、3名中2名の前後誤判定率は改善しなかった。
- ◆ トレーニング後の定位精度は、61 日間は維持された



2-11-3

2-11-3 耳介のサイズ・硬度が軟骨伝導の伝搬 成分に及ぼす影響

Effects of auricular size and hardness on propagation components of cartilage conduction

☆田村 茜, 大塚 翔, 中川 誠司(千葉大)

- ◆骨伝導の装用性の問題を解決する方法として"軟骨伝導"が提案され、 既に補聴器やスマートホンなどに応用されている。
- ◆ 軟骨伝導では骨伝導同様に 4 つの伝搬成分が生じるが、鼓膜までに 通常の聴覚経路に入る気導成分と外耳道内放射成分が支配的となる。
- ◆耳介特性と各伝搬成分の解明を目指し、耳介硬度 (Fig. 1)、耳介サイ ズおよび頭部サイズ、耳介・頭部振動を計測し、相関関係を調べた。
- ◆耳介サイズ-耳介振動間の有意な負相関が認められ、耳介サイズが小さいほど効率よく振動することが示された。
- ◆耳介振動と各部位の硬度の関係については、耳輪/対珠で負相関 (Fig. 2)、対輪上部で正相関が確認された。
- ◆これらの結果は、耳介サイズの減少が耳介振動を増大させること、耳 介硬度は部位によって耳介振動に異なる影響を与えることを示して いる。



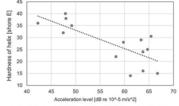


Fig. 1: Measurements locations for the hardness of pinna.

Fig. 2: A scatter plot of helix hardness vs. acceleration level at 500 Hz.

2-11-5

2-11-5 音楽聴取及びその他の生活習慣と 拡張高周波域聴力との関係

Relationship between lifestyle including music listening and hearing ability at extended high frequencies

〇倉片憲治, 原 和輝(早稲田大·人間科学)

- ◆音楽聴取及びその他の生活習慣と聴力との関係を明らかにするため に、21~25歳の男女25名を対象として、日常の生活習慣等に関する アンケート調査、拡張高周波域(8~20 kHz)を含む純音聴力測定、 及び音楽聴取音量の測定を行った。
- ◆その結果、聴力に悪い影響を及ぼすと考えられる生活習慣等(音楽聴取による音響曝露量の多さ、及び飲酒習慣)のある者は、そうでない者に比べて、拡張高周波域の聴力が悪い傾向が見られた。
- ◆一方、聴力に良い影響を及ぼすと考えられる生活習慣等(楽器演奏歴, 運動習慣,及び食事に対する高い意識)のある者は、そうでない者に 比べて、拡張高周波域の聴力は良い傾向が見られた(Fig.1)。
- ◆拡張高周波聴力測定によって,通常の聴力測定周波数(0.125~8 kHz)

における測定では明らかにできなかった, 聴力に悪影響を及ぼす生活習慣等を特定できる可能性が示唆された。

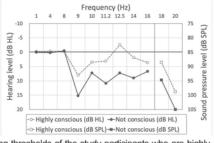


Fig.1: Median hearing thresholds of the study participants who are highly conscious/not conscious about their daily diet.

2-11-4

2-11-4 スナック類の咀嚼音の音響的特徴と 食感の関係

Relationship between acoustic characteristics and texture of chewing sounds in snacks

☆橋本卓己(神奈川工科大), 三浦貴大(産総研), 上田麻理(神奈川工科大)

- ◆これまで音と食感の関係を明らかにするために、ポテトチップス咀 嚼音の音響解析と深層学習による種類判別を行った。しかし、グミや ガムなどの食感が硬い食べ物での咀嚼音との関係性については明ら かにされていない。
- ◆そこで本稿では、8 種類のグミを用いて音響解析と食感評価アンケートを行い、ポテトチップスとの食感の違いを咀嚼音に注目して明らかにした。

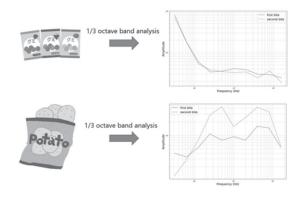


Fig. 1 Difference in chewing sound between gummies and potato chips by 1/3 octave band analysis

2-11-6

2-11-6 一側性難聴者の残響下での時間分解能の 検討一狭帯域雑音による無音検出課題を 用いて—

An investigation of temporal resolution under reverberation for those with unilateral hearing loss: Using gap detection tasks of narrow band noise ☆辻慎也, 荒井隆行(上智大•理工)

- ◆片側耳は正常だが、対耳側に難聴がある状態を一側性難聴という。
- ◆近年では良聴耳の時間分解能が両側聴力正常者と比べて、劣って いることが困難の原因の一つであると指摘されている
- ◆本稿では一側性難聴者・健聴者のモノラル受聴・両耳聴を対象に、 無響 / 残響下での狭帯域雑音

による無音検出課題を行い, 次の結果を得た (Fig. 1).

1)同じ中心周波数の狭帯域雑音に よる測定では、聴取条件間で閾値に 有意差はみられなかった。

2) 違う中心周波数の狭帯域雑音に よる測定では、両耳聴は残響下でも 無響と有意差なしの閾値であった 一方、一側性難聴者では残響による 有意な閾値の上昇がみられた

3) 一側性難聴者は残響下での聴取 のため、両耳聴より SN 比が必要で ある事と関連する可能性がある.

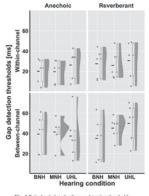


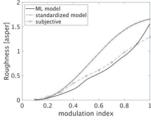
Fig. I. Raincloud plots for the gap detection thresholds including may litered data, how-whise, plots, and split-half violins for three hearing conditions; BNH (n = 5), MNH (n = 5), and UHL (n = 5). The solid and broken horizontal lines indicate the median and mean, respectively.
***p < 0.001, *p < 0.01, *p < 0.05</p>

2-11-7

2-11-7 Prediction of Psychoacoustic Roughness Model Using Machine Learning

☆Syuto Yoshida and Julián Villegas (University of Aizu)

- •We present a Machine Learning (ML) model that can predict psychoacoustic roughness of several phenomena.
- •We used Convolutional Neural Networks (CNNs) to build our roughness prediction model and trained it using previously published subjective roughness data.
- ◆Dataset creation, ML-model design and training were performed in Matlab R2022b and its toolboxes.
- ◆The figure on the right shows the predictions for an amplitude modulated tone with varying frequency modulation index.
- ◆This is one example where the ML predictions were more accurate than those computed by the ISO standardized model.



However, our model is less accurate predicting the influences of phase shifts, and has some overfitting problems.

2-P-1

2-P-1 佐藤工業技術センター 音響実験施設の音響特性

Acoustic characteristics of Acoustics Laboratory of Sato Kogyo Technology Center

◎平岡千春, 吉岡清(佐藤工業)

- ◆佐藤工業株技術センターに新設された音響実験施設の概要を紹介し、 主要な音響特性について報告する。本実験施設は、無響室と2つの残 響室、床衝撃音の測定を行う箱型実験室を主に、模型実験室や計測室 を収容する建物である。
- ◆残響室-残響室-無響室のそれぞれの室間には移動式の試料カセットを挿入して、試料の音響透過損失を測定する。
- ◆計画地は、西側に鉄道が、北側には自動車専用道路が位置しており、 振動及び騒音の影響が懸念されたため、無響室と残響室の床は、防振 浮き構造と計画するなど、計測への影響を最小化した。
- ◆残響室の拡散性については、拡散体を用いて実現する計画とした。
- ◆本報では、無響室の逆二乗則特性、残響室及び箱型実験室の残響時間、 試料カセットの音響透過損失に関して報告する。



Fig. 1 Floor plan and cross section plan

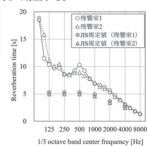


Fig. 2 Reverberation time (Reverberation chamber 1,2)

2-11-8

2-11-8 時間分解能検査の作成 一(6)振幅変調検出における難聴者の確率 密度分布の推定一

Tests of Human Auditory Temporal Resolution— (6) Estimation of probability density distribution of thresholds of hearing impaired listeners in amplitude modulation detection—

☆森田健志(九州大), 森本隆司(リオン), △岡本康秀(済生会中央病院・慶應 大), △神崎晶(慶應大), 森周司(九州大)

- ◆聴覚時間分解能の評価方法の1つである振幅変調検出の高速化を目 指し、ZESTによる測定を検討している。
- ◆ZEST では閾値の事前推定を用意する必要がある。事前推定が実際の 閾値分布を正確に表すことができていれば、測定効率がより上昇する。
- ◆難聴者 24 名に対して測定を行い、得られたデータからカーネル密度 推定法によって導出した確率密度分布に正規分布と modified hyperbolic secant でフィッティングを行った。
- ◆ どちらの分布もよく当てはまったが、8,16Hz では正規分布、64,128 Hz では modified hyperbolic secant の方が決定係数の値が高かった。
- ◆以前の研究で健聴者群の閾値分布を推定したため、本研究では難聴者 群の閾値分布を推定し、今後は両者を組み合わせた分布でコンピュー ターシミュレーションによる評価を実施する。

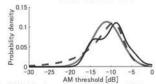


Fig.1:Threshold distribution of hearing impaired listeners at modulation frequency of 128 Hz (black) and fitted probability density distribution models(red: normal distribution, blue: modified hyperbolic secant)

2-P-2

2-P-2 小田原三の丸ホールの音響設計

Acoustical Design and Characteristics of Odawara Sannomaru Hall 〇鈴木航輔, 箱崎文子, 福地智子(永田音響設計)

- ◆小田原三の丸ホールは神奈川県小田原市に建設された公共施設で、 2021 年9 月に開館した。大ホール(1,105 席)、小ホール(296 席) の他、ダンスや音楽の練習スペース、展示スペースを擁し、多種多様 に利用されている。本稿では、本施設の音響計画と音響特性の概要を 報告する。
- ◆大ホールはフライタワーを備えた多目的ホールである。生音のオーケストラやオペラ、拡声設備を用いたポピュラー音楽、演劇、講演など様々なジャンルの公演に利用される場として計画された。客席は主階席とバルコニー席の2層構成である。
 - ▶ 主階席の壁際にブロック席を設け、中央付近にも初期反射音が届 きやすい室形状とした。
 - ▶ 客席勾配が大きく、視線が良好で聴覚的にも舞台と客席の親密感が高い空間である。
- ◆小ホールはリハーサル利用や、展示室としての利用も可能な場として 計画された多目的ホールである。音楽や演劇での利用に加え、移動観 覧席を収納することで大ホール舞台のアクティングエリアと同じ広 さの平土間空間とすることができる。
- ◆和太鼓の市民団体に協力いただき、大ホールと小ホールで測定した和 太鼓演奏時の発生音量も報告する。

2 - P - 3

2-P-3 東急歌舞伎町タワーの音響設計 -THEATER MILANO-Za·109シネマズ プレミアム新宿·Zepp Shinjuku (TOKYO)-

Acoustical design and characteristics of TOKYU KABUKICHO TOWER

〇箱崎文子, 服部暢彦, 小野朗(永田音響設計)

- ◆2023 年 4 月,東京都新宿区歌舞伎町にエンターテインメント施設とホテルからなる超高層ビル「東急歌舞伎町タワー」がオープンした。
- ◆本建物は Fig.1 に示すように建物の中層階に 900 席規模の劇場 「THEATER MILANO-Za」,映画館「109シネマズプレミアム新宿」, 地下にライブホール「Zepp Shinjuku (TOKYO)」が配置されている。 本稿では、劇場・映画館・ライブホールの遮音計画と劇場の室内音響 計画及び音響性能・特性について述べる。
- ◆上下に積層する劇場一映画館間では特に高い遮音性能が求められ、両室間に2 重スラブを計画すると共に、劇場と映画館(8室)の全室を防振遮音構造とした。 (室間遮音性能:65~71 dB/63 Hz)
- ◆演劇やミュージカルを主目的とする劇場の室内音響計画では、直接音の明瞭性を高める初期反射音(直接音からの遅れ時間:50 ms 以内)が得られるように客席天井反射板を設けた。一方で、遅い反射音を生じさせる上部の側壁や後壁を吸音仕上げとした。

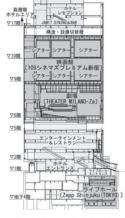


Fig.1 Section of TOKYU KABUKICHO TOWER

2 - P - 5

2-P-5 なでしこ芸術文化センターの音響設計

Acoustical design and characteristics of Nadeshiko Art and Culture Center

〇和田竜一, 小野朗(永田音響設計)

2022 年 11 月、兵庫県の神戸市営地下鉄西神中央駅前に開館した「なでしこ芸術文化センター」は、西神中央ホール(以下、ホール)と図書館からなる複合文化施設である。

- 遮音計画:ホール周囲には、リハーサルや楽器練習などに利用されるスタジオが4室配置され、いずれの室もホールとの距離が近い。各室間の同時使用を出来るだけ可能にするため、各スタジオに防振遮音構造を採用した。
- ホールの室内音響計画:ホールは、エンドステージ型を基本とした ワンスロープのホールで、クラシックからポップスまでの音楽利用を 主目的としている。

[天井] 十分な室容積を確保するため、天井を音響的に透過なルーバー で構成し、それより上部の空間も有効に利用した。またルーバー上の 傾いた屋根スラブからの初期反射音の偏在を解消するため、スラブ下

の一部に角度と高さを調整した反射 面を別途設けて、初期反射音の空間 的・時間的なバランスを整えた。

[壁面] 低域まで充実した反射音と、 音場の適度な拡散性を得るため、主 要壁面はコンクリートによる凹凸壁 面で構成し、高音域の反射音の散乱 を意図してランダムリブを設けた。



Fig.1 Section of Seishin Chuo Hall

2-P-4

2-P-4 NHK ホール天井耐震化工事の音響設計

Acoustic design for NHK hall ceiling earthquake resistance construction 〇内藤洋一(NHK), 荒田耕平, 北村浩一(NHK テクノロジーズ)

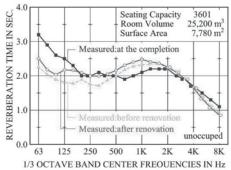
◆NHKホールは2021年3月から2022年6月まで休館し、新しい耐震 基準に準拠した天井脱落防止の改修工事を行った。ホール客席内での 改修工事概要と音響設計について報告する。





Fig.1:before renovation

Fig.2:after renovation



1/3 OCTAVE BAND CENTER FREOUENCIES IN Hz Fig.3:reverberation time (Concert mode)

2-P-6

2-P-6 野口遵記念館の音響設計 一音楽ホールにおける拡散形状の 組み合わせによる反射音の最適化—

Acoustic Design of the NOGUCHI Shitagau Memorial Hall – Optimization of reflected sound by combining diffusion shapes in concert hall -

〇日根野翔太, 宮崎秀生(ヤマハ),豊田政弘(関西大学),岸永伸二

- ◆本施設は、音楽ホールやリハーサル室、練習室などを備える公共文化施設である。
- ◆音楽ホールの内装については、地域文化をモチーフにしたデザインが 各所に散りばめられており、延岡の文化を強く感じられる意匠を有し ている。
- ◆音響設計においては、この特徴的な意匠の拡散性に着目し、意匠性と 音響性能が両立した拡散形状の検討を行った。
- ◆上記により、どの席においても反射音がバランスよく届く、音楽演奏 に適した音場を目指して各種検討を行った。
- ◆拡散形状については、数値解析及び残響室吸音率測定を実施し、効果 や影響について詳細検討を行った。

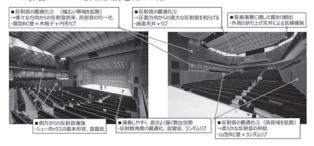


Fig.1 Acoustical Concept

2-P-7

2-P-7 FMBEM 入射指向性解析に基づく 拡散性指標のオクターブバンド分析

Octave band analysis of sound-field diffusion indices based on FMBEM incidence directivity analysis

☆萩原諒、佐久間哲哉(東大・エ)、安田洋介(神奈川大・建築)、桝本貴之(サイパネット)

- ◆ 既報では室内の定常音場を対象とし、FMBEM 入射指向性解析に基づいて単一周波数における拡散性指標を計算した上で傾向を把握した。
- ◆本報では、複数の周波数における入射指向性のエネルギーを方向毎に 平均することでオクターブバンドの入射指向性を定義し、さらにその 値を拡散性指標の計算に適用した結果について考察を行った。
- ◆まずオクターブバンドの入射指向性について、反射性の音場ではディップが少なくなだらかな形になる一方で、吸音性の音場では単一周波数とオクターブバンドで同方向にピークが生じ、同様の形が現れた。
- ◆オクターブバンドの入射指向性を用いて計算した拡散性指標は、反射性の音場で単一周波数の指標よりも高い値となり、吸音性の音場では 単一周波数の指標と同程度の値となった。

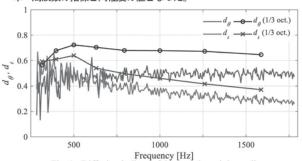


Fig. 1 : Diffusion indices of the sound receiving cell.

2-P-9

2-P-9 有孔板を用いた四重壁の共鳴透過

Mass-air-mass resonance for quadruple-leaf partition using perforated plates

〇杉江聡, 新田龍馬(小林理研)

- ◆両替所等の防犯用の仕切り板の遮音性能を低下させる目的に、 Fig.11に示すような、中央二層が無孔板、その両側に有孔板を 用いた断面左右対称の四重壁における、低域共鳴透過周波数の 設計方法を示す。

$$\begin{split} f_{r,quad,2} &= \sqrt{\frac{1+b}{2ab}} f_{r,dbd} \\ f_{r,quad,n} &= \sqrt{\frac{1+b+2ab \mp \sqrt{(1+b+2ab)^2-8ab}}{4ab}} f_{r,dbb}, \ \ n=1,3 \end{split}$$

- igaplus aとbを適切に調整すれば、 $f_{r,quad,i}$ の分散・集中が可能。
- ◆有孔板の実効面密度が容易に調整可能な点を活かして、 $f_{r,quad,2}$ =約1000 比として共鳴周波数を含む帯域で極力低下するように設計した遮音性能の実測結果をFig. 2に示す。

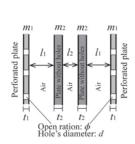
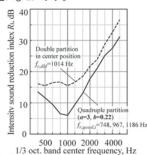


Fig.1 Symmetric quadruple leaf partition using perforated plates.



1/3 oct. band center frequency, Hz
Fig.2 Measured results for quadruple-leaf partition using perforated plate.

2-P-8

2-P-8 小規模ホールにおける音響状態の弁別に 関する研究

A study on discrimination of acoustic condition in small hall

☆小林優暉(舞鶴高専), 徳永泰伸(舞鶴高専), 寺島貴根(三重大)

- ◆本研究では、ホール客席部で採取されたインパルス応答とソロ演奏畳 み込み音源を被験者が受聴した際の主観的な差異知覚の傾向を明ら かにすることを目的とし、30名を対象に被験者実験を行った。
- ◆被験者は、ヘッドホン (MDR-CD900ST, SONY) からインパルス応答の対 (IR) と畳み込み音源の対 (Vn) を範囲ごとに受聴し、その対に対して、全く同じ [0]・少し違う [1]・おおむね違う [2]・かなり違う [3]・全く違う [4]。の5段階で回答を行った。
- ◆被験者の受聴レベルは各範囲の基準位置(L2, M2, S2)においてIRではL_{Anex}が70 dB, VnではL_{Aeq}が60 dBとなるように設定した。
- ◆ホール客席部の全体を対象としたL範囲では、インパルス応答と楽曲ともに被験者の差異の知覚には明確な傾向が見られた。M範囲では、判断の傾向の規則性は残るものの、座席が隣接するS範囲では布置の規則性は無くなり、差異の知覚判断はあいまいになる可能性がある。

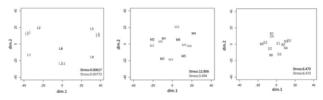


Fig.1: Results of MDS.

2-P-10

2-P-10 原始根に基づく円形リフレクタの 試作と評価

Prototype and evaluation of circular acoustic reflector based on primitive roots

☆齋藤優真,△牧野潔夫,高橋義典(工学院大)

- ◆数論の音響分野への応用として,原始根に基づく位相回折格子による リフレクタが提案されている.
- ◆本稿では対象となる波長が異なる2種類の円形リフレクタを試作し、 不整形模型残響室を用いて scattering coefficient を評価した.

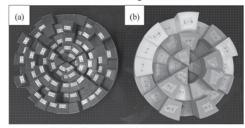


Fig. 1 Prototyped circulars reflector (a) for λ =136 mm and (b) for λ =226.7 mm

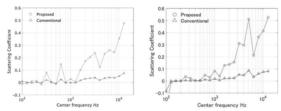


Fig. 2 Scattering coefficient (a) and (b)

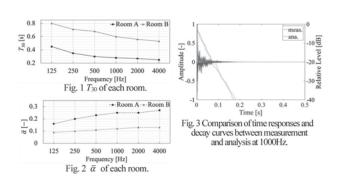
2-P-11

2-P-11 三味線練習室の音環境に関する実測 および数値解析的研究

Measured and numerical analysis on the sound environment of Shamisen practice room

☆吉田新(大分大院・工学研), 富来礼次, 岡本則子, 大鶴徹(大分大・理工), △鮎川晴季(大分大院・工学研)

- ◆三味線の練習室として適切な音環境の構築を目指し、実測および数値 解析による検討を行った。
- ◆実測の結果、演奏の際に不快に感じる室の T₂₀は 500Hz で 約 0.7s (Fig.1) 、 Eyring の残響式より算出した室の平均吸音率は 0.1 程度 (Fig.2) 、C₆₀ は5-10 dB 程度であった。
- ◆実測された室の残響時間の平均値に基づき境界の吸音率を設定した 数値解析の結果(ana.)は各受音点で時間応答波形と減衰曲線が実測値 (meas.)とよく対応した (Fig. 3)。



2-P-13

2-P-13 偏光高速度干渉計を用いたインパクト レンチ使用時に発生する音の可視化計測

Measurement and visualization of sound generated by impact wrench using high-speed polarization interferometer

☆山崎匠, 坂之上大喜, 及川靖広(早大理工), 木谷正一(小松製作所)

<u>目的</u>: インパクトレンチ使用時に発生する打撃音の特性を把握するため、偏光高速度干渉計による可視化計測を行う。

実験:

- 4000Hz と 9000Hz あたりに特徴的なピークがみられ、それらを中 心周波数とする音場の可視化計測を行う。
- 2. 制振材を作業箇所に貼付し、音場の変化を捉えることで振動抑制 の重要性を確認する。

結果: ネジ締め部分から振動が伝わり, 金属板の端から大きな放射音が発生していた。制振材による高周波数帯域の低減効果を確認した。

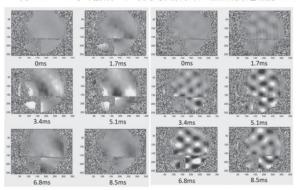


Fig.1 Visualization results for center frequencies of 4000 and 9000Hz

2-P-12

2-P-12 マイクロホン信号とその位置情報を用いた 深層学習に基づく壁面の吸音率推定

Deep learning based estimation of sound absorption coefficients using microphone signals and its positions

☆新井大斗, 佐藤元, 池田雄介(東京電機大)

◆背层

- ➤ 実環境に即した音場シミュレーションを行なうためには、境界条件の把握が必要
- ▶ 近年、計測の効率化を目的とした深層学習を用いた吸音率の推定 が提案

◆提案手法

四辺形の自由な部屋形状に対して、自由なマイクロホンアレイ位置で計測れた複素音圧を用いて、吸音率を正解とする教師あり学習を行い、吸音率を推定

◆実験

相対誤差の中央値において10%程度で吸音率の推定が可能である事がわかった。

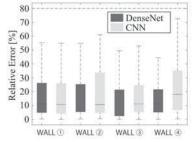


Fig.1: Relative error in estimated sound absorption coefficient

2-P-14

ミンミンゼミの周波数と体長の相関分析

A correlation analysis between frequency of climax and body length of each robust cicada

☆御園玲央, 武藤憲司(芝浦工大)

- ◆本研究は、ミンミンゼミの鳴き声で最も音圧レベルが大きい高潮音の 1周期「ミン」のうち「ミ」の部分の周波数特性を分析した.
- ◆これまでのセミの鳴き声は屋外録音のデータを用いて分析されてきたため暗騒音の影響が大きいという課題がある。本研究では自由音場で1匹ごとのミンミンゼミの鳴き声を録音し、周波数特性の算出とピーク周波数の統計分析をした。また、体長と周波数の相関分析をした。
- ◆ミンミンゼミ 75 匹の高潮音の周波数特性を統計分析した. ピーク周波数の平均値は4658 Hz,標準偏差は268 Hz であった. 体長平均値は 34.0 mm で,体長と鳴き声のピーク周波数の間で相関係数R = -0.28の相関を持っていた.

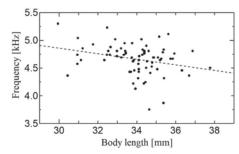


Fig.1: Scatterplot of body length and frequency of robust cicadas (R = -0.28, n = 75)

2-P-15

2-P-15 低周波音モニタリングシステムの検討

- 風雑音判別と上空音源を対象とした方向推定 - Examination of low frequency sound monitoring system

〇土肥哲也, 岩永景一郎, 小林知尋(小林理研), 中山 紬, 中島康貴(リオン)

- ◆低周波音を屋外でモニタリングする場合、計測対象音に混入する風雑音を判別して除外する必要がある。
- ◆Fig. 1 に示すように 3 つのマイクロホンを同じ風防内の地表面高さに 配置したモニタリング装置を試作した。
- ◆風雑音の判別性能をスピーカ試験と自然風を用いたフィールド試験 で検討した。
- ◆上空音源を対象とした場合の音の到来方向の推定をシミュレーションにより検討した。本装置を用いて仰角の違いを判別できる可能性が示唆された。

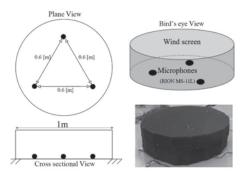


Fig.1: Improved low frequency sound monitoring device.

2-P-18

2-P-18 見通し領域と見通し外領域における 水中音響通信路の特性計測

Measurement of Underwater Acoustic Communication Channel Characteristics in Line-of-Sight and Non-Line-of-Sight.

☆脇 拓哉, 原 優人(筑波大・院・シス情工), 海老原格, 若槻尚斗, 水谷孝一(筑波大・シス情)

◆背景・目的

水中音響通信は、様々な海洋開発を支えるための技術として着目されている。安定した水中音響通信を実現するために、様々な信号変復調技術が検討されてきた。しかし、これらの多くは見通し内環境を想定したものであり、見通し外環境を想定したものは多くない。そこで本研究では、見通し内環境と見通し外環境における通信路の特性をFig. 1 に示す環境で調査した結果について報告する。

◆結果

実験の結果、Fig. 2 に示すように LoS 環境と NLoS 環境におけるインパルス応答には明確な違いが存在することが確認できた。これらの結果より、反射や回折による信号の減衰をふまえた信号強度の制御が必要であることが分かった。

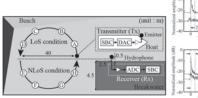


Fig. 1 Experimental environment.

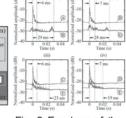


Fig. 2 Envelope of the impulse response.

2-P-17

2-P-17

SPR 型超音波センサにおけるセンシング径と 感度の検討

Experimental study on sensing diameter and frequency characteristics of SPR ultrasonic sensors

☆出竿康太, 仲辻衆登, 市橋隼人, 松川真美(同志社大)

- ◆近年、表面プラズモン共鳴(SPR)は様々な分野への応用が期待されて おり、超音波検出もその一つである。
- ◆従来のSPRセンサは後述するようにセンシング形状が楕円形かつ大き な領域となり、検出対象の音場に問題がある。
- ◆本研究では、センシング領域を真円形状にした SPR センサを製作し、 センシング径と周波数感度の検討をした。
- ◆センシング領域を真円形状にすることで、高周波での周波数応答が改善されたが、プリズム内に侵入した音波による光の屈折等により、観測波形が影響を受ける可能性がある。

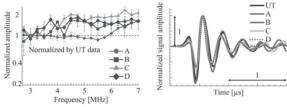


Fig. 1 Observed ultrasonic frequency spectra. Fig. 2 Observed ultrasonic waveforms.

2-P-19

2-P-19

ウマ脚部骨の超音波評価の検討

Ultrasonic evaluation of a horse leg.

☆津幡大聖 1、△北嶋昇太 1、児玉周大 1、 △三田宇宙 2、△田村周久 2、△千葉恒 3、松川真美 1 (1. 同志社大, 2. JRA 総研, 3. 長崎大)

- ◆安価で安全性、可搬性に優れ、骨の強度と関連する弾性的性質 を定量的に測定できる超音波法の競走馬骨診断への適用が期待 される。
- ◆馬のトレーニングなどで生じるソエ評価の適用可能性を<u>弾性</u> FDTD 法による2次元シミュレーションを用いて検討した。
- ◆競走馬の第三中足骨の CT 画像から、骨軸方向および円周方向 断面の骨ディジタルモデルを作成した。
- ◆骨表面の骨軸方向および円周方向に送受波器を設置し、 ソエ表面の凹凸が、観測される音波の最大振幅値に及ぼす影響を評価した(Fig. 1(a))。
- ◆ソエ表面の凹凸により、見かけの照射角度が変わると音波の散乱が 発生して、最大振幅値が変化することが分かり(Fig. 1(b)、Fig. 1 (c))、 簡易に**ソエを検出できる可能性**がある。

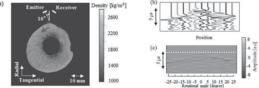


Fig.1(a) Simulation conditions in the axial-radial cross-section. (b) Observed waveforms in the tangential-radial cross-section simulation, and (c) B-scan image

2-P-20

凹型超音波ゲルレンズの径方向形状制御

- ◆ 超音波振動を用いることにより、機械的稼働部を必要としない凹凸両 用可変焦点ゲルレンズを開発した。
- ◆手ぶれ補正機構の小型・薄型化を実現するため、凹型型超音波ゲルレンズの径方向形状制御に関する検討を行なった.
- ◆レンズはガラス円板、4つの超音波振動子、透明シリコーンゲルフィルムで構成される。
- ◆共振周波数近辺の周波数において、4つの振動子の入力電圧比によって振動モードを保ったまま振動分布を径方向に制御できた(Fig. 1(a)).
- ◆振動分布と同様に、ゲルの変形形状も径方向に制御することが可能であった。これらの結果より、手振れ補正機能への応用が期待される。 (Fig. 1(b))。
- ◆今後レンズ像を通して手ぶれ補正機能に関する光学的実験を行う予 定である.

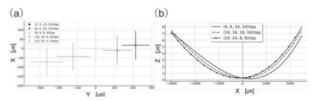


Fig.1:(a) Distributions of vibrational displacement of the lens at 67.0 kHz. (b)Relationship between voltage ratio and the center position of the gel.

2-P-22

2-P-22 皮質骨表面の微細構造による光音響特 性の変化

Changes in photoacoustic signal due to surface microstructure of cortical bone

☆黒岩丈琉, △服部大志, 松川真美 (同志社大)

- ◆糖尿病の合併症の一つとして、骨折リスクが高まる事例がある。この 要因としてコラーゲン分子間の構造を脆くする終末糖化産物(AGEs) が指摘されている。したがって骨中のコラーゲンの走査が重要となり、 光音響法に注目した。
- ◆ 本研究では、皮質骨、鏡面研磨したアルミニウム試料、熱硬化性フェノール樹脂試料を用いて光音響波を観測した。
- ◆ アルミニウム表面で発生した波の peak-peak 値はほぼ一定であったが、皮質骨表面では peak-peak 値が大きく増加した。 骨表面構造の不均一性の影響が考えられる。
- ◆小さい傷をもつ樹脂表面の光音響波形を計測した. 傷付近で振幅値が 増大し. 微小な孔などの表面形状により波形が異なることを確認した.

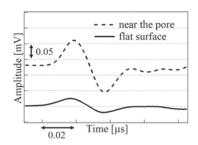


Fig.1: Photoacoustic signal changes due to surface microstructure of bone.

2-P-21

2-P- ポータブルオーディオプレーヤーを用いた 21 小型音響浮揚デバイスの実装

Implementation of Portable Audio Player-Based Small Acoustic Levitation Device

☆横山裕正、大久保電(東京都立大院・システムデザイン)

- ◆6個の超音波トランスデューサで半球状アレイの小型音響浮揚デバイ スを製作する。
- ◆ 浮揚物体の捕捉位置を変動させられるものの中では、省電力かつ小規模な構成となっている。
- ◆ポータブルオーディオプレーヤーを利用した小型音響浮揚デバイス の制御方法と内部回路について述べる。
- ◆作製した超音波信号音源を再生するだけの簡単操作で、音響浮揚を手軽に体験できるデバイスを実現している。
- ◆音源に音楽信号を加えることで、同時に音楽再生も可能である。

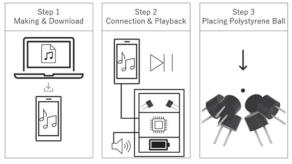


Fig.1: Portable Audio Player-Based Small Acoustic Levitation Device Configuration

2-P-23

2-P- Attempt to enhance underwater acousticstreaming using disks with acoustic cavity

☆Yimeng Wang¹, Kohei Aono², Manabu Aoyagi¹ (¹Muroran Inst. Tech.; ²SEIDENSHA ELECTRONICS CO.,LTD)

- Underwater acoustic streaming generated by the vibration source and disks with a cavity at 28 kHz in longitudinal vibration was observed by particle image velocimetry (PIV) experiments, as shown in Fig. 1.
- Vibration source consists of a bolt-clamped Langevin transducer (BLT) and a duralumin horn. The amplitude of the bottom vibrating surface in water was measured, and the relationship with the top surface was determined.
- ◆Disks with a cavity were made of transparent acrylic, and results are shown in Fig. 2. Tracers flew from the gap, passed through the cavity, and then jetted out. This flow distribution was very different from that in air, as shown in Fig. 3. Additionally, the tracer velocity in the gap was much higher than jet velocity.

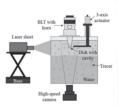


Fig. 1 Experimental setup for PIV experiments.

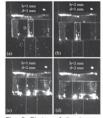


Fig. 2 Photos of the tracers around the disk with a cavity.



Fig. 3 Acoustic streaming distribution in water.

2-P-24 低周波超音波による経皮薬物導入のため の小型超音波トランスデューサ

A compact ultrasound transducer for transdermal drug delivery by low-frequency ultrasound

☆山本真也, △冨岡圭太, 杉田直広, △進士忠彦(東工大)

皮膚表面上で超音波キャビテーションを発生することで、経皮薬物 導入を可能とするウェアラブル超音波デバイスの開発を目指す。半径 方向に振動する円盤型圧電素子と椀状の音響振動子を接合すること で直径 28 mm. 厚さ 10 mm の小形でありながら、キャビテーション の発生に有効な 100 kHz 以下の低周波 (80 kHz) で駆動する超音波 トランスデューサを開発した。トランスデューサの概略を図1に示す。 光学的観察によって、皮膚表面に位置する音響振動子開口部において、 従来の振動子よりも広範囲にキャビテーション気泡が分布すること を確認した。また、疎水性の人工皮膚に超音波を照射し、水溶性染料 (アルラレッド)を溶かした染色液で染色する実験を行った。未処理 の人工皮膚は染色されないが、キャビテーションの効果によって改質 された領域は染色される。試作トランスデューサで音波を照射した人 工皮膚は、従来の振動子で音波照射した場合よりも広範囲が染色され t=.



Fig. 1 Outline of proposal transducer. (a) Concept of transducer and (b) Picture of fabricated transducer

2-P-26

2-P-26

L1 ノルム最適化による 超音波断層像の高空間分解能化

Improvement of spatial resolution of medical ultrasound images by L1 norm optimization

☆青嶋 亮人 (富山大・工), 大村 眞朗, 長岡 亮, 長谷川 英之 (富山大)

- ◆超音波画像の空間分解能が向上することによる生体組織での境界の 鮮明化を目的として、ビームフォーミング後の RF 信号における L1 ノルム最適化を用いたデコンボリューションに関する検討を行った.
- ◆本手法の性能評価として、半値幅、コントラスト、CNR 比較のための 基礎実験、生体組織の鮮明化を調べるための in vivo 実験の2つの実 験を行った
- ◆基礎実験では結果として、コントラストの変化はほどんどなく CNR が低下してしまうが、半値幅の減少が確認された. (Table 1)
- ◆in vivo 実験では包絡線の比較により、生体組織での境界の鮮明化が 確認された. (Fig. 1)

Table 1 Performance evaluation by phantom experiments

	DAS	Proposed
FWHM (range	0.20	0.14
direction) [mm]		
Contrast	0.89	0.91
CNR [dB]	6.97	2.04

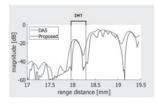


Fig. 1 Envelope of ultrasound RF signal in carotid artery (the IMT indicated in the figure is a rough guideline)

2-P-25

2-P-25 模倣学習モデルを使用した

エコーロケーションコウモリの飛行経路の予測

Prediction of Echolocation Bat Flight Path by Imitation Learning Model ☆源田祥子!,手嶋優風2,△小原大知1,△青木耀大1,△藤井慶輔3,飛龍志津子!

(1:同志社大, 2:海洋研究開発機構, 3:名古屋大)

- ◆コウモリには、エコーロケーションと飛行を高度に協調した行動ルー ルを持つことが予想される. 先行研究ではコウモリが獲物に向かう飛 行経路に対して数理モデルを構築し、その行動ルールを読み取る試み も行われた. しかし. 課題や環境との相互作用を加味した複雑なモデ ルを手動で作成するには限界がある.
- ◆本研究では、コウモリの障害物回避行動に着目し、模倣学習によって 飛行経路の予測モデルを開発した. シミュレーションで予測した飛行 経路と実際のコウモリの飛行経路を比較し、そのモデルの有用性につ いて検討した。
- ◆シミュレーションの結果、未知の環境であっても限定された情報から 飛行経路を予測することができ、モデルの汎用性と工学応用への可能 性が確認された.

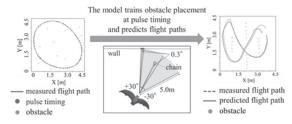


Fig. 1 Overview of this study.

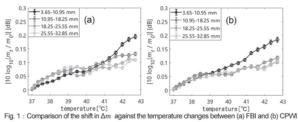
2-P-27

2-P-27 温熱療法の温度解析における 超音波ビーム形成法の最適化

Optimization of ultrasound beam formation method for temperature analysis in hyperthermia therapy

☆中村航大, 大村眞朗(富山大·工), 竹内道雄(立山科学), 長岡亮. 長谷川英之(富山大·工)

- ◆コンパウンド平面波イメージング (CPWI) での送受信音場特性およ び仲上分布を用いた振幅包絡の統計的性質を数値シミュレーション によって評価するとともに集束ビーム(FBI)との比較を行った.
- ◆192 素子, 7.5 MHz のリニアアレイプローブを再現した FBI および CPWI を行った、受信音場による影響を調べるため、各送信手法で受 信のビームフォーミングの F 値を変えてシミュレーションを行った.
- ◆散乱体数を 20 点/mm³でランダム配置し、37.0 から 42.7 °Cの温度変 化における散乱体分布変化を有限要素法により計算し、対応する超音 波データの生成および仲上形状パラメータmの変化量を算出した
- ◆F 値 1 におけるmの変化量の深度依存性を Fig. 1 に示す. CPWI は FBIに比べ、深さによって値の変化が小さく、温度上昇に伴い安定し て増加することが確認できた



2-P-28

医用超音波画像の

制約付き最小二乗法による高分解能化

High resolution medical ultrasound imaging using constrained least squares method

☆土居佳生, 大村眞朗, 長岡亮, 長谷川英之(富山大·工)

- ◆一般に困難とされている測定信号中のノイズ成分の特定を省略できる制約付き最小二乗法(constraint least square method)でのフィルタ 導出による高分解能手法による性能の検討を行った.
- ◆本研究では制約付き最小二乗法の算出に用いられる受信信号の共分 散行列である R の深さ方向の平均回数 K に注目し、K の値による制 約付き最小二乗法の性能への影響について検証する.
- ◆ファントム実験での性能検証を基に平均回数 K の最適化し、DAS の 画像を比較したものが Fig. 1 で、血管後壁部分を拡大した図から、本 手法による画像では血管内腔と後壁の境界が明瞭に描出された。

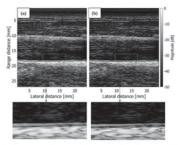


Fig. 1:B-mode images of a carotid artery of a 46-year-old healthy male obtained by (a) DAS beamforming and (b) proposed method with K value of 9.

2-P-33

Numerical analysis of transient response of a bubble cloud to ultrasound irradiation

☆An Junseok, 杉田直広, △進士忠彦(東工大)

超解像度超音波イメージングは、超音波造影剤である微小気泡の位置検出によって微細血管を描出する技術である。しかし、造影剤の濃度に制限があるため、気泡の検出確率が低く、音響データの取得時間が長いという問題がある。複数の気泡が存在する場合、気泡間の相互作用によって、気泡のエコー強度が低下することが知られており、この気泡クラウド内の干渉効果が検出確率低下の要因の一つと考えられる。本研究では、照射する超音波ビームに空間分布を与えることにより、従来の音波照射を用いた場合とは異なる気泡クラウドの振動モードを励起する手法の提案と原理検証を行う。音響イメージングシミュレーションにより、パターニングした超音波ビームに対する個々の気泡のエコー強度を確認した。本研究の提案の音波照射手法は、気泡検出率の向上に有効であることを確認した。

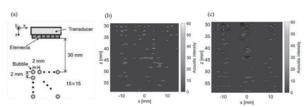


Fig.1 (a) Schematic representation of ultrasound imaging of a bubble cloud. Simulated acoustic images for (b) a plane wave and (c) a patterned plane ultrasound beam.

2-P-29

2-P-29 プライベート空間のための FM コウモリ型見守りセンサの開発

Development of a monitoring sensor system modeled on FM-bats for private spaces

○藪謙一郎(東大), 伊福部達(東大)

- ◆ 高齢者人口の増加に伴い、転倒や体調不良を早期に発見し、対応を促 すための見守りセンサシステムの必要性が高まっている。
- ◆周囲の物体の動きや状況を非接触で比較的広範囲に把握できる手段 として、FM コウモリをモデルとした超音波センサを開発している。
- ◆より扱いやすく実際の環境に設置できるセンサシステムの構築を目 的として、USB 接続のアンプ内蔵型送受信機、超音波送受信機およ び汎用パソコンのみで動作可能なシステムを開発した。
- ◆実験により、下降 FM 超音波の送受信によりシャワーの水のノイズ等がある環境でも、アクリルパイプを検出できることが示された。

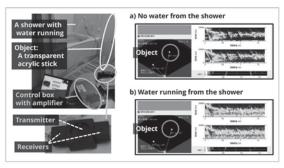


Fig.1: An experiment of FM bat-type ultrasonic sensor in a bathroom

2-P-34

2-P-34 熱音響発電システムの実用化に向けた研究 - 直管型段差熱音響発電システムの基礎検討*-

Study for Practical Application of Thermoacoustic Power Generation System
-Basic Study of Straight Tube Type Stepped Thermoacoustic Power
Generation System*-

〇坂本眞一, △七里晃弥(滋賀県立大)

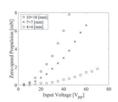
- ◆熱音響現象を応用したシステムの実用化に向けた研究を継続している。熱音響現象これまでにない、魅力的な振る舞いをする。これらの 特長を活かすことによって、これまでにない新しいシステムの実現が 可能となる
- ◆稼動部を必要としないことから長寿命でメンテナンスフリーなシステムである。また、外燃機関であるため太陽熱や工場廃熱などの未利 用熱を駆動熱源として用いることができるシステムなど、魅力の多い システムとなる可能性が高い。
- ◆電気エネルギーの利用に向けた熱音響システムの研究について着目 し、継続している。
- ◆単一内径の直管型システムではこれまでの検討条件においては発振 が難しい状況であった。
- ◆直管型段差熱音響発電システムを採用し、検討を進めた
- ◆非常に少ない電力ではあるが発電に成功した.
- ◆熱エネルギーを音エネルギーに変換し、その音エネルギーを電気エネルギーに変換することに成功した.
- ◆今後は、システム内の音場や諸条件について今後は検討を継続していく予定である。

2-P-35 LN 厚み振動子による液中推進システムの小型化の検討

Miniature acoustic propulsion system with 36° Y-cut lithium niobate thickness-vibration-mode transducer

◎孔徳卿、平田拓己(室蘭工大·院)、黒澤実(東工大)、青柳学(室蘭工大院)

- ◆36° Y-cut ニオブ酸リチウム (LiNbO3, LN) 基板 (厚み: 0.35 mm) を 用いた 10×10 mm, 7×7 mm, 4×4 mm の正方板厚み振動子を作製し、 音響推進システムを検討する。
- ◆Fig. 1 に示すように、ゼロ速度推進力が増えていき、ゼロ速度推進力 は入力電力に比例していることが分かった。駆動電圧は 70 Vpp の場 合、4×4×0.35 mm の小型化のゼロ速度推進力は 1.8 mN を測定した。 振動面の実効面積が小さいため、同じ入力電圧の時にゼロ速度推進力 が小さいである。
- ◆Fig. 2 を示すように、入力電力の増加に伴い、単位面積当たりの推力 も増加していることが分かった。また、振動子が小さいほど単位面積 当たりの推力は大きくなった。入力電力は3 W の場合、4×4×0.35 mm の小型化の音響推進システムの単位面積あたりのゼロ速度推進力は 最高 11.8 mN/cm² が得られた。



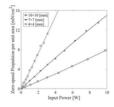


Fig. 1 Zero-speed propulsion with input voltage

Fig. 2 Zero-speed propulsion per unit area with input power

2-P-37

2-P-37 キュベット内に導入したレーザ誘起応力波 の圧力分布

Pressure distribution of laser induced stress wave introduced in a cuvette 〇宮崎晃汰(金沢工大院), 會澤康治(金沢工大院)

- ◆ レーザ誘起応力波のキュベット内部の圧力の分布を調べる
- ◆キュベット内部の圧力分布を明確にし、酵母菌増殖実験において酵母 菌にどの程度の圧力が照射されているのかを調べる。
- ◆キュベットの底に近い酵母菌とキュベットの上方に浮遊している酵母菌では10 MPa 程度差があることが分かった。
- ◆浮遊した酵母菌に増殖促進効果のあるレーザ誘起応力波を照射する際には何発か照射する必要があるのではないかと思われる。

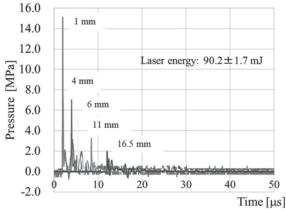


Fig.1: Typical waveforms at several hydrophone position

2-P-36

2-P-36 中間電極挿入構造 ScAIN・YbAIN 薄膜/ 高音速基板上を伝搬する RSAW 特性解析

Analysis of Sezawa-mode RSAWs on ScAIN or YbAIN films/high velocity substrates with intermediate electrode layers

☆福永慶, 鈴木雅視, 垣尾省司(山梨大)

- ◆本研究では、Famell と Adler の SAW 伝搬特性解析および FEM を用いて、分極反転構造を模擬した中間電極挿入構造 Sc_{0.4}Al_{0.6}N、Yb_{0.33}Al_{0.67}N 薄膜/高音速基板(Diamond、6H-SiC、AIN、Al₂O₃、Si)上のセザワモード RSAW 伝搬特性を解析した。
- ◆中間電極挿入 Sc_{0.4}Al_{0.6}N 薄膜基板上のセザワモードではすべての高音 速基板との組み合わせにおいて、h_dλ 増加に伴い ピ が増加、最適 h_dλ で ピ が最大となり、Sc_{0.4}Al_{0.6}N 単層薄膜構造の約 1.4-1.7 倍となった。 粒子変位を比較すると、単層薄膜や分極反転薄膜と比較して中間電極 挿入構造は SAW エネルギーがより薄膜に集中していることが判明した。この Sc_{0.4}Al_{0.6}N 薄膜への SAW エネルギーの集中が中間電極挿入 構造での高結合化につながったと考えられる。また、中間電極挿入 Yb_{0.33}Al_{0.67}N 薄膜基板上のセザワモードでも同様に ピ が増加した。

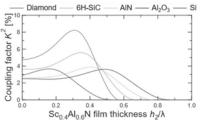


Fig.1: K^2 of Sezawa-mode RSAW on Sc_{0.4}Al_{0.6}N film with intermediate electrode/high velocity substrates as a function of h_2/λ .

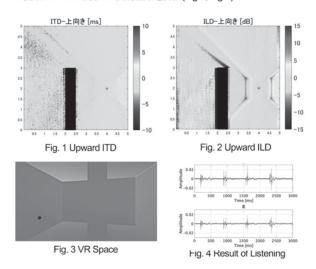
2-P-40

2-P-40 伝達関数のシミュレーションを用いた VR 空間における立体音響の再現とその評価

Reproduction of Stereophonic sound in VR space using transfer function simulation.

☆片桐明日香, △日比野祐吾, 中島弘史(工学院大)

- ◆個人ユーザーでも簡単に立体音響を再現することを目的に、FDTD 法 を用いた音響シミュレーションで両耳のインパルス応答を計算し、リ アルタイムに音源に畳み込み聴取する Unity 用の SDK を開発.
- ◆両耳間時間差(Fig.1)と両耳間レベル差(Fig.2)でシミュレーション結果 の音環境を評価
- ◆構築した VR 環境での受聴結果を評価(Fig.3, Fig.4)



2-P-41 オープンイヤー型イヤホンと室内スピーカを 組合せた聴取における残響制御音の 空間知覚評価

Evaluation of spatial perception of reverberation-controlled sound in listening with a combination of open-ear earphones and room loudspeakers.

〇野口 賢一, 千葉 大将, 小塚 詩穂里, 加古 達也, 伊藤 弘章 (NTT), 亀川 徹 (東京藝術大)

- ◆周囲で発生する音を自然に聞くことが可能であるオープンイヤー型 イヤホンの特徴に着目し、イヤホン再生音と室内スピーカ再生音を同 時に聴取することによる音像定位の距離制御を試みる.
- ◆HRIR を畳み込んだ音源信号をイヤホンから、RIR を畳み込んだ音源 信号をスピーカから再生する。イヤホン再生音に対するスピーカ再生 音のレベル比で直間比を制御し、音像定位の距離制御を図る。
- ◆スピーカ再生による残響付与で、音像距離が変化する可能性がみられ

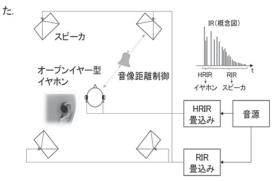


Fig.1:System diagram of proposed method

2-P-43

2-P-43 イマーシブサウンドシステムを用いた 新たな音楽表現の試み

一残響制御と音像制御を組み込んだクラシック楽曲の制作と実演— A challenge for new musical expression using an immersive sound system: unprecedent music composition and state-of-the-art sound design based on the latest reverberation and sound image control system

◎大木 大夢, 宮崎 秀生(ヤマハ), Kim Sungyoung(KAIST/RIT), △Uhm SiHyun(ジュリアード音楽院)

- ◆イマーシブサウンドシステムの使用を前提としたクラシック楽曲の 制作を試みた。
- ◆ 残響制御・音像制御システムをクリエータの表現を拡大するツール として用いることで、新たな音楽表現が生み出されると考えた。
- ◆制作した楽曲は実験会で実際に演奏を行い、来場者のアンケートから その効果を確認した。
- ◆本発表では、楽曲に用いたシステムの概要と効果を報告し、 今後の高臨場感音響システムのあり方について考察する。



Fig.1: Music score of the produced piece. The letters Setting indicate the pattern of reverberation.



Fig.2: Experimental Concert

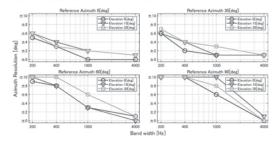
2-P-42

2-P-42 音源の仰角及び帯域が 音源定位分解能に与える影響

Effect of elevation angle and bandwidth of sound source on localization resolution

☆齋藤優季, 西口正之, 渡邉貫治, 安倍幸治(秋田県大)

- ◆ 音源信号の方位・仰角・帯域の違いが人の定位分解能にどの 様に影響するのか明らかにすることを目的とした。
- ◆11 個のスピーカを並べ、中央スピーカの方位である基準方位4種(0°,30°,60°,90°)、仰角3種(0°,15°,30°) の12 条件について、4種の帯域雑音(200,400,1k,4k [Hz]でLPF した白色雑音)を、基準方位との差分角10種(2°間隔で±10°0°を除く)のスピーカからランダムに提示し、3-AFCを用いて音源方向を回答させた。
 - ▶ 被験者の正面ほど分解能が高くなり、側方ほど低くなった
 - ▶ 仰角が上がるにつれ分解能が低くなった
 - ▶ 低域は分解能が低くなり、高域ほど高くなった



2-P-44

2-P-44 頭部伝達関数で合成された超近接音像の 心理的距離に関する研究

A Study on the psychological distance of Super Near Field sound images synthesized by the Head Related Transfer Functions

☆佐藤 秀英, 下倉 良太, △飯國 洋二(大阪大)

- ◆ 音の3次元空間表現が求められる中で、距離表現、特に近接領域および超近接 領域における距離定位をHRTFにより実現することを目的とする。
- ◆本研究ではヘッドホン着用を想定した環境において、超近接音を再現可能な頭部伝達関数(SNF-HRTF)を合成する手法を提案する。今回、健聴者を対象とした音像距離に関する実験を実施した。
- ◆本報告では、無響室で測定された外耳道入口部から3,5,8,12cmの参照音と提案手法により作成した試験音を比較した.6名の被験者はFig.1のGUIに従い、提案手法の心理的距離を調整するパラメータの調整を自身で行なった。そして、音像距離における両者の心理的距離を一致させた。
- ◆Fig. 2 に結果を示す。図は各参照音に対する被験者の解答結果を示しており。 回答結果の中央値を結んだものである。参照音が外耳道入口部に近づくにつれて、パラメータは大きくなることが確認できた。



Fig.1: GUI of Adjustment Method

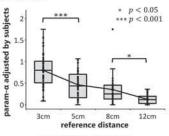


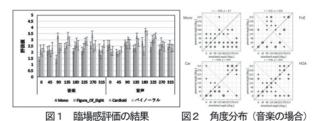
Fig.2: Subjects' answers (N=48 for each reference stimulus)

2-P-45 8ch 水平方向スピーカアレイを用いた単一 音源の最適再生方法の検討

Study of the optimal reproduction method for single sources using horizontal loudspeaker array.

☆岡田拓真, 山高正烈 (愛知工科大)

- ◆本研究では、スピーカによる聴取環境を想定し、Ambisonics 録音方式によって収録した音空間を再生する場合の最適な再生方法の構築を目指した。具体的には、Ambisonics マイクを使用して音を収録し、収録した音信号を複数の再生方式に変換させ、音の種類と最適な再生方法の関係について検討を行った。
- ◆音源は、音楽と音声の2種類であり、再生方式は、モノフォニックと ステレオ再生方式2種類 (Figure-Of-Eight, Cardioid), HOA の計 4 条件を採用した。
- ◆高次感性評価および音像定位精度を調べた結果、後方方向の音が前方 方向に比べて臨場感の評価が高いことが分かった。方向知覚の結果からは、後方の音像定位の精度を高めるには HOA 方式を,前方方向の 場合は Cardioid 方式を使用するとよいことが分かった。



2-P-47

2-P-47 オープンイヤー型イヤホンにおける 耳介形状の特性を補正した仰角知覚評価

Evaluation of elevation perception in open-ear headphones compensated for the acoustic characteristics of auricle shape.

◎千葉 大将, 小塚 詩穂里, 加古 達也, 渡邊 悠希, 伊藤 弘章, 野口 賢一 (NTT)

- ◆ 外耳孔より上に音波の放射孔があるオープンイヤー型イヤホンでは、イヤホン から外耳孔までの特性であるヘッドホン伝達関数 (HpTF) の影響により、再生 音の音像が元の仰角方向よりも上方に定位しているように知覚される。
- ◆ HpTF による音像定位精度の劣化を改善する方法として、ダミーヘッドや聴取 者自身の HpTF から算出した補正フィルタを利用する手法がある。そこで、オ ープンイヤー型イヤホンの再生音に補正フィルタを適用し、聴取者の仰角知覚 に違いが生じるか、また、音像定位精度を改善できるかを調査した。
- ◆ 主観評価実験の結果、ステレオ再生では補正フィルタの適用によって仰角知覚 が有意に低くなった。また、バイノーラル再生では補正フィルタを適用するこ とで仰角の高さに関する音像定位精度が改善する傾向を確認した。

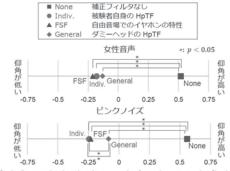


Fig.1:Scale diagram showing elevation perception for each compensation filter in stereo reproduction.

2-P-46

2-P-46 斜放射パネルスピーカを用いるサラウンド システムの音場解析

Sound field analysis of a surround system using oblique-radiation panel loudspeakers

☆鈴木優斗 (筑波大・院・シス情工)。

若槻尚斗、海老原格、善甫啓一、水谷孝一(筑波大・シス情系)

◆ 背景·目的

少チャネルかつ広い受聴範囲をもつサラウンドシステムを実現するため、平面波を斜めに放射するパネルスピーカに注目し、これを矩形状に配置することで空間中に均一な音圧と伝搬方向を持つ音波を伝搬させるスピーカーシステムを提案した。そこで、パネルスピーカを L 字状に配置した場合の音場を計測し、解析を行うことで提案システムを実現する上で望ましい特性を有することを確認する。

◆ 結果

観測領域全体にわたり、概ね一様な方向かつ音圧差が小さい分布が観測された。これは広い受聴範囲の実現に資する特性である。

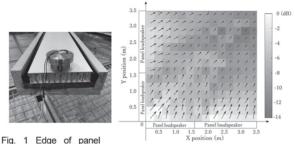


Fig. 1 Edge of pane loudspeaker

Fig. 2 Direction of sound and sound pressure

2-P-48

2-P-48 オープンイヤー型イヤホンにおける 耳介形状に基づく伝達特性の個人性分析

A study on the personalization of transfer function in open-ear earphones based on auricular shape measurement.

◎渡邊悠希, 千葉大将, 小塚詩穂里, 加古達也, 伊藤弘章, 野口賢一 (NTT)

- ◆ イヤホンでの立体音響再生では、ヘッドホン伝達関数(HpTF: Headphone Transfer Function)の影響で音像定位精度が劣化するため補正が必要である。
- ◆ HpTF は聴取者の個人性にも依存するため、これを考慮した補正を行うことは 音像定位精度向上のために有用である。一方、HpTF の測定と補正フィルタ 生成を各聴取者に対し個別に行うことは実用の観点でコストが大きい。
- ◆ 耳介形状から適切な補正フィルタを推定できればより低コストに音像定位精度 を改善できると考え、HpTF の形状分析と耳介形状との関連調査を行った。
- ◆ その結果,HpTF の形状と耳の大きさやくぼみに関連がみられた.

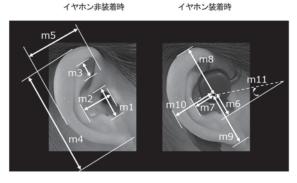


Fig.1: Measurement point of auricular shape

2-P-49 正中面における仰角の異なる 音源間の空間的マスキング効果

* Spatial auditory-masking between source signals with different elevation angles on the median plane

☆藤城博人, 西口正之, 渡邉貫治, 安倍幸治(秋田県立大),

- ◆「空間的マスキング効果」を3D音響信号の効率的な符号化へ活かす ことを目的に正中面内の異なる仰角の2音源間でのマスキング閾値 について調査を行った。
- ◆正中面内でマスカとマスキの音源配置を移動させてもITD/ILDが変化 しないためマスキング閾値が大きく変化しないことを予想していた。
- ◆聴取実験の結果、マスカ、マスキの周波数によっては、その方向によってマスキング閾値が変化するものが存在した。Fig. 2 下図の条件ではマスカに対してマスキの仰角が変化してもマスキング閾値が大きく変化しておらず予想通りの結果となった。一方で、Fig. 2 上図ではマスカに対してマスキが天長位置に移動するにつれてマスキング閾値が下降するという結果となった。
- ◆本研究では聴取実験の結果とともに、結果を考察するために追加で実施した測定についても述べる。

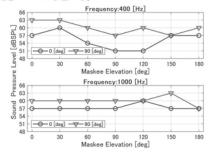


Fig. 1: Maximum masking threshold for each elevation angle of the maskee relative to the elevation angle of the masker

2 - Q - 1

2-Q-1 マスクおよびフェイスシールド着用話者の 音声コーパス構築

Construction of a speech corpus from speakers wearing masks and face shields

〇福森隆寛, 永瀬亮太郎, △許凱, 山下洋一(立命館大)

- ◆マスクとフェイスシールド着用話者の音声コーパスを構築
 - ▶ 三種類の着用状態(不織布マスクを着用,不織布マスクとフェイスシールドを着用,未着用)で収録
 - ▶ 話者: 20代の大学院生と学部生8名(男性5名,女性3名)
 - ▶ 発話文: ITA コーパスの Recitation セット (324 文)
 - ▶ 合計 7,776 個の音声データを収録
- ◆構築した音声コーパスを用いて、マスクの着用状態の分類実験と、声質変換モデルによる音声明瞭度の改善効果の有無を確認するための音声認識実験を実施(Fig. 1 は実験に使用した音声サンプルの一部)

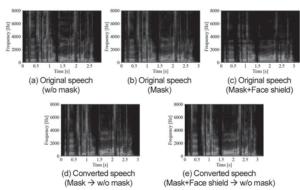


Fig.1: Spectrograms of recorded speech and converted speech samples.

2-P-50

2-P-50 異なる深層学習モデルによる基準 HRTF の 推定と仰角方向への拡張

Estimation of reference HRTF and extension in elevation angle by different deep learning models

☆齋藤匡紀,下倉良太,△飯國洋二(大阪大)

- ◆頭部伝達関数(HRTF)によって、ヘッドホン再生で立体音響を実現できるが、HRTFには個人差があり測定も困難である。
- ◆200 以上のサンプル点をもつ HRTF を DNN で直接推定すると、過学 習の恐れがあるため、HRTF を PK フィルタで近似することで情報量 を削減した。
- ◆そして、正面方向から他仰角への HRTF の変化を PK フィルタの変化 として学習する DNN モデルを提案した(Fig.1).
- ◆NFD という指標で、提案した DNN モデルの有用性を確認した.
- ◆さらに入力角度を変更することで、DNN モデルを複数作成して、ア ンサンブル学習を行った。
- ◆その結果, 先行研究よりも推定精度が向上した.

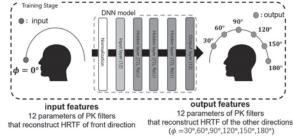


Fig.1: Proposed DNN model

2 - Q - 2

2-Q-2 潜在変数モデルを用いた CTC による 非自己回帰型音声認識

Non-autoregressive automatic speech recognition based on connectionist temporal classification and latent variable models

○藤田悠哉(ヤフー)、渡部晋治(カーネギーメロン大)

- ◆機械翻訳の研究分野で良好な結果が報告されている潜在変数モデルとCTCを組み合わせた新しい非自己回帰型音声認識手法を提案する.
- ◆ニューラルネットワークの構造の工夫や, Intermediate CTC といった 既存手法と組み合わせることで, 自己回帰型のモデルより良い WER を達成した。å

Table 1: Comparison of WERs on TED-LIUM2

or WERS on .	LED-L	TUMZ
Beam or	dev	test
#Iter.		
1	7.7	8.4
10	7.3	7.8
1	8.1	8.1
10	8.0	8.0
1	8.1	7.8
1	7.2	7.5
1	7.3	7.7
1	7.4	7.6
3	7.0	7.2
	Beam or #Iter. 1	#Iter. 7.7 10 7.3 1 8.1 10 8.0 1 7.2 1 7.3 1 7.4

2-Q-3 言語現象と非言語現象も検出する 音声認識システムの提案

Proposal for automatic speech recognition to detect verbal/non-verbal phenomena

〇塩根凪人, 若林佑幸, 北岡教英 (豊橋技術科学大学)

背景

- ◆会話を理解する上で言語・非言語現象の認識は必要
- ◆一般的な音声認識モデルは言語情報だけ認識

目的

◆ 言語情報と 9 種類の言語・非言語現象の同時認識を行う 音声認識モデルを構築

実験

- ◆日本語日常会話コーパスを用いて音声認識モデルを構築
- ◆言語・非言語現象を示すタグの付与位置の変更による 音声認識への影響調査
- ◆Shallow Fusion による外部言語モデルの統合

結果

- ◆ 言語情報と言語・非言語現象の同時認識を行う音声認識モデルが実現
- ◆ 言語情報に対して言語・非言語現象を示すタグを前に付与する 音声認識モデルが最適
- ◆ 言語情報と言語・非言語現象の同時認識により音声認識の性能が向上
- ◆外部言語モデルの統合により音声認識の性能が低下

2 - 0 - 5

2-Q-5 Lighter Training and Faster Decoding for RNN-Transducers: A Dynamic Pooling Mechanism based on Continuous Integrate-and-Fire

☆ Wen Shen TEO, Yasuhiro MINAMI (UEC)

- •We aggressively reduce the acoustic vector length (T) in RNN-T down to the number of symbols (Sym) or the number of morphemes (Mor) in the transcript.
- We do this using CIF as a segment-wise pooling mechanism with dynamic boundaries.
- •We reduced training memory footprint and increased decoding speeds, with varying impacts across different datasets.
- •We propose Ragged Attention (R.Attn) Downsampling that outperformed the conventional Cascade method and minimized the CER degradation associated with aggressive downsampling.

	Model	CIF-less	Cas	cade	R.A	Attn
	Add. Param.	0	446,850		447,234	
	M^* Type		Sym	Mor	Sym	Mor
Ð	test-clean (CER)	4.26	4.60	5.34	4.39	4.63
Librispeech	test-other (CER)	13.02	13.83	14.79	13.41	13.52
	Mem. Red. (%)	100	99	100	098	102
	Speedup (%)	100	128	154	123	149
	eval1 (CER)	4.03	4.37	4.98	4.28	4.44
_	eval2 (CER)	3.42	3.71	4.22	3.45	3.87
S	eval3 (CER)	3.58	3.88	4.30	3.82	4.27
	Mem. Red. (%)	100	149	180	148	179
	Speedup (%)	100	115	139	111	134

Table 1: CERs, additional parameters (Add. Params.), memory reduction (Mem. Red.), and decoding speedup factor (Speedup) comparison among ragged downsampling submodules and target length Mr definitions

2 - Q - 4

2-Q-4 話し言葉テキストと書き言葉テキストの統合 系列を出力する End-to-End 音声認識

End-to-End Automatic Speech Recognition to Output Dual Text of Spoken and Written Texts

◎庵 愛, 佐藤 宏, 田中 智大, 増村 亮, 水野 沙希, 北条 伸克 (日本電信電話株式会社, NTT コンピュータ&データサイエンス研究所)

- ◆本稿では、話し言葉テキストと書き言葉テキストを同時に獲得する必要のある場合に、1モデルで両テキストを出力する方法を提案する。
 - 従来、話し言葉テキストと書き言葉テキストの両方を獲得するためには、音声認識モデルと話し言葉・書き言葉変換モデル、または書き言葉テキスト推定モデルの2つのモデルが必要であった。
 - 提案手法では、音声を入力すると話し言葉テキストと書き言葉テキストの統合系列を出力するため、メモリ確保の観点で効率的に両テキストを獲得できると考えられる。

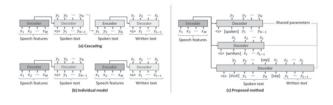


図1:音声認識と音声からの書き言葉テキスト推定における 従来手法と提案手法

2 - Q - 6

2-Q-6

Neural Transducer を用いた 目的話者音声認識システムの検討

A Study on Target-Speaker ASR with Neural Transducer 森谷崇史, 佐藤宏, 落合翼, デルクロアマーク (NTT),

- ◆本研究では複数話者の音声が含まれる混合音から目的話者の音声 のみを認識する目的話者音声認識タスクを実施
- ◆先行研究では目的話者音声抽出モデル (TSE) を用いて目的話者の音声を推定し、抽出された音声を後段の音声認識システム (例. RNNT) によって認識するカスケード方式 (TSE+RNNT) を提案 × TSE によって認識処理に大きな遅延が発生
- ◆本研究では音声認識における RNNT と TSE モデルの要素を組み合わせた Target-Speaker RNNT (TS-RNNT) を提案
- ◆実験により以下の条件を満たす結果が得られることにより提案手法 である TS-RNNT の有効性を確認
 - > TSE が不要のため RNNT と同程度の処理時間で動作可能
 - ➤ TSE+RNNT と同等以上の認識精度を達成

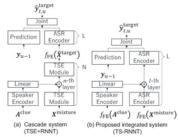


Figure 1: The overview of target-speaker ASR systems

Table 1 : Results of each target-speaker ASR system.

System	CER [%]	RTF
RNNT	76.1	0.40
TSE+RNNT	16.5	1.22
TS-RNNT	15.8	0.40

講演取消

2 - Q - 8

2-Q-8 Mecon Audio: Medical Conference Audio corpus

井上 顧基(Elith·AI先端科学研)

- ◆医療分野での音声技術の応用は大量の医療音声データが必要である。
- ◆現状、利用可能な日本語の医療ドメイン音声コーパスは非常に少ない状況である。
- ◆我々は厚生労働省の先進医療会議議事録を利用し、医療用語を 含むコーパスを構築した。
- ◆日本語を母国語とする女性話者一人を雇い、先進医療会議議事 録の一部 (2022年~2023年途中)及びITAコーパスの読み上げ音 声を収録した。
- ◆コーパス構成を以下に示す。

dataset/ __wav/ __m001-0001.wav m005-xxxx.wav __text/ __transcribe-text.txt __transcribe-hiragana.txt

◆現時点では女性話者一人の音声に留まっているが、今後が話者 数を増やしていき、より大規模な医療ドメイン音声コーパスの実現 を目指す。

2 - 0 - 9

2-Q-9 SSL に基づく音声データのみを用いた 音声認識モデルのドメイン適応

Domain adaptation with a nonparallel target domain corpus.

〇木内貴浩 1 小川厚徳 2 若林幸弘 1 北岡教英 1 (1 豊橋技科大, 2 NTT)

目的

- ◆ ASR モデルの精度は、学習データのドメインに大きく依存
- ◆ ドメイン内音声と書き起こしデータの準備は高コスト.
- ◆ ドメイン内の音声データのみであれば、低コストで収集可能 手法
- ① 自己教師あり学習モデルをドメイン内音声で追加事前学習し、ドメイン外のパラレルデータで Fine-tuning することにより、ドメイン内のパラレルデータを一切使わずにドメイン適応
- ② さらに、ノンパラレルコーパスでドメイン適応するために、①に加え て Density Ratio Approach (DER)によってドメイン内テキストのみ でドメイン適応

結果

ドメイン内コーパスを用いない場合と比べて、文字誤り率(CER)は

- ◆ 音声だけを用いた手法の場合 26.6%の相対 CER の削減を達成
- ◆ DRA も併用した場合 33.6%の相対 CER の削減を達成.

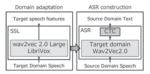
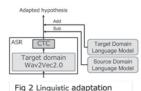


Fig 1 Acoustic adaptation using additional pre-training



using language model fusion

2 - Q - 10

2-Q-10 Density Ratio Approach に基づく複数音声認識モデルの統合による頑健な音声認識システムの構築

Constructing Robust Speech Recognition System by Integrating Multiple Speech Recognition Models based on density ratio approach.

☆北條圭悟, 森大輝, 若林佑幸(豊橋技科大),

小川厚徳(NTT), 北岡教英(豊橋技科大)

- ◆複数の音声認識システムの統合は、認識精度向上に有効であるが、 End-to-End 音声認識で用いられることは少ない。
- ◆End-to-End 音声認識では、言語モデルの併用が精度向上に効果的で、 Density Ratio Approach(DRA)はその手法の1つである。
- ◆複数の音声認識システムの統合と DRA の考え方を利用した新たなド メインにおける効率的な音声認識システムの構築を提案する。
- ◆ 実験の結果、提案手法は新たな学習なしで、従来の言語モデル統合 手法の認識精度を上回ることが確認された。
- ◆また、提案手法は従来の統合手法である ROVER よりも効率的な統合 が可能であることが判明した。

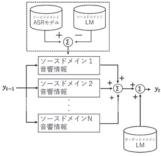


Fig.1: Overview of the proposed method

2-Q-11 End-to-end 音声認識器の中間層への言語 知識転移

Transferring Language Knowledge into Intermediate Layers of End-to-end Speech Recognisers

◎Michael Hentschel(ワークスモバイルジャパン)、西川勇太(奈良先端大)、小松 達也、藤田雄介(LINE)

- ◆CTC 型の音声認識(ASR)モデルへの BERT 言語モデルの知識蒸留 (KD) に対する新しい手法を提案する
- ◆提案手法は、中間層へ BERT の知識蒸留を行うための attention デコーダを導入する
- ◆提案の中間層への蒸留は、従来の最終層のみへの蒸留と比較して効率 良く言語知識を転移できる。
- ◆提案手法は、グリーディ CTC デコードの高速さを保ちつつ、外部 LM との shallow fusion を用いずに言語知識を考慮した認識ができる。
- ◆ Table 1 に示す LibriSpeech データセットを用いた実験により、提案手 法の有効性を確認した。

Model	clean		other	
	dev	test	dev	test
CTC ベースライン	3.74	3.89	9.24	9.09
提案法	2.75	2.94	6.68	7.02

Table 1: LibriSpeech 960におけるグリーディCTCデコードの WER[%]

2 - Q - 13

2-Q-13 データの分布マッチングによる End-to-End 音声認識モデルのドメイン適応

Domain adaptation by data distribution matching for end-to-end speech recognition

〇篠原雄介(ヤフー), △渡部晋治(CMU)

- ◆ターゲットドメインのテキストが与えられた時, 音声認識モデルをドメイン適応する問題について検討する.
- ◆従来は、所与のテキストから構築した言語モデルを統合する「shallow fusion」がよく用いられた。しかし、モデルサイズ・推論演算量が増加するため、実応用(特にオンデバイス)での利用が難しかった。
- ◆本講演では、既存のマルチドメイン学習データから、ターゲットドメ インの分布と適合するサブセットを選択して、モデルを fine-tune す る方法を提案する. この時、劣モジュラ最適化に基づくアルゴリズム でサブセットを選択する.
- ◆LibriSpeech での実験で、モデルサイズ・推論演算量を増やすことな く、ターゲットドメインで認識精度が改善することを報告する。

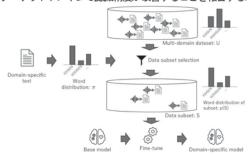


Fig. 1: Workflow of the proposed method.

2 - Q - 12

2-Q-12 深層強化学習を用いた特徴量変換に よる音声認識システムの適応法

Adaptation method for speech recognition systems based on conversion of spectrum using deep reinforcement learning 〇鈴木 基之, △石田 涼(大阪工大・情報科学)

研究の目的

入力された音声をより認識しやすい音声へ変換することで,音声認識 システム自体に手を加えることなく,利用環境に適応させて認識精度を 向上させるシステムを開発

特徴量変換システムの概要

- 特徴量変換器は深層強化学習を用いて構築
 - 変換が大きくなりすぎないよう学習を制御
- 変換後音声の認識結果を元に報酬を定義
 - 音素誤り率 (PER) に基づく定義 (定義 1)
 - 認識音素の正誤に基づく定義(定義2)

音声認識実験の結果

- 外国訛りを含む日本語音声を用いて実験
- 音声を変換することで音素誤り率を改善
 - 定義1:2.3% 削減,定義2:3.6% 削減
- 変換前音声のPERが高い(認識が困難な)音声に対しては定義 1が、PERが低い音声に対しては定義2が有効

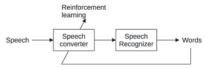


Fig.1 Block diagram of the system

2-Q-14

2-Q-14 耐雑音音声認識のウェーブレット レベル内分割による改善の基礎的検討

A basic study of noisy speech recognition using the divisions in the level on wavelet spectrum

〇緑川洋一(大分大・理工)

音声認識技術が一般に広がるため更なる改善が必要な技術の一つとして、雑音環境下音声認識の認識率の改善があげられると思われる。入力装置としての音声入力の一般へのさらなる普及のためには不可欠な技術の一つである。しかし、まだ信頼できる決定的な技術が現段階では確立されてはいないと思われる重要な課題の一つであると考える。我々は以前よりこのような点に着目し、雑音環境下における音声認識について研究を行ってきている。一般的に信号処理で用いられる常套手段としてはフーリエ変換による周波数領域における処理があるが、これと少し異なる解析手法であるウェーブレット変換を用いることにより雑音環境下音声認識の認識率改善ができないか基礎的検討を行っている。

研究の1つとして、我々は雑音環境下音声認識の認識率の改善のため、認識時の標準データとする無雑音の音声データにはウェーブレット変換を用いて強調することにより擬似的に雑音を加えるような処理を行い、入力の雑音の重畳した音声データにはウェーブレット変換を用いて圧縮することにより雑音の低減をし、両者を併用することにより特徴を近づけ、認識率の向上を狙う方法において、今回はウェーブレットレベル内を分割して閾値条件を設けて行う方法でも認識率改善ができるか基礎的検討を行ったので報告する。

2-Q-15

Investigation on Multi-task Universal Speech Models

O Peng Shen, Xugang Lu, Hisashi Kawai (NICT)

- We conducted an investigation on improving the performance of language identification (LID) and speech recognition (ASR) by using the Whisper model to automatically identify languages for building multilingual ASR systems.
- To enhance the system's performance, we proposed a method of joint learning fine-tuning aimed at not only improving LID performance but also maintaining the ASR performance.
- Experimental results demonstrated that the proposed method could improve performance on in-domain datasets. However, for out-of-domain datasets, although the proposed method could enhance LID performance, challenges still remain in the ASR tasks.

Method	Dataset	Acc-1s	Acc-F	WER	WER-GT
whisper	MLS	47.47	99.32	31.51	30.78
FT 0.0	MLS	79.47	99.97	36.34	36.33
FT 0.001	MLS	81.04	99.97	17.96	17.95
FT 0.01	MLS	80.95	99.97	17.09	17.08
FT 0.1	MLS	80.13	99.95	17.12	17.11
FT 0.3	MLS	81.04	99.94	17.16	17.15
FT 1.0	MLS	79.96	99.92	17.19	17.12
whisper	CV	18.33	90.26	39.48	35.24
FT 0.0	CV	65.29	92.26	52.55	48.30
FT 0.001	CV	68.21	94.44	47.28	45.45
FT 0.01	CV	66.53	93.88	46.16	44.81
FT 0.1	CV	65.86	93.96	46.16	45.28
FT 0.3	CV	67.26	94.21	46.57	45.56
FT 1.0	CV	67.16	93.70	47.23	45.91

Fig.1: Results of the baseline and proposed methods: Acc is the evaluation for LID, and WER is for ASR tasks. Acc-1s represents duth 1-second utherances, Acc-F is data with full-length utherances, and WER-GT refers to results obtained by setting the ground-turth language label.

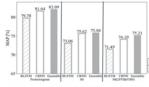
2 - 0 - 17

2-Q-17 音声中の音声検索語検出におけるアンサンブル学習モデルを用いた検索精度向上

Improving Retrieval Accuracy Using Ensemble Model in Query-by-example

☆山賀光, △小嶋和徳(岩手県立大), 李時旭(産総研), 伊藤慶明(岩手県立大)

- ◆ 音声中の音声検索語検出 (SQ-STD) の代表的な実現方式の Posteriorgram 照合 方式は、深層学習モデルに BLSTM 等を用いた場合、高い精度が得られる一方、 検索時間が長くメモリ量が大きい。
- ◆検索精度の向上及び検索時間とメモリ量の削減のため、様々な手法が提案された(最尤系列化照合、フレーム数削減、スコア統合)。
- ◆ 提案手法: BLSTM とは異なる特徴抽出アプローチを持つ CRNN を新たに導入 し、BLSTM とのアンサンブル学習により Posteriorgram を再構成する。提案モデル及び他モデルに先行手法を適用し、スコア統合により精度向上を図る。
- ◆実験結果: Posteriorgram 照合では、CRNN の導入により MAP が BLSTM の 79.78%から 81.64%と 1.86pt 向上し、さらにアンサンブルにより 82.09%と 0.45pt 向上した。最尤系列化照合を用いた場合も同様の傾向を確認した。スコア統合では、ベースラインと比較して検索時間は約 0.2 秒削減し 0.7 秒以下、メモリ使用量は同等で MAP は NTCIR-10 で 90.48%、 NTCIR-12 で 88.02%と これまでで最良の結果となった。以上より提案手法の有効性を確認した。



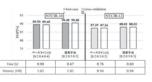


Fig.1: Results of BLSTM, CRNN and Ensemble

Fig.2: Results of Score Integration

2 - Q - 16

2-Q-16 言語情報埋め込みベクトルを利用した 音声表現学習の検討

Cross-Lingual Speech Representation Learning with Language ID Embeddings

〇田中 智大, 増村 亮, 佐藤 宏, 庵 愛, 山根 大河 松浦 孝平, 芦原 孝典, 森谷 崇史 (NTT)

- ◆音声の教師なし事前学習によって音声認識の精度が改善するが、高精度なモデルを構築するには大量のデータが必要
- ◆大量の音声データを利用する方法として、多言語音声のデータを利用 する方法が考えられる
- ◆ これまで提案されてきた多言語音声データを用いた手法では、言語情報を直接利用されていない
- ◆本稿では言語情報埋め込みベクトルを入力として利用する音声表現 学習手法を提案
- ◆言語情報埋め込みベクトルを入力する方法として、Transformer に足 し合わせる方法と、時間方向に結合する方法の2つを検証
- ◆実験は日本語と英語データで行い日本語は文字誤り率 (CER), 英語 は単語誤り率 (WER) で評価
- ◆データを単純に足し合わせる場合と比較して、後段の音声認識タスク の精度が改善することを確認

Model Data			CSJ (%CER)			LS (%WER)			
Model	for SSL	dev eval1 eval2 eval3	dev-clean	dev-other	test-clean	test-other			
	CSJ	3.60	4.76	3.49	3.79	3.32	8.89	3.42	9.11
Baseline	LS	3.66	4.77	3.56	3.81	3.23	8.61	3.30	8.98
	CSJ+LS	3.61	4.80	3.36	3.78	3.20	8.48	3.31	8.93
LID-ADD	CSJ+LS	3.54	4.60	3.39	3.72	3.08	8.27	3.13	8.14
LID-CONC	COU+LO	3.46	4.56	3.36	3.61	2.98	8.29	3.24	8.11

Table. 1: CER and WER results. LS: Librispeech.

2 - Q - 18

2-Q-18 ASR 自動字幕の低コストな修正支援技術の リアルタイム化とオンライン評価

Real-time and Online Evaluation of Low-cost Correction Support Techniques for ASR Automated Subtitling

☆片岡諒弥(静岡大院・総合科学技術研), 甲斐充彦(静岡大)

- ◆本稿ではASR システムを用いた自動的な字幕作成システムと字幕修正作業の低コスト化のための字幕の半自動的な修正技術をリアルタイム化し、実際に被験者がシステムを用いて字幕の修正作業を行うことでシステムの操作性や修正精度の評価・検証を行う。
- ◆短時間単位で最新のデコード状態を出力することで発話セグメント 単位ではなく、各単語単位のデコード状態の表示へと字幕表示までの 遅延を改善。
- ◆修正作業において修正後の単語のみを入力する操作法により、修正時の作業量を低減、修正作業を行うことで音声認識結果の文字誤り率、 名詞誤り率、名詞句部分の文字誤り率を減少した。

Table. 1: Comparison of recognition accuracy before and after subtitle modification [%]

	before	after
Character error rate	20.6	20.2
Noun-phrase error rate	29.0	27.7
Character error rate for	21.5	20.2
noun phrase parts	21.0	20.2

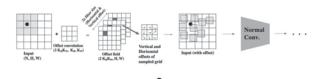
2-Q-19

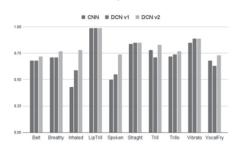
変形可能受容野に基づく深層演奏テクニック識別

〇山本雄也(筑波大), Nam Juhan (KAIST), 寺澤洋子(筑波大学)

スペクトログラム上変則的なパターンが観測できる

テクニックに対し、<u>Deformable Convolution</u>を適用して<u>受容野を変形</u>させた CNN を使えば、テクニック識別問題の性能向上に役立つか?





2 - Q - 21

2-Q-21 ランダム直交行列を用いた秘密鍵による 音声のプライバシー保護法

Speech privacy-preserving method with secret key using random orthogonal matrix ☆丹羽祥子,塩田さやか、貴家仁志(都立大)

◆ 背景と目的

- クラウドサービスにおけるデータ漏洩の可能性
- 発話内容の秘匿など、音声のプライバシーの保護が必要
- ▶ 先行研究における秘密鍵の予測が容易
 - → Fig. 1 に示すシナリオを想定した、 ランダム直交行列を用いた秘密鍵による音声の暗号化法の提案

◆ 提案法の評価

- ▶ 正規ユーザのみが正しくモデルを使用可能
- ▶ 暗号化した音声 (Fig.2) の発話内容を聞き取ることは困難
- ▶ 先行研究よりも鍵空間が大きくなり、さらに秘密鍵の予測が困難
 - → 秘密鍵の安全性を高めながら、音声のプライバシー保護が可能

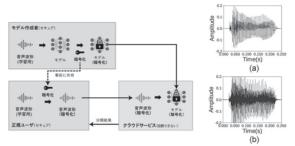


Fig.1: Privacy-preserving Scenario

Fig.2: Audio encryption (a): original, (b): encrypted waveform

2 - Q - 20

2-Q-20 人間は認識して機械は誤認識する 音声コマンドの研究

Research on audio commands that are recognized by humans and misrecognized by machines ☆信川裕介, 下倉良太, △飯國洋二(大阪大)

- ◆音声認識の分野に注目が集まる中、Fig. 1 に示すような音声を使用した音声 CAPTCHA の安全性が低下してしまっている.
- ◆Audio Adversarial Examples と呼ばれる。データに意図的に小さな摂動を加えることによって、聴感上元音声とほぼ変わらないが認識結果は異なるような音声に着目した。
- ◆本報告では、Carlini らが実装を行った Audio Adversarial Examples の 作成の限界を探るために実験を行った。その結果、元音声の認識結果 から大きく離れたセンテンスを目標として実行を行うと、不完全かつ 摂動に大きく違和感を持ってしまうような音声の生成が確認できた。
- ◆結果を考慮しつつ、音声 CAPTCHA に Audio Adversarial Examples を導入することで、安全性の低下を防ぐことができると考える。そこで、Fig. 2 に示す音響モデルに Julius を利用し、似ている子音(k/と/n/など)など、小さな変化を行うことでストレスのない安全な日本語の音声 CAPTCHA の作成が可能になると考え、実装を目指す。

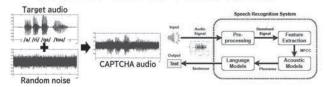


Fig.1: Example of CAPTCHA audio

Fig.2:Architecture of Speech Recognition System

2 - Q - 22

2-Q-22 VBx クラスタリングのマルチストリーム 拡張による EEND-VC の 話者ダイアライゼーションの改善

Improving EEND-VC diarization performance with multi-stream extension of VBx

○デルクロア マーク¹, 俵直弘¹, Mireia Diez², Federico Landini², Anna Silnova², 小川厚徳¹, 中谷智広¹, Lukas Burget², 荒木章子¹ (¹NTT, ²Brno University of Technology)

- We introduce a multi-stream extension of VBx (MS-VBx) for clustering speaker embeddings derived from a end-to-end speaker diarization (EEND) system.
- ◆The proposed MS-VBx reduces diarization error rate and speaker counting errors for various tasks.

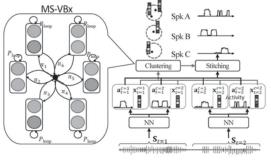


Fig.1:Schematic diagram of the end-to-end speaker diarization with vector clustering (EEND-VC), with the proposed MS-VBx for clustering the speaker embeddings

2-Q-23

2-Q-23 ラベル誤りに頑健な話者年齢推定のための サンプル選択に基づくモデル学習法の検討

Investigating model training based on sample selection for robust speaker age estimation against label error

〇俵直弘, 小川厚徳, 北岸佑樹(NTT)

- ◆ 一部サンプルに誤った年齢ラベルが付与された学習コーパスを用いて 話者年齢推定モデルを頑健に推定する方法として、サンプル選択に 基づく以下の3手法の有効性を検証
 - 1. Small-loss selection:学習時の損失値が小さいサンプルを選択し学習に利用
 - Cross update (co-teaching):同じ構造で初期パラメタが異なる二つのモデルを 遠入し、一方のモデルで選択したサンプルでもう一方のモデルを学習
 - Joint update:同じ構造で初期パラメタが異なる二つのモデルを導入し、損失値が小さく、かつ両モデルの出力の差が小さいサンプルを選択し同時に学習
- ◆ SRE08/10 から作成した擬似的なラベル誤りを含むデータセットと、 AgeVoxCeleb を用いた年齢推定実験により、いずれのサンプル選択 手法も有効であることを確認(表1,表2)

Figure 1. Mean absolute errors and correlations on SRE08/10

Methods	Noisy-label samples ratio				
Methods	0%	20%	50%		
Baseline		7.58 (.671)	13.31 (.575)		
+ Small-loss selection	5.63 (.816)	5.79 (.822)	6.35 (.783)		
+ Cross update		5.88 (.815)	6.30 (.789)		
+ Joint training		5.76 (.826)	6.17 (.800)		
Baseline wo/ noised samp.	5.63 (.816)	5.76 (.807)	6.01 (.779)		

Figure 2. Mean absolute errors and correlations on AgeVoxCeleb

Method	MAE (ρ)
Baseline	8.71 (.684)
+ Small-loss selection	8.37 (.713)
+ Cross update	8.30 (.717)
+ Joint training	8.38 (.712)

2 - Q - 25

2-Q-25 深層話者埋め込みを用いた歌唱者の照合 に関する検討

A Study on Singer Verification Using Deep Speaker Embedding

☆当間佐耶佳,有賀智輝,樋口陽祐(早大).

△早坂一寿, △岡本直紀 (第一興商), 小川哲司 (早大)

- ◆TDNN による歌唱音声の埋め込み(x-vector)による歌唱者照合

 - ☆ <u>調査2</u>: 照合に用いる歌唱音声の長さが性能に与える影響を明らかにする。

◆評価実験

- ◇ 実験1:発話音声と歌唱音声の二つのデータセットで x-vector 抽 出器の学習を行い歌唱者照合の性能を比較したところ、x-vector 抽出器の学習に歌唱音声を用いることの有効性を確認した(表1).
- ⇒ 実験2: 照合時、歌唱冒頭Nフレーズの音声を用いてx-vector を 抽出し照合性能を調査したところ、一曲全てを用いる場合と比較 して大きく性能が劣化することを確認した(表2)。

Table 1: Verification performance for TDNN's Table 2: Verification performance for number training data. of singing phrases.

Training	EER [%]	minDCF	#phrases	EER [%]	minDCF
Voxceleb	6.56	0.39	4	5.11	0.37
Singing Voice	2.63	0.11	8	3.87	0.26
			16	3.84	0.18
			full	2.78	0.10

2 - Q - 24

2-Q-24 短い発話に頑健な overlapped speech detection モデルの検討と一考察

Improved Overlapped Speech Detection for Shorter Utterances

☆町田裕璃奈(筑波大), △山岡大志(Poetics)

- ◆複数の話者が同時に発話する「オーバーラップ(overlapped speech)」は、 会議や日常会話などの多人数参加型の場面で発生し、特に相槌をはじ めとした短い特線時間のオーバーラップは頻繁に生じる。
- ◆本研究では、オーバーラップの持続時間に着目し、短い発話にも頑健なオーバーラップ検出モデルの学習方法について検討する.
- ◆実験として、まず日本語の自然発話音声から2話者のオーバーラップ を含む音声を合成する.この音声のうち、オーバーラップの持続時間 が短いサブセットを用いて、Wav2Vec2-Conformer の finetuning を実施 し、その検出率を持続時間の長短で比較した.
- ◆ finetuning を通して、従来法の pyannote より recall の値がオーバーラップの特線時間の長短に関わらず高くなった.

Table 1: evaluation with 0.5/1.0s-overlapped speech

	accuracy	recall	precision
pyannote	77.26%	11.01%	81.29%
$Wav 2 Vec 2\hbox{-}Conformer$	61.51%	22.47%	28.90%
proposed	96.43%	98.01%	76.89%

Table2: evaluation with 3.0/5.0s-overlapped speech

	accuracy	recall	precision	
pyannote	90.34%	32.36%	97.59%	
Wav2Vec2-Conformer	52.17%	22.12%	31.91%	
proposed	73.81%	48.97%	91.60%	

2 - Q - 26

2-Q-26 単語発声による同一話者判定 DNN の 学習と話者照合

Text-dependent speaker verification using Same Speaker Identification DNN trained on short phrases ☆亀田健太郎, 黒岩眞吾, 堀内靖雄(千葉大),

柘植覚(大同大), 西田昌史(静岡大)

- ◆同一話者判定 DNN の学習を、単音節音声に加えて 2~5 モーラの単語音声で行う手法を検討した。
- ◆2 つの単語音声の入力長を揃える手法として、0 埋めと DTW を用 いた対応付けの比較をした
- ◆科学警察研究所『大規模話者骨導音声データベース』の単語音声 および単母音音声を対象に話者照合実験を行った.
- ◆単音節学習と比較して、DTW を用いた単語学習によって音声長の 長い単語で EER が削減された。しかし、学習単語よりも長い7モ ーラの単語(サービスエリア)では EER の改善は見られなかった。

Tablel 単語音声を対象とした話者照合結果(EER[%])

	同時期				異時期				
	単音節学習		単語学習		単音節学習		単語学習		
	0埋め	DTW	0埋め	DTW	0埋め	DTW	0埋め	DTW	
おい	6.4	2.8	5.0	2.1	9.8	5.9	8.3	6.4	
はい	9.1	3.2	5.7	2.8	13.0	6.7	11.5	6.7	
おまえ	2.9	1.2	2.4	1.2	6.2	4.0	5.7	4.0	
くるま	5.5	1.0	3.6	0.7	9.9	3.5	6.5	3.2	
もしもし	6.0	1.9	3.4	0.3	9.7	4.5	6.5	2.4	
サービスエリア	7.4	1.0	2.3	0.6	11.4	2.9	6.6	1.8	

2-Q-27

2-Q-27 Laboro-ASV: a small yet effective add-on dataset for Japanese speaker verification

Xinyi Zhao, Hiromasa Fujihara (Laboro.Al Inc.)

- Aiming to improve the performance on the Japanese speaker verification task, we create the <u>Laboro-ASV dataset</u> from Japanese terrestrial television programs and make it open-source.
- ♦We add Laboro-ASV as a supplemental dataset to other datasets and compare the outcomes. The combination of VoxCeleb 1 and Laboro-ASV yields the best performance.
- Furthermore, we discuss the key attributes of an effective supplemental dataset.

2-Q-28

2-Q-28 話者数無制限

ブロックオンラインダイアライゼーション

Block-online speaker diarization for unlimited numbers of speakers

◎堀口翔太¹, 渡部晋治², △Paola Garcia³, 高島悠樹¹, 川口洋平¹

¹ 日立²CMⅡ ³, 旧Ⅱ

- ◆話者数を制限しない話者ダイアライゼーション手法である EEND-GLA のブロックオンライン推論を実現するために、 Block-Wise Speaker-Tracing Buffer (BW-STB) を提案した。
- ◆EEND-GLAでは、十分な短時間内(本研究では5秒以内)に発話する話者数は高々数名という仮定を用いて、全体として話者数を制限しない話者ダイアライゼーションを実現している。BW-STBではこの仮定を利用できるように、バッファをブロックに分割し、各ブロック内では時間的に連続したフレームの特徴量と推論結果を保持する。
- ◆提案不法はシミュレーションデータおよび実データにおいて既存手 法のダイアライゼーションエラー率を大きく改善した。

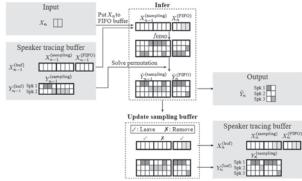


Fig.1 Block-online diarization using EEND-GLA with BW-STB

2 - Q - 29

2-Q- 自己教師ありモデル特徴量から音声波形を 29 生成するニューラルボコーダの実験的評価

An Empirical Study of Self-Supervised Learning Model Features for Speech Waveform Reconstruction

◎中田 亘, 佐伯 高明, 齋藤 佑樹, 高道 慎之介, 猿渡 洋 (東大院・情報理工)

SSLモデル を探しています

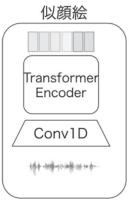


層数: 12層 or 24層 持技:

音声情報処理

特徴:

再合成音声の自然性か



※知見をお持ちの方は2-Q-29まで

2 - Q - 30

2-Q-30 対戦ゲーム動画の 実況解説音声の分析と合成の検討

Analysis and synthesis of commentary audio for competitive game videos ☆井浦 昂太, 齋藤 佑樹, 猿渡 洋(東大院・情報理工)

- ◆スポーツ実況解説を始めとして、実況解説者は対象とする映像に音声を追加することで、視聴者の認識や興奮度に影響を与える。
- ◆実況解説音声を再現できる音声合成手法の実現はエンターテインメント分野への応用が期待できる。
- ◆エンターテインメント分野として注目される対戦ゲームプレイの実 況解説音声に注目し、発話内容と音響的特徴の分析を行い、実況解説 音声の発話スタイルを再現する音声合成手法の検討を行う。

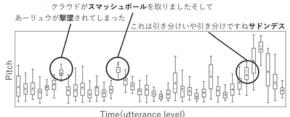


Fig.1: Analysis of a commentary speech.

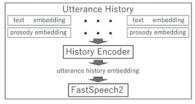
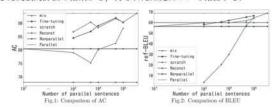


Fig. 2: An example of commentary speech synthesis model.

2-Q-31 注意機構付き VAE を用いた テキスト発話スタイル変換における 少量パラレルデータの活用

Utilization of a small amount of parallel data in spoken text style transfer with VAE and attention mechanism. 会吉岡大貴, 安田裕介, 戸田智基 (名大)

テキスト音声合成において、感情やキャラクタ性などの「スタイル」を付与した音声を合成するため、音声だけでなく「発話するテキスト」自体のスタイルも制御可能な技術が有用であると考えられる。しかし、テキストのスタイル制御を行うテキストスタイル変換では、利用可能な日本語のデータセットが整備されておらず、また人手による元テキストと正解変換文のペア(パラレルデータ)作成はコストが高い、一方、大量のパラレルデータを人手で準備するのは難しいが、少量ならば利用できる場合が想定できる。前回の音響学会で我々は、既存モデルで生成したスタイル変換テキストを疑似的な正解変換文として利用し、再構成だけでなく、疑似正解変換文を元のスタイルに戻すことで変換も学習する CycleCVAE+Attn+CWAPE を提案した。本稿では、CycleCVAE+Attn+CWAPE の枠組みで、疑似パラレルデータだけでなく少量のパラレルデータを用いることができる条件においての性能改善について調査を行った。日本語のテキストスタイル変換タスクを対象とした実験的評価の結果から、得られた知見について報告する。



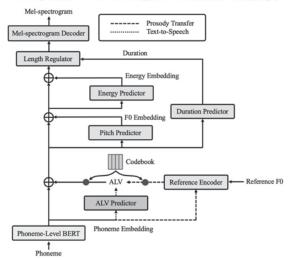
2 - Q - 33

2-Q-33 アクセント潜在変数の予測と制御が可能な TTS モデルによる方言音声合成の検討

Investigation of dialect speech synthesis using a TTS model that can predict and control Accent Latent Variable

☆山内一輝, 齋藤佑樹, 猿渡洋(東大院・情報理工)

- ◆方言音声合成には以下のような課題がある:
 - ▶ 話者数の限られた方言におけるアクセント辞書の欠如
 - > TTS モデルの学習に使える十分な品質の方言音声収録の困難性
- ◆そこで、以下の機能をもつ提案モデルによる方言音声合成を検討:
 - ▶ テキストからのアクセント潜在変数(ALV)予測
 - ▶ 入力参照音声からの ALV 自動抽出による合成音声の韻律制御



2 - Q - 32

2-Q-32 Emotion transfer with intensity control for text-to-speech synthesis based on self-supervised learning model

O Wei Li, Minematsu, Daisuke Saito (The University of Tokyo)

- In this research, we used self-supervised learning model, wav2vec 2.0, to extract acoustic features highly related to speech emotions for emotional text-to-speech task.
- ◆Relative attributes method was used to predict the emotion intensity values for controlling the strength of speech emotion.
- •We trained the speech synthesis model on emotional speech data (ESD) which contains 10 speakers and 5 emotional states (neutral, angry, happy, sad, surprised).
- We utilized several subjective and objective metrics to evaluate the performance.
- Compared with traditional reference audio encoders, wav2vec 2.0-based emotion transfer model has higher similarity between the reference speech and the synthesized speech.

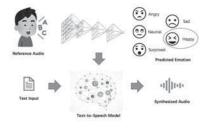


Fig.1: Overall structure of emotion transfer for text-to-speech synthesis

2 - Q - 34

2-Q-34 音声合成システムの入力表現に関する 分析的検討

Analytical Study on Input Representation of Speech Synthesis System

☆朴 浚鎔、齋藤 大輔、峯松 信明 (東大院・エ)

- ◆ 自己教師あり学習(Self-Supervised Learning, SSL)モデルを通じてラベルが指定されていない音声からの「入力表現」をより容易に得ることができ、このようなSSL モデルからラベルのない音声からの音声合成モデルと、このようなモデルがどの程度の性能を出せるかに対する研究も進行された。
- ◆本論文では、音声合成システムにおける音声特徴抽出のための方法について調査し、SSL により音声情報から得られる入力表現に関して分析し伝統的な入力表現と比較する。
 - 英語と日本語の合成音声からの指標を通じて結果を評価することにより、 上記の2つの入力表現間の比較研究を実行する。
 - ▶ 明瞭性と自然性に対する実験結果に加えて音響品質を追加で測定できる 指標を追加して評価する。

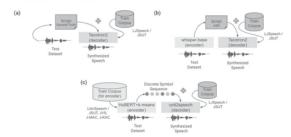


Fig. 1 Construction of speech synthesis system using (a) Ground-Truth label, (b) speech recognition (ASR) model, (c) self-supervised learning (SSL) model

2-Q-35 自己回帰及び非自己回帰モデルによる口唇動画を 用いた音声合成

Speech synthesis using lip video with autoregressive and non-autoregressive models.

☆南汰翼,藤田直明、鏑木時彦(九州大)

- ・ 喉頭癌等によって声帯を摘出した場合、これまで通りの発声が不可能になってしまう。これに対して本研究では、深層学習を用いて口唇の動画から直接音声を合成することによる、新たな代用音声の可能性を検討する。
- 口唇動画からの音声合成を検討した先行研究は、音声を自己回帰的に 推定する手法と、非自己回帰的に推定する手法の二つに分類できる。 しかし、これまで各手法の推定精度の比較検討は行われてきたものの、 推定方法の違いに着目した比較は行われていない。
- 時系列データである音声を推定する上で自己回帰的な推定は有効と考えられ、実際にテキスト音声合成ではその有効性が確認されている。しかし、動画からの音声合成では音声の話者性やイントネーションだけでなく、発話内容が十分理解できる程度の明瞭性を達成する必要がある。この難易度の高さから、自己回帰的な推定によって時間方向に誤差が累積してしまうことが、推定全体に悪影響を及ぼす可能性がある。
- 上述の影響を検討するため推定方法の違いに着目し、シンプルなモデル構造で精度を比較した結果、音声の明瞭性については自己回帰的な推定が有効である一方、ピッチの推定については非自己回帰的な推定が有効であることが示された。そのため、両者の弱みを補い合うようなモデルを構築することでさらなる精度の改善が可能かどうか、今後検討する必要がある。

2 - Q - 37

2-Q-37 単言語話者による 多言語音声合成モデルの構築と評価

Construction and evaluation of a multilingual speech synthesis model with monolingual speakers

☆古藤快都, 西村良太(徳島大), 太田健吾(阿南高専), 北岡教英(豊橋技科大)

- ◆本研究では、話者性を保持したまま、その話者の母語以外の言語で音 声を合成する多言語音声合成モデルの構築を行う。
- ◆提案モデルの構築には、ESPnet2のフレームワークを使用しEnd-to-End 型音声合成モデルVITS を用いた。
- ◆多言語音声合成モデルの実現のため、VITS に話者情報と言語情報の 埋め込みを使用した。
- ◆学習にはバイリンガル話者と単言語話者のデータを使用する. これに より、提案モデルが異なる言語間の関係を学習することが期待される.
- ◆提案モデルは、生成した音声の自然性と話者類似性に対して MOS と DMOS の指標を用いて主観的に評価された.

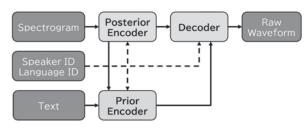


Fig. 1: Overall view of VITS using speaker ID and language ID.

2 - Q - 36

2-Q-36 テキスト音声合成における データサブセット選択のための指標検討

Study on data selection metrics for text-to-speech

☆関 健太郎, 高道 慎之介, 佐伯 高明, 猿渡 洋(東大院·情報理工)

- ◆TTS コーパスの大規模化
 - ▶ オーディオブックやダークデータを用いたコーパス構築
 - ▶ 多様なデータを用いることで TTS モデルの表現力向上に寄与
 - ▶ 実際のコーパスサイズは記憶媒体サイズ・学習時間によって制限 される:効率的なサブセットの抽出が重要
- ◆サブセット評価指標
 - ▶ 枠組み:データセット構築・選択の主要な方法を検討
 - ☆ エントロピー: バランスよく選択を行う
 - ◇ 多様性: 重複するデータを避けて選択を行う
 - ▶ 特徴量:モデルへの入力・モデルの出力で分類
 - ◆ 入力側:音素,話者 ID,言語特徵量,話者表現特徵量
 - ◇ 出力側: 自己教師あり学習モデルによる音響特徴量
- ◆どの枠組みにどの特徴量を用いるのが効果的か?
 - ▶ 古典的方法(音素エントロピー)との比較
 - ▶ 日本語・英語の多話者コーパスを用いた実験的検証



2-0-38

2-Q-38 自然言語による説明文に基づく キャラクタ性を考慮した音声合成

Character-based Text-to-Speech Based on Natural Language Descriptions.

☆根岸慧. 藤江真也(千葉工大)

- ◆自然言語の説明文により指示されたキャラクタ性に適した音声を出 力するテキスト音声合成システムについて検討する。
- ◆従来のテキスト音声合成は、学習データに存在しない話者の再現が困難なことや、使用したい音声を直感的に選択できない問題があった。
- ◆提案モデル (Fig. 1) は話者ベクトル推定器と複数話者テキスト音声 合成モデルで構成され、話者識別で用いられる特徴量の x-vector を用 いて複数話者の再現を行う。
- ◆話者ベクトル推定器は、自然言語で書かれた説明文を入力とし個人性 を指定するための x-vector を出力とするもので、BERT をファインチ ューニングすることで学習した。
- ◆「高い声の優しそうなお姉さん」といった説明文を入力することで、 学習データに含まれていない音声であっても適切に合成できること を確認した。

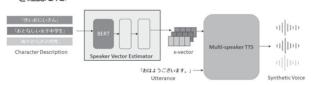


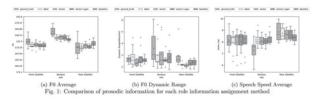
Fig.1:Proposed Model

2-Q-39 情報伝達のための音声合成における発話 文の役割情報付与手法の検討

A Study of Methods for Assigning Role Information in Speech Synthesis for Information Transfer

☆謝佳臻*, 藤江真也*†, 小林哲則*(*早大, †千葉工大)

- ◆ 役割情報を考慮した音声合成
 - ▶ 目的: 段落において重要な発話文を強調する話し方の実現
 - 役割情報の定義: 重要な発話文を核, その前後の発話文を前置き, 補足とする
- ◆ 検討する役割情報付与手法
 - ▶ 手法1:特殊トークンとして発話文に付与する
 - ▶ 手法2: Global Condition として中間層特徴量へ畳み込む
 - ▶ 手法3:手法2に加え、Conditional GAN を用いて生成された音 声に対し、役割情報が適切かどうかの判別を行う
- ◆ 実験: 役割情報に対する韻律情報の分析
 - ▶ 基本周波数の平均において役割情報を考慮する手法が有効
 - ▶ 基本周波数ダイナミックレンジにおいて Global Condition を用いる手法が比較的良好
 - ▶ 話速はどの手法でも差が見られなかった

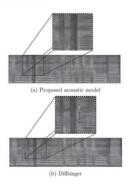


2 - Q - 41

2-Q-41 Preliminary Investigation of Poisson Flow Model for Singing Voice Synthesis

O ZHU Tingyuan (TokyoTech), OKAMOTO Takuma (NICT), SHINOZAKI Takahiro (TokyoTech)

- Singing voice synthesis aims to synthesize high-fidelity natural vocals for a given score. The current state-of-the-art SVS model based on the diffusion model faces the problem of slow synthesis speed and insufficient model robustness.
- ◆The Poisson flow model, which is inspired by electrostatic mechanics, has achieved faster generation speed than the diffusion model on image generation tasks.
- •We propose an acoustic model for the SVS task based on the Poisson flow model. The proposed model has a faster synthesis speed and higher model size tolerance than the state-of-the-art diffusion-based SVS model.



2 - Q - 40

2-Q-40 音声対話システムのテキスト音声合成に おける声質変換と x-vector 埋め込みを 用いた感情制御方式の検討

An Investigation of a Method to Control Emotion for Text-to-Speech in Spoken Dialogue Systems Using Voice Conversion and x-vector Embedding ☆小原俊一, 阿部匡伸, 原直(岡山大院・HS 統合科学研)

- ◆本稿では、声質変換とx-vector埋め込みを用いて、音声対話システム の入力音声に応じて、合成音声の感情とその強さを制御する方式を提 案する。
- ◆評価実験では、提案方式が感情を制御可能かの検証を目的とした感情 認識テストと、さらに感情の強さをも表現できるかを検証した一対比 較実験を実施した。
- ◆実験結果から、提案方式は喜びと悲しみの感情を制御し、その強さを も表現可能であることが示された。

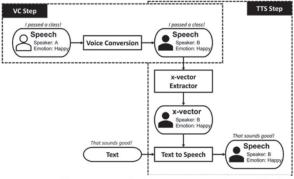


Fig.1:Outline of the proposed method. The proposed method consists of a VC Step and a TTS Step.

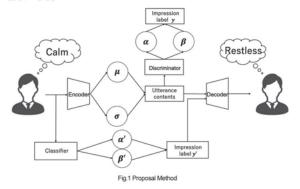
2 - Q - 42

2-Q-42 ベータ分布に基づく FaderNet を用いた音声 印象変換の性能評価

Performance evaluation of voice impression conversion using FaderNet based on beta distribution

☆釘本 咲, 中鹿 亘(電通大)

- ◆話者性はそのまま音声の印象のみを変換する印象変換を目的とする
- ◆ FaderNetVIC: 声質変換手法である FaderNetVC における話者ラベルを印象ラベルに置き換えることで印象の変換が可能
- ◆Classifier と Discriminator における印象ラベルの識別・推定にベータ 分布を用いることで印象変換精度は向上
- ◆本稿ではベータ分布に基づいた印象変換モデルについてコーパスの 違いによる印象変換精度、Decoderに挿入する印象ラベルの違いによる印象変換精度と話者性の保持性、モデル自体の話者再現性について 評価・検討



2-Q-43 Diff-SVC を用いたオペラ歌唱音声合成に おける中高域強調ネットワークの検討

A mid-high frequency enhancement network for opera-singing voice synthesis using Diff-SVC.

☆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 ('神戸大学, ²メック株式会社)

- ◆本研究では、声質変換手法を用いてオペラ歌唱未経験ユーザーの発話 音声からオペラ歌唱音声を合成可能なシステムの実現を目的とする。
- ◆マルチ受容野混合層(MRF)を用いて、2.2kHz~4.0kHz帯の中高音域を 強調するネットワークを導入することで、プロのオペラ歌唱音声の特 徴の付与を行い、合成音声のオペラ性の向上を検討する。
- ◆中高域強調ネットワークを導入することで、導入しない場合と比較 して話者性の低下が見られたがオペラ性が向上した。

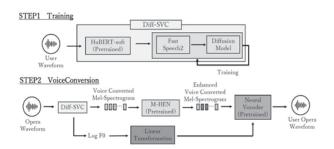


Fig.1 Training procedure using Diff-SVC with mid-high frequency enhancement network(M-HEN).

2 - Q - 45

2-Q-45

アバターに適した話し方および 声質への変換手法の検討

Investigation of conversion methods for speech style and voice characteristics suitable for avatars

△石川拓磨, ○實廣貴敏(愛知工科大)

- ◆VR 空間でのアバターを使うコミュニケーションの場合, アニメによ くいるような、そのキャラクターに合った独特な口調や声で話したい。

性能を評価する。「音声翻訳」とみなすことができる。

- ◆音声認識に Google Cloud Speech API, 機械翻訳に Transformer [Vaswani, et al. 2017], 音声合成に VITS [Kim, et al. 2021]を用いた。
- ◆アニメ番組の主人公「おじゃる 丸」を変換対象とした。「マロ は~でおじゃる」など、独特な ロ調がある。
- ◆機械翻訳の客観評価 BLEU では65.5, 提案法全体の主観評価では4.12 など, 比較的高い値を得ることができ, 可能性を示せた。

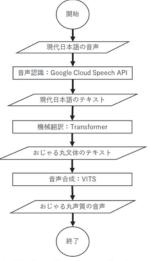


Fig.1: The flow of our proposed method.

2 - Q - 44

2-Q-44 StarGANv2-VC への 2D-1-2D 構造適用 による精度変化の検証

Verification of Accuracy Changes through the Application of 2D-1-2D Structure to StarGANv2-VC

☆杉本晃輔, 澤田隼, 大村英史, 桂田 浩一(東京理科大院·創域理工学研)

高精度に変換可能な音声変換モデル StarGANv2-VC を拡張し、感情音声への変換でも良好に動作するモデルを提案する. StarGANv2-VC では変換の際に学習可能なパラメータで分布変換を行う Adaptive Instance Normalization (AdalN) という手法を取り入れているが、2 次元特徴マップに対しての適用では感情音声のような複雑な分布への変換は苦手な傾向にある. そこで本研究では生成器の構造を 2D-1-2D 構造へ変更することで、AdalN を 1 次元特徴ベクトルに対して適用し、複雑な分布への変換でも良好な結果を残すことに成功した.

実験には Emotional Speech Dataset (ESD) を用いた. このデ ータセットのうち, 男性話者 4 名, 女性話者 2 名, 感情は Angry, Neutral, Happy を用いて感情音声に対する変換の精度 を評価する.変換精度の評価には声道特徴の近さを表す指標で ある Mel Cepstral Distortion (MCD) を, 生成された音声の明瞭 度の評価には Character Error Rate (CER) を用いた. MCD は StarGANv2-VC が 8.59, 提案モデルが 8.18 と StarGANv2-VC を上回っているものの良好とは言えない結果となった. これは StarGANv2-VC が間の開け方を保持しつつ声音のみを変換する という特徴を持つことから, 感情音声のように間の開け方もか なり変化するような音声と比較すると無視できない差異が生じ てしまうことが原因であると考えられる. また, CER について は StarGANv2-VC が 18.62, 提案モデルが 12.84 とこちらも StarGANv2-VC を上回る結果となった. これらの結果から, 2D-1-2D 構造の導入により StarGANv2-VC が感情音声への変換 において良好に動作することを確認できた.

2 - Q - 46

2-Q-46

リアルタイム声質変換のための 骨伝導を介した自己発声音の アクティブキャンセリング

Active cancellation of self-vocalization sounds via bone conduction for real-time voice quality conversion.

☆長谷川郁弥、橋本佳、南角吉彦、徳田恵一、田中由浩、加藤昇平 (名工大) 本稿では、リアルタイム声質変換において所望の聴覚フィードバック をユーザに与えるための、骨伝導イヤフォンによる自己発声骨導音の キャンセリング手法を提案する。遅延・変換を含む聴覚フィードバック の影響を緩和するだけでなく、聴覚フィードバックによって積極的にユ ーザに働きかけ、より話しやすい歌いやすい、さらには歌唱の上達など も視野に入れた聴覚フィードバックの生成技術の開発を目指す。

提案システムの実現可能性を検証するため、主観評価実験によって、 自発声を模擬した骨伝導スピーカーから生成した正弦波が位相シフト した骨導音によって低減可能であることを示した。

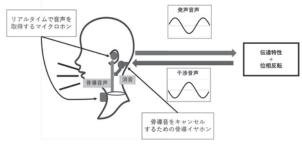


Fig. 1: Proposal method

2-Q-47

2-Q-47 End-to-End 系列変換型声質変換への高速 ニューラル波形生成モデル導入の検討

Exploring the Introduction of High-Speed Neural Waveform Generation Models to End-to-End Sequence-to-Sequence Voice Conversion

- ◎山下陽生(神戸大/NICT), 岡本拓磨(NICT), 高島遼一(神戸大), 大谷大和 (NICT), 滝口哲也(神戸大), 戸田智基(名大), 河井恒(NICT)
- ◆End-to-End で学習可能な系列長変換型声質変換モデルである JETS-VC の品質を保ちながらの高速化のために、高速ニューラル波 形生成モデルである MS-iSTFT-HiFiGAN と MS-FC-HiFiGAN の導入 を検討する。
- ◆また、reduction factor の値の検討と、Encoder/Decoder に E-Branchformer を導入することによる品質向上も確認した
- ◆高速ニューラル波形生成モデルを用いたところ、MS-FC-HiFiGAN は reduction factor=2,3 のときに品質を保っており、約4倍の高速化も達成した。
- ◆また、E-Branchformer を用いることで、推論速度は落ちるものの、 さらなる高品質化を達成した。

Table1.	Result	of	Subjective	evaluation
---------	--------	----	------------	------------

encoder/decoder block			drf	$female \rightarrow male$		$male \rightarrow female$	
	vocoder	erf		MOS	similarity[%]	MOS	similarity[%]
Transformer	HFG	2	1	3.83 ± 0.31	73.3	3.57 ± 0.37	56.7
		3	1	3.91 ± 0.27	83.3	3.77 ± 0.30	53.3
		4	1	3.80 ± 0.30	66.7	3.63 ± 0.30	70.0
		4	2	3.51 ± 0.37	63.3	3.26 ± 0.40	53.3
	MS-iSTFT	2	1	3.89 ± 0.36	73.3	3.43 ± 0.45	63.3
		3	1	3.11 ± 0.38	70.0	3.51 ± 0.43	50.0
		4	1	3.74 ± 0.39	46.7	4.03 ± 0.32	80.0
	MS-FC	2	1	3.74 ± 0.37	70.0	3.83 ± 0.32	80.0
		3	1	3.74 ± 0.33	63.3	3.71 ± 0.34	73.3
		4	1	3.11 ± 0.37	63.3	3.66 ± 0.36	63.3
E-Branchformer	HFG	2	1	4.09 ± 0.29	75.9	3.54 ± 0.44	50.0
		3	1	3.80 ± 0.38	73.3	4.20 ± 0.30	73.3
original	-	-	-	4.14 ± 0.39	-	4.66 ± 0.19	-

2 - Q - 49

2-Q-49 深層学習モデルを用いた話者年齢埋め込 みベクトルに関する分析

Analysis of deep learning based speaker age embedded vector
☆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)
 - ➤ 話者年齢を推定するように NN を学習
 - ▶ 学習したNN を用いて発話音声から Age-vector を抽出
- ◆Age-vector の解析的な分析
 - ▶ 2歳から59歳までの実環境発話データ2,302個を対象
 - ➤ 512 次元固定長ベクトルの Age-vector を主成分分析で次元削減
 - ◆ 180 次元で累積寄与率が99.0%となることを確認
 - ➤ 次元削減後の Age-vector を 7 つの年齢層クラスに分割
 - ♦ 寄与率が確保されている 180 次元までの平均で重心を算出
 - > 各年齢層の重心間距離を比較
 - ◆ 各年齢層のベクトルの重心間距離をユークリッド距離で比較
 - ▶ 男性が12歳以下と13歳以上、女性が8歳以下と9歳以上を境に 重心間距離が変化
 - ▶ 女性は年齢層の変化に伴うベクトル重心の変化が小さい
- ◆Age-vector の各年齢層の重心間距離(男性・女性)(Table 1)
 Table 1: Vector distance between age groups (male/female). Color indicates similarity.



2 - Q - 48

2-Q-48 表現力の異なる話者埋め込み空間と 主観的話者間類似度の比較

Comparison of inter-speaker similarity metrics between different speaker embedding spaces and human subjectivity.

☆森田湧大, 齋藤大輔, 峯松信明(東大·工)

Keywords: 話者埋め込み・音声合成・人間の知覚・Triplet Loss

- ◆複数話者音声合成の研究において、「『話者埋め込み特徴量を近づける こと』と『合成音声の話者性を近づけること』がどの程度相等しているか」については議論が少ない。
- ◆本発表では、複数話者音声合成研究への応用を想定し、以下の二つの 対応関係について分析を試みた結果を報告する。
 - ▶ 埋め込み特徴量に基づく話者間の類似度
 - ▶ 人間の知覚に基づく話者間の主観的類似度

