

1-1-1

1-1-1 複数台の Kinect を用いた 音声収録時の頭部運動測定

Head movement measurement during voice recording
using multiple Kinect units.

☆倉本有紗, 矢田部浩平, 及川靖広 (早大理工)

- ◆これまで1台の Azure Kinect を用いて、音声収録時における発声者の頭部運動測定を行うシステムが構築されてきた。
- ◆発声者の身体を大きく回転させた場合や、発声者の位置が Kinect から大きく離れた場合では、適切に測定することができなかった。
- ◆本研究では、2台の Kinect を用いて、それぞれから関節情報を取得し、球面線形補間を用いてそれらの関節情報を1つに統合した。
- ◆2台の Kinect を用いる本手法と、1台の Kinect を用いる従来手法で、ポディトラッキングを行うことが可能な範囲がどれほど変化するかを計測し、測定可能な範囲が広がったことを確認した (Fig. 1).

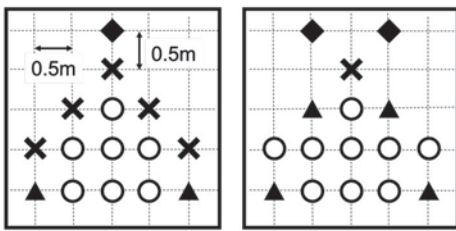


Fig.1: Comparison of the range where body tracking is possible.
In the left figure, one Kinect is used.
In the right figure, two Kinects are used.
(◆: Kinect, ○: possible, ▲: freezing, ×: impossible)

1-1-3

1-1-3 ガス管メンテナンスを目的とした音響信号 駆動磁場の可視化と磁場発生源位置の推定

Visualization of acoustic signal driven magnetic field
and estimation of magnetic field source location
for gas pipe maintenance.

☆黒沢琢登, 及川靖広 (早大理工)

浅川正俊, 前田亮, 佐土原奨, 高村知広, 小野寺文也 (東京ガス(株))

- ガス管内メンテナンス箇所への推定
 - メンテナンス箇所で音響信号に対応した磁場を生成
 - 作業員が壁越しにコイルで受信し音の強弱から推定
- 提案手法
 - 音の強弱から磁場分布を可視化
 - ◆ iPad Pro を用いて AR 技術により現実世界に重畳して表示
 - 分布の対称性から磁場発生源位置を自動推定
- 実験結果
 - アンテナと受信コイルの位置関係 (Fig. 1) により異なる分布を確認 (Fig. 2 上)
 - 分布によらず磁場発生源位置を特定可能 (Fig. 2 下)

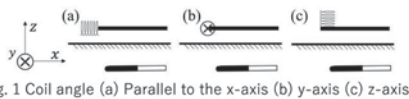


Fig. 1 Coil angle (a) Parallel to the x-axis (b) y-axis (c) z-axis

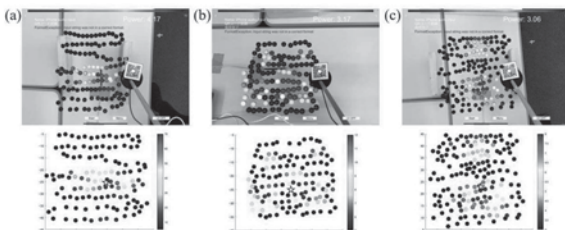


Fig. 2 Visualization and estimation of magnetic field source location
(a) Parallel to the x-axis (b) y-axis (c) z-axis

1-1-2

1-1-2 音響レンズの音声認識への応用と評価

Application of acoustic lenses
for speech recognition and evaluation of speech recognition ratio.

☆村上龍斗(産技高専), 高橋義典(工学院大学), △高野邦彦(産技高専)

- ◆筆者等は、小林等によって提案された距離選択型音響レンズの最適收音距離を短くすることで、話者の背後にある雑音源を抑え、音声認識への応用を試みた。
- ◆最適收音距離を 0.1 m として設計した音響レンズを試作し、背景雑音として白色雑音および、あらかじめ録音した環境雑音を再生している状況で、音源とマイクロホンの距離に対する音声認識率を調べた。

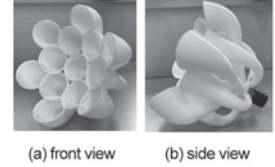
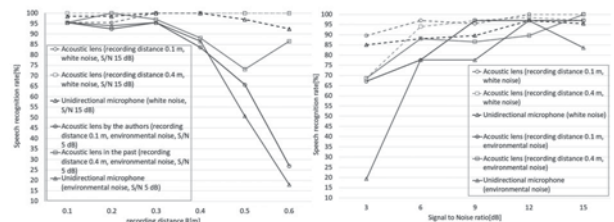


Fig.1 Acoustic lens manufactured by 3Dprinter (recording distance 0.1 m).



(a) Trend of speech recognition rate according to recording distance
(b) Speech recognition rate with change in Signal to Noise ratio.

Fig.2 Evaluation of speech recognition rate.

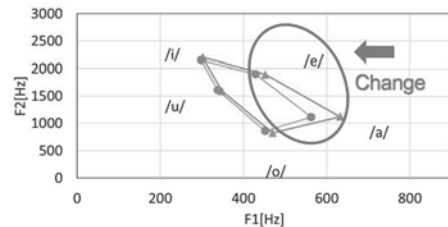
1-1-4

1-1-4 マスク着用時の発声変化表現 パラメータの検討*

The research of the feature parameters expressing the change of the Japanese vowel pronunciation by influence of the face mask

○秋田昌憲, 工藤悠夏, 大坪心椰, 緑川洋一(大分大)

- ◆新型コロナウイルスの蔓延によりその着用が強制されているマスクの発音に与える影響を調査している。
- ◆先の 2021 年春季研究発表会においての報告では、不織布マスク着用の場合下図のように第一・第二フォルマント周波数による母音フォルマント五角形の面積の縮小傾向がマスク着用時に見られることを示した。



◆ Not wearing the surgical mask ◆ Wearing the surgical mask

Fig.1 Example illustrating of the distribution of Japanese 5 vowels in case that the examinee wear the surgical mask (Speaker D)

- ◆ここでは、特定母音のフォルマント座標移動を変化特徴パラメータとして定義し、不織布マスク着用の不快感と発生時の声道形状変化の特徴について検討する。

1-1-5

1-1-5 偏光高速度干渉計を用いた口笛演奏音の収録

Recording of whistling sounds using parallel phase-shifting interferometry.
☆野澤澤, 及川靖広 (早大理工)

- ◆偏光高速度干渉計によって口笛演奏音の伝搬の可視化するとともに、それに基づく収録を行った。
- ◆偏光高速度干渉計による計測結果に対して、フィルタバンク用いて口笛発生音の抽出を行い、それらを合成し収録音とした。
- ◆マイク収録による結果と比較し、第3倍音までの口笛演奏音が収録されていることを確認した。

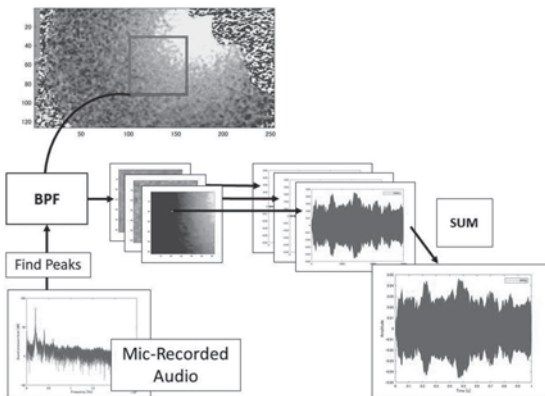


Fig.1 Outline of recording using high-speed polarized Interferometer.

1-1-7

1-1-7 ハンガリー法と欠損帯域補完に基づく周波数領域ブロックパーミュテーション解決法

Frequency-domain block permutation method using Hungarian algorithm and missing component restoration
©李莉(NTT, 名大), 亀岡弘和(NTT), 関翔悟(NTT)

- ◆周波数領域で定式化されるブラインド音源分離(BSS)のアプローチは効率の良いアルゴリズムを実現できる利点がある一方で、周波数ごとに得られた分離信号の順番に任意性があるというパーミュテーション問題が存在する。
- ◆音源の周波数間の成分の依存関係をモデル化し、BSSの最適化問題に制約やコストの形で取り入れることでパーミュテーション問題と音源分離を同時に解決する手法が多数提案されており、その効果が示されている。しかし、これらの手法では、帯域ブロックごとにパーミュテーション不整合がしばしば生じることが報告されている。
- ◆本稿では、帯域マスキングと欠損帯域補完による音源スペクトログラム推定値の修正処理と割当問題の求解法であるハンガリー法を用いたブロックパーミュテーション問題の解決法(HBP)を提案し、その有効性を実験により示す。

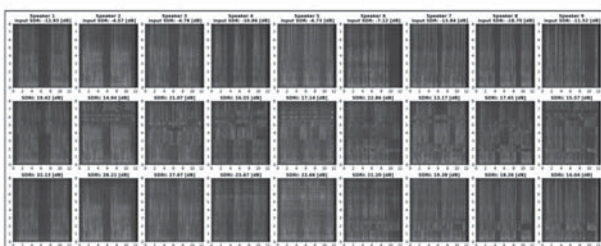


図1. 9音源の分離例. 上からは正解信号, HBPを用いないFastMVAE2による分離信号, HBPを用いたFastMVAE2による分離信号のスペクトログラム.

1-1-6

1-1-6 ChimeraACVAEによる高速多チャンネル変分自己符号化器法

Faster multichannel variational autoencoder method with ChimeraACVAE
©李莉(NTT, 名大), 亀岡弘和(NTT), 牧野昭二(筑波大)

- ◆多チャンネル変分自己符号化器法(MVAE)は、各音源のスペクトログラムの生成過程を変分自己符号化器(CVAE)を用いてモデル化した混合信号のモデルであり、これを用いることで収束性が保証された反復アルゴリズムにより高精度な音源分離を実現できる一方で、高い計算コストを要していた。
- ◆MVAEの計算コストの削減を目的としたFastMVAE法では、クラス識別器つきVAE(ACVAE)を音源モデルとして用いることにより高速な分離をできた一方で、テスト時において学習時と条件が一致しない場合に分離性能が低下する傾向があった。
- ◆本稿では、知識蒸留を用いた新たなACVAE音源モデル(ChimeraACVAE)の学習法とモデル構造を提案する。
- ◆18音源の未知話者の分離実験により、提案手法はILRMAより高速に分離を行えることと、ILRMAとFastMVAEより高い分離精度が得られることを確認できた。

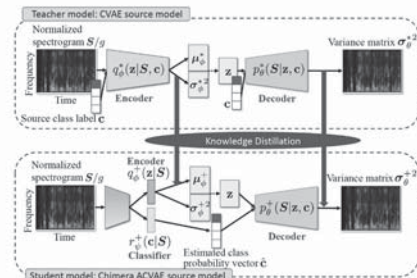


図1. ChimeraACVAEの学習概念図

1-1-8

1-1-8 深層フルランク空間相関分析に基づくブラインド音源分離

Amortized inference of neural blind source separation

©坂東 宜昭(産総研 / 理研 AIP), 関口 航平, Aditya Arie Nugraha, Mathieu Fontaine(理研 AIP), 吉井 和佳(京大 / 理研 AIP)

- ◆目的: 高精度かつ安定して動作するブラインド音源分離(BSS)
 - 多くのBSSは単一の混合音に対し推論 → 局所解に陥りやすい
- ◆従来法: 大量の混合音を事前に準備して、大域的な特徴を学習
 - 従来のBSSの分離結果を模倣するDNNを学習
 - DNNは訓練データの平均的な結果を記憶するので安定に分離
 - しかし線形モデルに基づくBSSが教師となるので性能に制限
- ◆提案手法: 深層音源モデルに基づくBSSを教師なし学習
 - 混合音の深層生成モデルとその推論モデルを同時学習
 - 非線形音源モデルにより精緻に音源を表現可能
 - 一度学習が済めば未知の混合音を高速に分離可能に
- ◆実験結果
 - 従来のBSS/教師なし多チャンネル深層分離を上回るSDRを達成

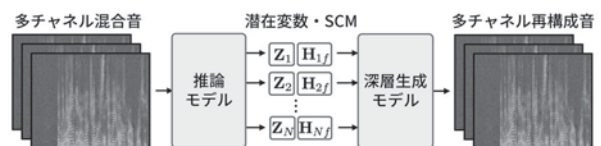


図1: 多チャンネル混合音のみから音源分離を教師なし学習する深層フルランク空間相関分析の概要

1-1-9

1-1-9 因果的MPDRビームフォーマの近接分離最適化による設計

Causal minimum power distortionless response beamforming via proximal splitting algorithm

© 升山義紀, 山岡洗瑛, 木下裕磨, 小野順貴 (都立大)

- ◆ 目的: 低遅延なビームフォーミング
 - 時間周波数 (T-F) 領域でのビームフォーミングでは、短時間フーリエ変換の窓幅に依存したアルゴリズム遅延が発生
- ◆ 従来手法: 空間フィルタを時間領域へ変換し、非因果成分を除去
 - T-F 領域で設計したフィルタの非因果成分を除去し、時間領域で混合音に畳み込むことでアルゴリズム遅延を削減
 - **非因果成分を除去したフィルタの特性を保証できないのが課題**
- ◆ 提案手法: **無歪み性と因果性を同時に満たす空間フィルタの設計**
 - 最小パワー無歪応答 (MPDR) ビームフォーマの設計問題に、空間フィルタの因果性を制約として追加
 - 交互方向乗数法 (ADMM) により、周波数またはマイクごとの小規模な部分問題を反復的に解くことで効率的に空間フィルタを最適化
 - 各制約の緩和を提案し、分離性能を改善 (Prop-relax in Fig1.)

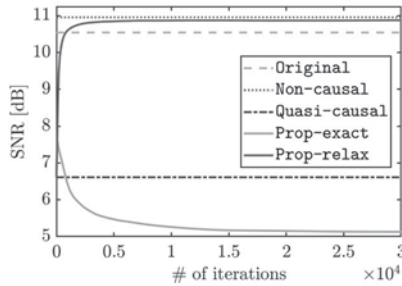


Fig.1 Signal-to-noise ratio (SNR) of extracted signals averaged over 10 mixtures.

1-1-11

1-1-11 ヴァーチャルマイクロフォンの内挿における位相及び振幅補間の音声強調性能への影響の評価

Evaluation of the effect of phase and amplitude interpolation on speech enhancement performance in virtual microphone interpolation

☆ 瀬川華子 (筑波大), 李莉 (NTT), 牧野昭二 (筑波大・早稲田大), 山田武志 (筑波大)

- ◆ 劣決定条件における音声強調性能の改善に有効である、従来のヴァーチャルマイクロフォン技術による位相および振幅補間法では 2 マイクの信号を用いて補間を行う場合についてのみ検討が行われている。
- ◆ 本稿では、2 つ以上の実マイクロフォン信号を用いたヴァーチャルマイクロフォン信号の振幅と位相の補間法を提案した。また、補間の音声強調性能への影響を実振幅、実位相との比較により調べた。
- ◆ 提案法によって、任意のマイク数への拡張が容易に行えるようになり、実マイクロフォンの選択順による性能変動を無くすことができた。

	method	SDR Imp	SIR	SAR
simu	proposed	6.74	6.46	5.00
	prev((1,2),3)	6.73	6.44	5.01
	prev((1,3),2)	6.73	6.43	5.01
	prev((2,3),1)	6.73	6.43	5.01
real	proposed	11.71	14.58	8.24
	prev((1,2),3)	11.86	14.96	8.30
	prev((1,3),2)	11.51	14.29	8.04
	prev((2,3),1)	11.86	14.98	8.30

Table. 1 Proposed method and previous method in simulation

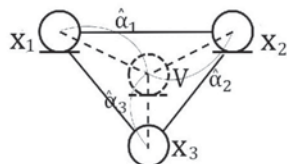


Fig. 1 Arrangement of 3 real and virtual microphones

1-1-10

1-1-10 多変量一般化 Gauss 分布に基づくランク制約付き空間共分散行列推定法における雑音欠落ランク空間基底推定

Deficient basis estimation of noise spatial covariance matrix for rank-constrained spatial covariance matrix estimation based on multivariate generalized Gaussian distribution

☆ 近藤祐斗, 久保優騎, 高宗典玄 (東大), 北村大地 (香川高専), 猿渡洋 (東大)

- ◆ ランク制約付き空間共分散行列推定法 (RCSCME) [Kubot, 2019] は、拡散性雑音中の方向性目的音源をブラインドに抽出するための、前段処理で求めたランクが欠如した雑音空間相関行列の補完に基づいた音源分離手法である。
- ◆ 従来の補完される基底も最適化する RCSCME [近藤+, 2021] に対し、本稿では生成モデルを多変量 Gauss 分布から多変量一般化 Gauss 分布に拡張し、majorization-minimization 及び majorization-equalization アルゴリズムに基づく更新則を導出し、理論的な性質を示す。
- ◆ 実験により、提案法が従来法より高い分離性能を示すことを確認する。Table 1: SDR improvements [dB] for each method and noise case. Each term represents "peak score / score after 200 iterations." ρ is the shape parameter in generalized Gaussian distribution.

Methods		Babble	Station	Traffic	Cafe
ILRMA		6.1 / -	6.2 / -	4.7 / -	6.4 / -
Fixed-basis MM	$\rho = 0.5$	8.6 / 7.8	10.2 / 9.6	8.2 / 7.8	10.0 / 9.3
	$\rho = 1$	8.5 / 7.7	9.7 / 9.3	7.8 / 7.6	9.7 / 9.2
	$\rho = 2$	8.4 / 7.6	9.7 / 9.1	7.8 / 7.5	9.7 / 9.0
Fixed-basis ME	$\rho = 0.5$	8.7 / 7.7	10.2 / 9.5	8.3 / 7.8	10.1 / 9.3
	$\rho = 1$	8.6 / 7.7	9.9 / 9.4	8.0 / 7.7	9.9 / 9.2
	$\rho = 2$	8.5 / 7.6	9.8 / 9.2	7.9 / 7.5	9.8 / 9.1
Proposed basis-optimizing MM	$\rho = 0.5$	9.3 / 8.8	11.4 / 11.0	9.4 / 9.2	11.0 / 10.5
	$\rho = 1$	9.0 / 8.6	11.2 / 10.8	9.1 / 9.0	10.9 / 10.4
	$\rho = 2$	8.9 / 8.6	10.9 / 10.7	8.9 / 8.9	10.7 / 10.4
Proposed basis-optimizing ME	$\rho = 0.5$	9.3 / 8.8	11.5 / 11.0	9.4 / 9.2	11.1 / 10.5
	$\rho = 1$	9.1 / 8.6	11.2 / 10.8	9.1 / 9.0	10.9 / 10.4
	$\rho = 2$	9.0 / 8.6	11.1 / 10.7	9.0 / 8.9	10.8 / 10.4

1-1-12

1-1-12 RealSense を用いた GCC-PHAT に基づく移動音源追跡精度評価

Output level calibration for bone conductive earphone

☆ 狩野拓充†, 高橋徹†, △ 赤塚俊洋†,

江川琢真†, 中山雅人† (†: 阪産大, ‡: 阪産大院)

- ◆ 移動音源の追跡精度を評価する枠組みを実装
 - 音源位置の真値を LiDAR (Light Detection and Ranging)
- ◆ 二組のマイクロホンアレイによる GCC-PHAT ベースの音源定位
 - 1m 間隔に配置し、三角測量に基づく音源定位
- ◆ 3m 以下の領域で、移動音源を誤差 0.1m 程度で追跡可能

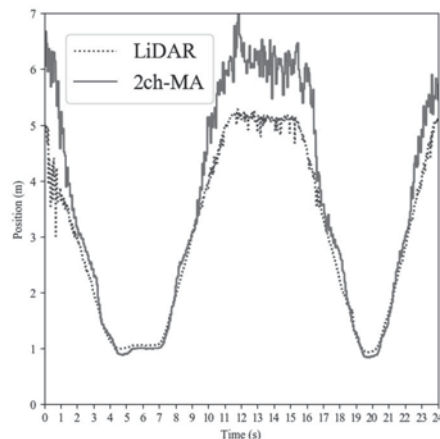


Fig. 1: Moving sound localization (Red solid line: two microphone-arrays based localization, Blue dotted line: LiDAR based localization).

1-1-13

1-1-13 ブラインド音源分離に基づく複数音源方向推定を用いた分離音源選択の検討

Selection of Separated Source Using Multiple Source Direction Estimation Based on Blind Source Separation

☆茂木 倅大, 中嶋 大志, 若林 佑幸, 小野 順貴 (都立大)

- ◆ブラインド音源分離 (blind source separation; BSS) は複数の音源が混合した観測信号のみを用いて、混合される前の音源信号を推定する技術である。
- ◆BSS を選択的聴取機能の実現や補聴器の開発などへ応用するためには音源の分離とあわせて、分離された複数音源から所望の音源をユーザが選択する必要がある。
- ◆本稿では、複数音源に対し、ユーザが指定した方向にある音源を選択することを想定し、音源方向を用いた音源選択について評価を行う。
- ◆実験の結果、表1のように、音源マイク間の距離・音源間の角度の範囲に依らず安定して、8割を超える精度でユーザ所望方向の音源の選択ができていることが確認できた。

Table 1 Correct answer rate of selection (proposed method/ideal sound image).

$\Delta\theta$	$r = 1.0\text{ m}$	$r = 2.0\text{ m}$	$r = 3.0\text{ m}$
$10^\circ \leq \Delta\theta < 20^\circ$	92/98	89/100	98/100
$20^\circ \leq \Delta\theta < 30^\circ$	95/100	93/100	100/100
$30^\circ \leq \Delta\theta < 40^\circ$	82/100	81/100	98/100

1-1-15

1-1-15 サンプリング周波数非依存音源分離モデルを用いた楽音分離の実験的評価

Music source separation experimental evaluation of sampling-frequency-independent audio source separation model

☆齋藤弘一, 中村友彦(東大), 矢田部浩平(早稲田大), 猿渡洋(東大)

- ◆我々はこれまでに、サンプリング周波数に非依存な音源分離モデル (SFI音源分離モデル) を提案してきた [Saito+, 2021]。
- ◆SFI音源分離モデルは、入力信号のサンプリング周波数に応じて畳み込み層の重みを生成することで、未学習のサンプリング周波数の信号に対しても動作する。
- ◆本稿では、SFI音源分離モデルとConv-TasNet [Luo+, 2019], 入力信号を学習データのサンプリング周波数へリサンプリングする手法 (SFI SR) に関し、分離性能を実験的に比較した。
- ◆また、これらの手法の聴感上の分離品質に関しても主観評価実験により比較を行った。

— Conv-TasNet — SFI SR — SFI Audio Source Separation Model

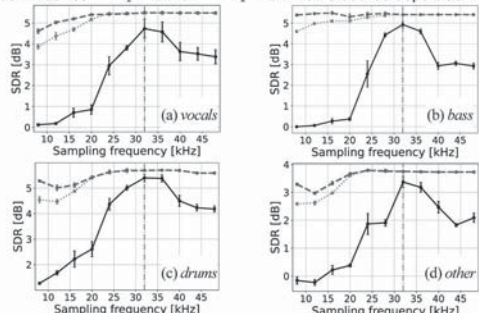


Fig.1: Signal-to-distortion ratios (SDRs) of comparison methods at various sampling frequencies. Dashed-dotted line shows trained sampling frequency.

1-1-14

1-1-14 深層学習に基づく間引きインジケータ付き周波数帯域補間手法による音源分離処理の高速化

Fast audio source separation based on deep-neural-network-based frequency component interpolation with decimation indicator

☆渡辺瑠伊, 北村大地 (香川高専), 中村友彦, 猿渡洋 (東大), 高橋祐, 近藤多伸 (ヤマハ)

先行研究[1]では多チャンネル音源分離 (MASS) 及び深層学習によって、音源分離処理を高速化するフレームワークを提案した。このフレームワークによって、音源分離の高速化は実現されたが、周波数ビンの間引き方によってその分離性能が変化した。本稿では、混合信号のパワーが大きい周波数帯域を MASS に適用するような分割手法を提案した。実験から、従来の分割手法と比較し音源分離性能が向上したことを確認した。

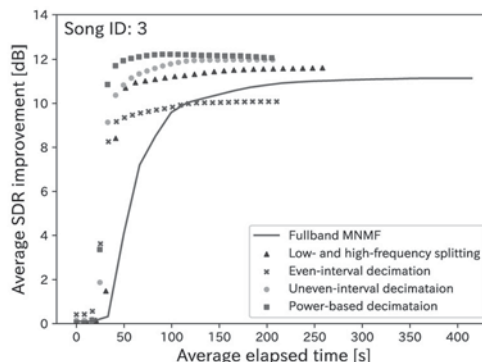


Fig. 1: Example of average SDR improvement and elapsed time for each method.

[1] 渡辺瑠伊, 北村大地, 猿渡洋, 高橋祐, 近藤多伸, “深層学習に基づく周波数帯域補間手法による音源分離処理の高速化” 日本音響学会 2021 年春季研究発表会講演論文集, pp. 213-216, 2021.

1-1-16

1-1-16 Product of Priors 型確率分布を導入した音源モデルに基づく独立深層学習行列分析による多チャンネル音源分離

Multichannel Audio Source Separation by Independent Deeply Learned Matrix Analysis Based on Source Model with Product-of-Prior-Type Probability Distribution

◎蓮実拓也, 中村友彦, 高宗典玄, 猿渡洋(東大) 北村大地(香川高専), 高橋祐, 近藤多伸(ヤマハ)

- ◆多チャンネル音源分離手法の独立深層学習行列分析 (IDLMA) は深層ニューラルネットワーク (DNN) の事前学習を要するため、学習時と音色の異なるデータを分離する際には分離性能が低下する。
- ◆本稿では音源の分散を潜在変数とし、それに対する事前分布として教師なし NMF と事前学習済み DNN 出力による確率分布の積で表現される product of priors (PoP) 型分布を用いた音源モデルを提案する。
- ◆PoP を導入した IDLMA (PoP-IDLMA) を構築し、majorization-minimization アルゴリズムを用いた分離アルゴリズムを導出する。
- ◆楽音分離実験により分離性能を評価し、提案法の有効性を確認した。

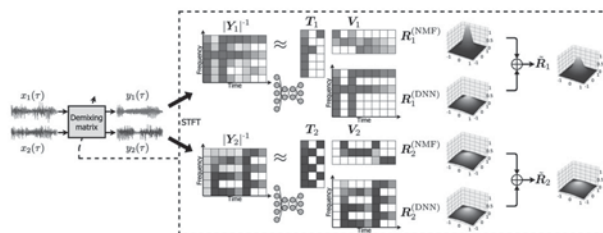


Fig.1: Separation process of PoP-IDLMA.

1-2-1

1-2-1 空間能動騒音制御のためのノイズ・スピーカ音場の個別補間に基づく適応フィルタアルゴリズム

Adaptive filtering algorithm based on individual interpolation of primary and secondary sound fields for spatial active noise control

☆有川和志, 小山翔一, 猿渡洋(東京大学 大学院情報理工学系研究科)

- ◆カーネル補間に基づく空間ANCの従来法では、(誤差マイクロフォン信号)=(ノイズと二次音源の合成音場)に対して直接補間を行っていた。
- ◆本研究では、誤差マイクロフォンの観測信号をノイズ成分と二次音源成分に分解した上で、それぞれの成分に対して個別にカーネル補間を行うというアプローチを提案する。
- ◆個別カーネル補間に基づく NLMS アルゴリズムを新たに導出し、数値実験によって従来法とANCの精度比較を行う。

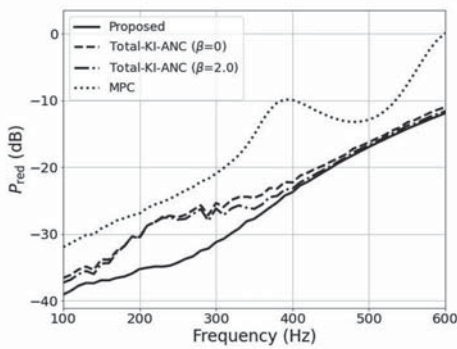


Fig. 1 Regional noise power reduction inside the target region with respect to noise frequency.

1-2-3

1-2-3 マルチゾーン音場合成のための差分ノルム正則化を用いた多点振幅制御法

Amplitude matching with differential norm for multizone sound field synthesis

☆阿部拓実, 小山翔一(東京大学),

植野夏樹(東京都立大学), 猿渡洋(東京大学)

- ◆マルチゾーン音場合成は、複数のスピーカを用いて、複数の分割した対象領域内にそれぞれ異なる音場を合成することを目的とする。
- ◆本研究では、対象領域を離散化し、各離散点での振幅が所望の振幅と一致するようにスピーカの駆動信号を求める**多点振幅制御**を扱った。
- ◆広帯域での多点振幅制御において、フィルタ長が過度に長くなることを抑制する**差分ノルム正則化**を提案した。
- ◆多点振幅制御の最適化問題を効率的に解くために、**交互方向乗数法に基づくアルゴリズム**を提案した。
- ◆実環境データを用いた実験により、提案手法を用いることで、従来手法よりも高い精度で音場が合成されることを確認した。

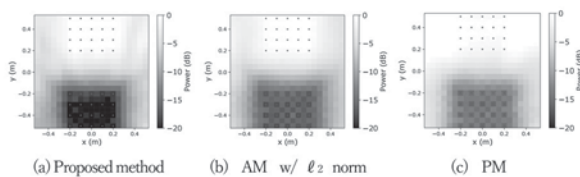


Fig. 1. Power distributions of synthesized sound fields by each method.

Desired amplitude was 1 in the upper region and 0 in the lower region.

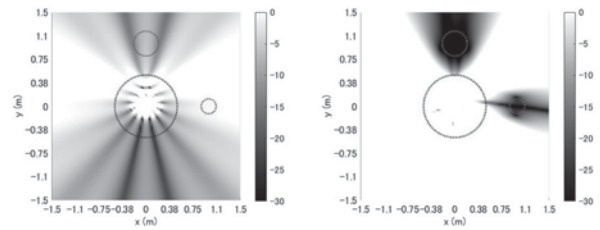
1-2-2

1-2-2 円筒関数の加法定理を用いた2次元外側局所音場再現法

A 2D Local Exterior Sound Field Reproduction Method Based on an Addition Theorem of Cylindrical Functions

☆任逸, 羽田陽一(電通大)

- ◆高臨場感再生技術のひとつである音場再現の研究の多くは、閉じた形状のスピーカアレイを用い、アレイの内側または外側の音場を制御対象とする。外側の音場を制御対象とする手法では、再現精度を確保することが難しい。
- ◆本研究では、外側音場を対象とした音場再現手法において、同じく制御範囲を犠牲にし、ローカルエリアを設定する。円筒関数の加法定理を用い、外部音場を小さな局所音場に分割し、限られた局所音場に高い精度で音場を再現する手法を提案する。
- ◆計算機シミュレーションおよび無響室での実験により、ローカルエリアにおいて、提案手法は従来手法より高い精度で音場を再現できることがわかった。



従来手法による再現誤差

提案手法による再現誤差

1-2-4

1-2-4 水平面配置スピーカを用いたパラメトリック頭部伝達関数に基づく上方音像制御の検討

Upper sound image control with horizontal-arranged loudspeakers based on parametric head-related transfer functions.

☆三浦峻平(立命館大院), 岩居健太, 西浦敬信(立命館大), 添田喜治(産総研)

22.2 マルチチャンネル音響は立体的なスピーカ配置によって立体的な音像を再現できるが、設置コストが高いという問題がある。一方で、スピーカを水平面のみ配置した音響システムは、立体配置の音響システムと比較して設置コストが低いが、上方音像の再現が困難という問題がある。そこで本研究では、パラメトリック頭部伝達関数に基づいた水平面配置スピーカによる上方音像制御を提案する。提案手法(Fig 1)では、正中面のパラメトリック頭部伝達関数に基づき、仰角に音像制御されたバイノーラル信号を生成する。次に、両耳間レベル差に基づき、バイノーラル信号の音像を方位角に制御する。最後に、Multi-input/output Inverse Theorem (MINT)に基づいて設計された多チャンネル逆システムを用いることにより、両耳信号から水平面に配置されたスピーカの出力信号を生成する。評価実験の結果、提案手法はバイノーラル再生と同等の精度で仰角に対する音像を制御できることが示された(Fig 2)。

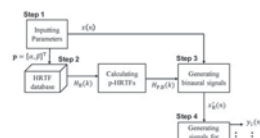


Fig 1. Overview of the proposed method

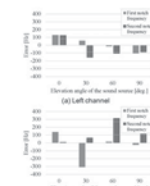


Fig 2. Notch frequencies errors between the binaural sounds and the evaluation sounds

1-2-5

1-2-5 音場分離と内部外部同時制御に基づくマルチ音場制御

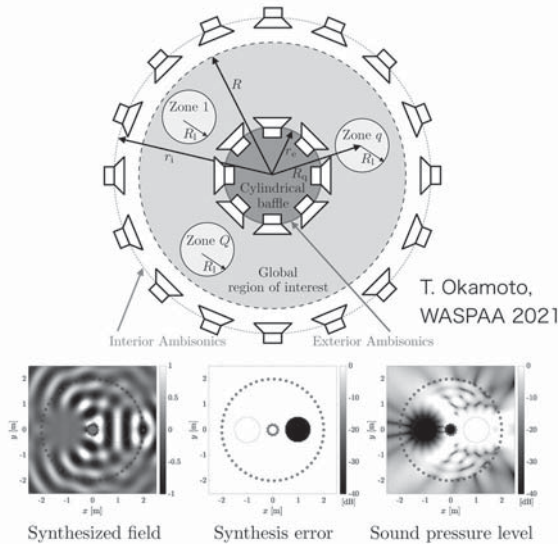
Multi-zone sound field synthesis based on sound field separation and interior-exterior control

○岡本拓磨 (NICT)

Q: 「マルチ音場制御におけるオクルージョン問題を低減したいのですが?」

オカモト: 「それならこれ!!」

1. 複数のローカル音場をグローバルな内部外部音場に変換
2. 内部外部同時モード制御により再生



1-2-7

講演取消

1-2-6

1-2-6 任意の指向性を有する移動音源によって形成される波面の合成

Synthesis of the sound field generated by a moving sound source with an arbitrary directivity.

◎佐々木陽, 松井健太郎, 中山靖茂 (NHK)

◆任意の球面調和スペクトルで規定された放射特性を有する音源が任意の軌道で移動・回転する音場の球面調和領域表現を導出した。

$$P(r, \omega) = \begin{cases} \sum_{\nu, \mu} \sum_{l, n, m} \hat{G}_{\nu, \mu}^{l, n, m} \int D_{mn\nu}^n(\alpha(\tau), \beta(\tau), \gamma(\tau)) S_{mn\nu}^{lm}(\mathbf{r}_s(\tau)) q(\tau) e^{-i\omega\tau} d\tau & \text{for } r < r_s(\tau) \\ \sum_{\nu, \mu} \sum_{l, n, m} \hat{G}_{\nu, \mu}^{l, n, m} \int D_{mn\nu}^n(\alpha(\tau), \beta(\tau), \gamma(\tau)) \hat{S}_{mn\nu}^{lm}(\mathbf{r}_s(\tau)) q(\tau) e^{-i\omega\tau} d\tau & \text{for } r > r_s(\tau) \end{cases}$$

$\hat{G}_{\nu, \mu}^{l, n, m}$: 球面調和スペクトル ω : 角周波数 \mathbf{r} : 受信点座標
 $\alpha(\tau), \beta(\tau), \gamma(\tau)$: ZYZオイラー角 τ : 時刻 $r_s(\tau)$: 音源座標
 $q(\tau)$: 音源信号

◆導出した球面調和スペクトルを, Okamotoの方法[1]で角度スペクトルに変換し, SDM (Spectral Division Method) によりラインアレイスピーカを用いて再現できることを, 数値シミュレーションで確認した。

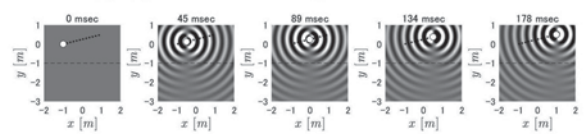


Fig. 1: Desired sound fields. The dot 'o' denotes the position of the sound source. The dotted line denotes the trajectory of the sound source. The dashed line denotes the desired line for SDM.

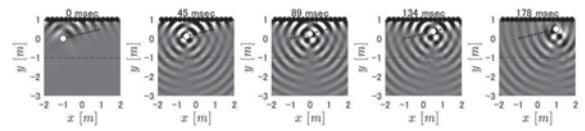


Fig. 2: Synthesized sound fields. The dot 'o' denotes the position of the virtual sound source. The dots '●' denote the secondary sound sources. The dotted line denotes the trajectory of the virtual sound source. The dashed line denotes the reference line for SDM.

[1] T. Okamoto, Proc. WASPAA 2017, pp. 180-184, 2017.

1-2-8

1-2-8 近距離場音響ホログラフィ法の境界音場制御の原理への適用による収録システムの小型化

—理論的検討—

Downsizing of the BoSC microphone array by applying the near-field acoustical holography to the boundary surface control principle -Theoretical study-

○伊勢史郎(東京電機大)

- ◆境界音場制御 (BoSC) の原理に基づく音場再現において再現される音場のエリアがマイクロホンアレイよりも広いことが経験的に知られている。
- ◆その事実から BoSC マイクの小型化が検討されている。
- ◆その理論的根拠を構築するために境界音場制御の原理に基づく音場再現に近接場音響ホログラフィ法 (NAH) を適用する。
- ◆すなわち BoSC マイクの小型化が可能であることの理論的根拠を示す。

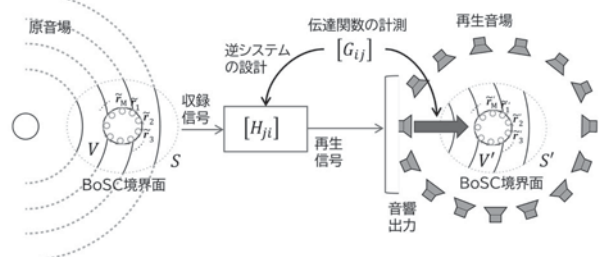


Fig. 1: Sound field reproduction system using a small microphone

1-2-9

1-2-9 近距離場音響ホログラフィ法の境界音場制御の原理への適用による収録システムの小型化—数値計算による検討—

Downsizing of the BoSC microphone array by applying the near-field acoustical holography to the boundary surface control principle -Numerical study-

☆岩内彩音, 伊勢史郎(東京電機大)

- ◆境界音場制御(BoSC)の原理に基づく音場再現における収録システムに小型剛球マイクが適用可能であるか検討を進めている。
- ◆従来の中空型と小型化した剛球型の BoSC マイクを想定し、数値シミュレーションにて音場再現を行う。
- ◆原音場におけるマイク中心と音源の間の距離を変化させたときの音場再現精度を比較する。
- ◆原音場でマイク中心から距離 (R) 1.5 m, 0.3 m に音源がある場合の音場再現精度分布を Fig. 1, Fig. 2 にそれぞれ示す。
- ◆小型剛球マイクは、マイク中心と音源の距離を小さくしたときにも、現在使用されている中空マイクに比べ、より音場再現精度が高いことを確認した。

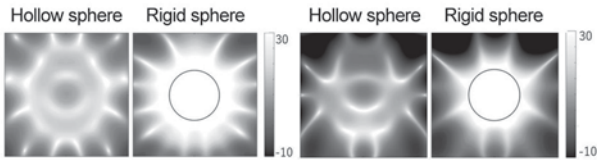


Fig. 1: Distribution of reproduction accuracy (R = 1.5 m)

Fig. 2: Distribution of reproduction accuracy (R = 0.3 m)

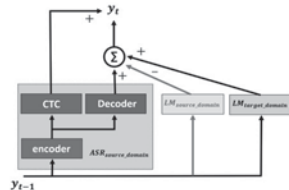
1-3-1

1-3-1 Encoder-Decoder 音声認識モデルにおける暗黙的言語情報の置換法

Language model replacement method for Encoder-Decoder speech recognition model which excludes implicit linguistic information

☆森大輝(豊橋技科大), 太田健吾(阿南高専), 西村良太(徳島大), 小川厚徳(NTT), 北岡教英(豊橋技科大)

- ◆最先端の言語モデル統合手法である Density ratio approach を日本語の Encoder-Decoder モデルに適用した。
- ◆現状では日本語の End-to-end 音声認識モデル内部の言語情報が外部言語モデルによって近似されているかどうかは不明である。
- ◆実験では日本語学術講演音声認識モデルと日本語模擬講演音声認識モデルを用いてクロスドメイン CER を算出した。
- ◆Density ratio approach (DRA)は、Shallow Fusion (SF)の CER を改善した。



Method	ASR Model	Test set	CER
Baseline	学術講演	模擬講演	16.2
SF			15.2
DRA			14.0
Baseline	模擬講演	学術講演	16.2
SF			14.6
DRA			13.0

1-2-10

1-2-10 高精度な局所再生を実現するエンクロージャレス・スピーカアレイの提案

Proposal of enclosureless speaker array for personalized sound zone generation

○福井勝宏, 小林和則, 鎌土記良(NTT)

- ◆本稿では、コンパクトで且つ音楽再生にも適した局所再生を実現できるエンクロージャレス・スピーカアレイを提案する。
- ◆エンクロージャレス・スピーカアレイの外観を Fig. 1 に示す。これは、エンクロージャのない、スピーカユニット2つをパツフル板に取り付けたのみのコンパクトな構成である。この構成により、信号処理なしで側面においてヌルを形成する。
- ◆さらに、アレイの上下側面方位に信号処理で双指向性を作ることで、アレイ近傍を除く全空間にわたる消音を実現する。
- ◆また、エンクロージャを使用しないため、本アレイはスピーカユニットのもつ本来の最低共振周波数まで再生することが可能となり、小さなスピーカユニットでも音楽再生に適したダイナミックレンジを実現できる。

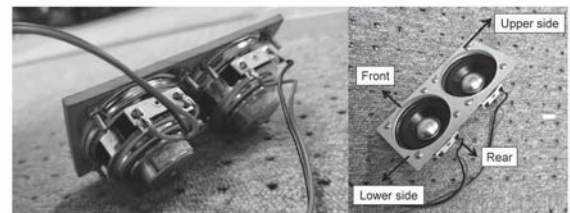


Fig.1: Enclosureless speaker array

1-3-2

1-3-2 擬似ラベリングと特徴表現学習を併用した構音障害者音声認識

Dysarthric speech recognition using pseudo-labeling and feature learning

☆澤佑哉, 富士原健斗(神戸大), 相原龍(三菱電機), 高島遼一, 滝口哲也(神戸大), △今井良枝(三菱電機)

- ◆近年、音声認識性能の向上のために、ラベルの無い音声も音声認識に活用することが注目されている。
- ◆ラベル無し音声も音声認識に活用する手法として、音声認識により擬似的にラベルを付与する手法や、ラベル無し音声のみで学習できる特徴表現学習を用いる手法が挙げられる。
- ◆構音障害者音声認識において擬似ラベルを使用する場合は、擬似ラベルの精度が低いことから、誤った正解ラベルの付与による音声認識性能の悪化が懸念される。
- ◆本研究では、音声認識モデルの訓練時にラベルに依存しない学習方法である特徴表現学習とのマルチタスク学習を行うことにより、ラベルの低信頼性を補完することを試みる。
- ◆実験の結果から、マルチタスク学習を行わない場合と比較して、提案手法は優れた性能を達成できることが分かった。



Fig.1: Training procedure of the automatic speech recognition (ASR) model.

1-3-3

1-3-3 言い淀みを考慮した自由発話の End-to-End 音声認識

End-to-End Spontaneous Speech Recognition Considering Hesitation
☆堀井こはる(豊橋技科大), 福田芽衣子(徳島大), 太田健吾(阿南高専), 西村良太(徳島大), 北岡教英(豊橋技科大)

- ◆自由発話に含まれる、吃音や言い直しといった発話の非流暢性を「言い淀み」ラベル '@' として付加し、音声認識モデルを学習。

実際の発音 : 「これを 生 生成突然変異と呼んでいます」

ラベル付け文 : これを @ 生成突然変異と呼んでいます

- ◆実際の認識結果を見ると

正解文 : 得要率は @ 事前に試行してある程度幅を @ 持たせたものに設定いたしました
 認識結果 : 得要率は @ 事前に試行してある程度幅を @ 持たせたものに設定いたしました
 目的文 : 得要率は事前に試行してある程度幅を持たせたものに設定いたしました

認識結果の言い淀みラベルは正しい位置に出力されており、モデルがラベルの意味を学習できていると考えられる。

- ◆ラベル部分を削除することで、正しい目的文を得られた。

	model	evaluation data	CER	SER
◆文字誤り率 (Character Error Rate : CER), 文誤り率 (Sentence Error Rate : CER)と もにラベル付けにより削減できた。	baseline	eval1 (CSJ)	6.1	57.5
		eval2 (CSJ)	4.4	54.0
		eval3 (CSJ)	4.7	37.5
◆ our model		eval1 (CSJ)	5.0	49.8
		eval2 (CSJ)	3.5	47.8
		eval3 (CSJ)	4.1	34.2

1-3-5

1-3-5 音声合成による wav2vec 2.0 を用いた音声認識のデータ拡張

Data augmentation for wav2vec 2.0-based speech recognition using text-to-speech
☆上乃聖, 河原達也 (京大)

- ◆ wav2vec 2.0 を用いた音声認識モデルは従来の対数メルフィルバンク (lmbf) 特徴量を入力とするモデルよりも性能が良いことが報告されている。
- ◆ 本研究では、wav2vec 2.0 を用いた音声認識モデルに対して、音声合成によるデータ拡張を行い、wav2vec 2.0 のパラメータ更新に関して比較を行う。
- ◆ 拡張実験において、単語誤り率・改善率ともに、wav2vec 2.0 を用い、自然音声・合成音声の区別なく finetuning することで従来手法よりも良いことを確認した。

ASR performance (WER (%)) for Tedlium 2 testset.
We used paired-data of train-clean-100 in LibriSpeech (100h).
We generated speech from transcriptions of Tedlium 2 (211h).

	dev	test
Baseline Model: Real (train-clean-100)		
lmbf feature [reference]	31.95	28.92
wav2vec 2.0 w/o finetuning	27.63	26.41
wav2vec 2.0 w/ finetuning	15.01	14.44
Augmented Model: Real (train-clean-100) + TTS (Tedlium 2)		
lmbf feature [reference]	19.33	18.96
wav2vec 2.0 w/o finetuning	17.99	16.46
wav2vec 2.0 w/ finetuning Real data only	20.14	18.92
wav2vec 2.0 w/ finetuning TTS data only	16.41	16.04
wav2vec 2.0 w/ finetuning All	11.04	11.02
Oracle Model: Real (Tedlium 2)		
lmbf feature [reference]	9.28	8.56
wav2vec 2.0 w/ finetuning	8.31	7.72

1-3-4

1-3-4 CTC を用いた音声認識のための 中間層予測による条件づけ

A conditioning method with intermediate predictions for CTC-based ASR
☆野崎樹文(LINE/京大), 小松達也(LINE)

- ◆CTC ベースの音声認識モデルは推論が高速というメリットがある一方、出力トークン間の条件独立性を仮定しているため、Attention ベースの Encoder-Decoder モデルと比較して精度が劣ることが多い。
- ◆提案法では、モデルの最終層に加えて中間層も CTC 損失関数で学習し、中間層で得られる事後確率分布を次の層の入力に加える。
- ◆これにより最終層の予測が中間層の予測に条件づけられ、CTC の条件独立の仮定の欠点を補うことを期待する。

- ◆英語 (TEDLIUM2)・中国語 (AISHELL-1) のコーパスを用いた実験の結果、提案モデルは CTC の高速な推論速度を保ったまま認識精度を改善した。

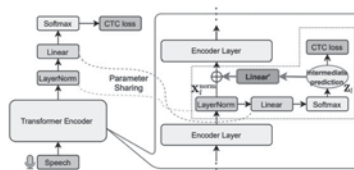


Fig. 1 An illustration of our proposed model.

Table 1 WER and RTF on TEDLIUM2 and CER and RTF on AISHELL-1.

	model	TEDLIUM2				AISHELL-1			
		dev	test	RTF	Speedup	dev	test	RTF	Speedup
Autoregressive	Transformer	12.6	10.2	0.290	1.00×	5.6	6.1	0.105	1.00×
	+ beam search	10.4	9.2	5.596	0.05×	4.9	5.4	1.604	0.07×
Non-autoregressive	CTC	12.8	12.2	0.039	7.44×	5.7	6.2	0.038	2.76×
	Mask CTC	11.9	10.7	0.058	5.00×	5.2	5.7	0.042	2.50×
	InterCTC	10.8	10.1	0.039	7.44×	5.3	5.7	0.038	2.76×
	Proposed	9.9	9.4	0.041	7.07×	4.9	5.3	0.050	2.10×

1-3-6

1-3-6 大規模言語モデルの知識蒸留による コンテキスト推定モデルを用いた 低遅延逐次音声合成

Low-Latency Streaming Text-to-Speech Synthesis with Distilled Context Estimation Network
☆佐伯高明, 高道慎之介, 猿渡洋 (東大院・情報理工)

未来予測機構を持つストリーミング・低遅延 End2End TTS

- ◆ 大規模言語モデル (GPT2) に基づく先読み機構 [Saeki+21] を持つストリーミングテキスト音声合成 (Text-to-Speech; TTS)
- ◆ GPT2 を用いたストリーミング TTS を教師モデルとし、軽量なコンテキスト予測ネットワークを持つ生徒モデルに知識蒸留
- ◆ 人間の英語話者の平均話速よりも遥かに高速な発話生成 (800 words per minute) を達成!! → リアルタイム用途も可能

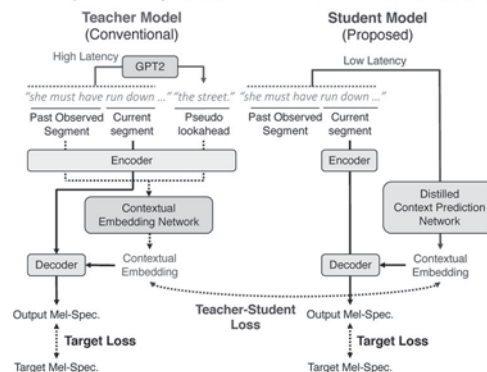


Fig. 1: Proposed teacher-student training method.

1-3-7

1-3-7 Cross-lingual, multi-speaker text-to-speech synthesis for low resource languages

○ Zolzaya Byambadorj (Tokushima Univ.), Ryota Nishimura (Tokushima Univ.), Altangerel Ayush (Mongolian Univ. of Science and Tech.), Kengo Ohta (National Institute of Tech., Anan College), Norihide Kitaoka (Toyoashi Univ.)

In this paper, we propose a cross-lingual, multi-speaker TTS system with limited target language data. However, high-quality, multi-speaker data is generally unavailable in most low-resource languages. As a result, it is impossible to synthesize natural speech for a target speaker without enough data. Therefore, we built a multilingual, multi-speaker model, which can be used to address data availability issues. We investigate the effects of using multilingual and augmented data on the performance of the low-resource TTS model. Moreover, we propose to transfer voices across languages to solve the lack of low-resource language speakers. We also combined the model with a technique for uncovering latent speech attributes to encourage it to synthesize more natural-sounding speech.

We evaluated only the target language synthesized speech using our model in this study. Our experimental results show that both multilingual and augmented data improved the performance of our TTS model to synthesize target language speech. Besides, our TTS model achieved reasonable performance in synthesizing target language speech using foreign speakers. The following Table 1 and 2 show the results of the synthesized speech using speakers in the target and foreign language, respectively.

Table 1. Naturalness MOS results with 95% confidence intervals for multilingual and monolingual TTS models. (+) and (-) indicate that the models were trained with and without augmented data or GST module

Model name	Augmented data	GST module	Naturalness MOS
Ground Truth			4.93 ± 0.060
M1 (multi)	-	-	2.50 ± 0.136
M2 (multi)	+	-	3.46 ± 0.208
M3 (multi)	+	+	3.70 ± 0.247
M4 (mono)	+	+	3.20 ± 0.199

Table 2. Naturalness and speaker similarity MOS results with 95% confidence intervals for cross-lingual TTS

Model name	Naturalness MOS	Speaker similarity MOS
M2	3.33 ± 0.183	3.05 ± 0.101
M3	3.47 ± 0.187	3.23 ± 0.172

1-3-9

1-3-9 因子分析に基づく HSMM を利用した構造化アテンション音声合成

A factor analyzed voice model for sequence-to-sequence speech synthesis using a hidden semi-Markov model based structured attention mechanism

○高木信二, △牛田光一, 橋本佳, 南角吉彦, 徳田恵一(名工大)

本稿では、隠れセマルコフモデル(HSMM)の構造を導入した構造化アテンション機構に基づく Seq2Seq モデルを拡張し、因子分析に基づく HSMM を利用した音響特徴量・音素継続長・発話スタイルの同時モデリング手法を提案する。

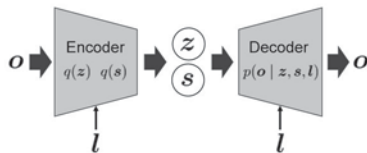


Fig. 1 提案法の概略図.

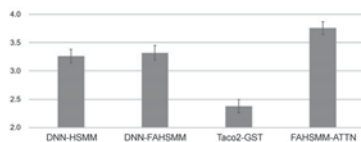


Fig. 2 主観評価実験結果(自然性).

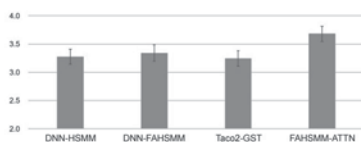


Fig. 3 主観評価実験結果(感情表現).

1-3-8

1-3-8 文横断コンテキストを用いた日本語オーディオブック音声合成の評価

Evaluation of Japanese Audiobook Speech Synthesis Conditioned by Cross-Sentence Context

☆中田亘(東大工学部), 郡山知樹, 高道慎之介(東大院・情報理工), 井島勇祐, 増村亮 (NTT), 猿渡洋(東大院・情報理工)

▶ 背景: オーディオブック音声合成

- 文横断コンテキストを用いることで、英語においてオーディオブック合成音声の品質が改善することが確認されている。
- 日本語においては未確認

▶ 目的: 日本語オーディオブック音声合成において、文横断コンテキストを用いた音声合成モデルを評価

▶ 実験: 客観評価, 主観評価を用いて合成音声の評価

- 連続した 5 文の合成音声において文横断コンテキストを用いる事により合成音声の品質が向上
- BERT を転移学習することにより更に合成音声の品質が向上

1-3-10

1-3-10 HiFi-GAN ボコーダにおける LPCNet 特徴量の検討

Investigation of LPCNet features in HiFi-GAN vocoder

◎松原圭亮^(1,2), 岡本拓磨⁽²⁾, 高島遼一⁽¹⁾, 滝口哲也⁽¹⁾, 戸田智基^(3,2), 河井恒⁽²⁾ (1 神戸大学, 2 NICT, 3 名古屋大学)

- ◆ 近年では HiFi-GAN 等の複数話者音声合成が可能かつ CPU によるリアルタイム合成が可能なニューラルボコーダが多く提案されている
- ◆ 本稿では HiFi-GAN をはじめとする高速な複数話者ニューラルボコーダについて、品質および合成速度について比較する
- ◆ またテキスト音声合成における頑健性が報告されている LPCNet 特徴量を用いた場合の品質についても比較を行う

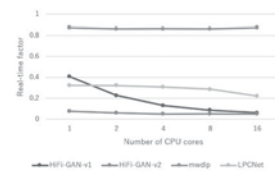


Fig. 1: Result of real-time factor for inference.

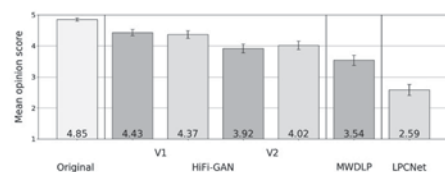


Fig. 2: Mos results of multi-speaker model using unseen male speaker.

1-3-11

1-3-11 発話の局所的類似性に基づく短発話向け話者照合

Speaker Verification for Short Utterances based on Local Similarities between Utterances.

○村田有実子, 安藤厚志(NTT)

- ◆本研究では、登録発話が長く照合発話が短い条件下でのテキスト非依存話者照合を行う。
- ◆従来、登録発話及び照合発話の部分区間(以後、セグメントと呼ぶ)ごとに話者表現を抽出し、それらの話者類似度の総当たり平均に基づいて話者照合を行う手法が提案されている。この手法は全てのセグメント組に対して等しく話者類似度を評価するが、特定のセグメント組、例えば類似した音韻を持つセグメント組に話者同一性の情報が強く表れる可能性がある。
- ◆本稿では、発話長が著しく異なる状況下においても頑健な話者照合を行うため、発話のセグメント単位の音韻情報の類似性に着目した。発話長が著しく異なる状況下におけるセグメント単位の話者表現に基づく話者照合の精度向上を目的とし、照合発話と音韻情報が類似した登録発話のセグメントを利用する話者照合手法を提案する。
- ◆実験の結果、登録発話が長い条件において提案手法は従来手法に比べて高い話者照合精度を示すことが明らかとなった。

Table1: Equal Error Rate (EER) for each enrollment / test utterance length.

	Enrollment 8 [sec]				Enrollment 30 [sec]				Enrollment 60 [sec]			
	1	2	4	8	1	2	4	8	1	2	4	8
Conventional	25.8	14.7	9.7	8.1	19.2	9.5	4.5	3.5	18.3	8.6	3.6	2.8
utfr-emb	10.1	7.5	5.8	5.4	7.9	5.3	3.7	3.3	7.8	5.0	3.4	3.1
seg-emb	12.8	9.2	7.0	6.2	10.1	6.0	3.8	3.0	9.5	5.6	3.2	2.5
Proposed	11.9	8.3	6.3	5.7	9.0	5.0	3.1	2.6	8.4	4.5	2.8	2.2
seg-emb+attn												
seg-emb+plm+attn												

1-3-13

発話リズムに基づく話者埋め込みを用いた音素継続時間長モデリングの検討

Phoneme Duration Modeling with Speaker Embedding Based on Speech Rhythm

○藤田健一, 安藤厚志, 井島勇祐(NTT)

- ◆本研究では、個人の再現性が高い音声合成の実現のため、発話リズムの個人性を表現する話者埋め込みベクトルの獲得手法を提案する。
- ◆提案法では、音素とその継続時間長を入力として話者を識別するDNNモデルを学習し、得られたモデルの中間層出力を話者埋め込みベクトルとして抽出する。このDNNモデルが話者の識別を行えば、ここから発話リズムの個人性を表現したベクトルが得られる。
- ◆本研究では、話者識別と音声合成に関する2つの実験を行った。話者識別実験では、提案した話者識別モデルが10.3%の等価誤り率(EER)を達成した。更に可視化により発話リズムの類似性が話者埋め込み空間における分布に関係することがわかった。音声合成実験では、提案した話者埋め込みベクトルを用いて生成した音声は、従来の話者埋め込み(x-vector)を用いた音声に対して、客観評価と主観評価の両方で優れていることを確認した。

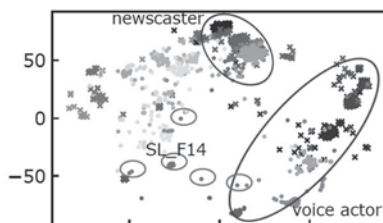


Fig.1: Distribution of speaker embedding vectors

1-3-12

1-3-12 ResNet/ResNeXtによる深層話者埋め込みの検討

Deep Speaker Embedding by ResNet/ResNeXt

☆鈴木克樹, 堀内靖雄(千葉大), 柘植寛(大同大), 黒岩真吾(千葉大)

- ◆近年、深層話者埋め込みを用いた話者認識手法が注目を集めている。従来の深層話者埋め込みの研究では、実験データセットに約100万発話、多言語からなるVoxCelebが主に用いられている。
- ◆本稿では、VoxCelebの約30分の1の日本語音声データを用いて、深層話者埋め込みの検討を行う。DNNの構造としては、ResNetとResNeXtの評価を行う。評価は、話者識別により行った。
- ◆話者オープン条件(DNNの学習に識別対象話者の音声を含まない)ではResNetの方が識別エラー率が低かった。ResNet, ResNeXt共通での誤りは11話者の延べ12発声だった。
- ◆話者クローズド条件(DNNの学習に識別対象話者の音声を含む)ではResNeXtの方が識別エラー率が低かった。なお、全結合のみで構成したDNNでの識別エラー率は非常に高かった。

Table 1: Segment-level identification error rate [%] (エラー数)

構造	セグメントサイズ	話者オープン	話者クローズド
ResNet	1秒	14.4 (148)	1.0 (11)
	2秒	6.9 (71)	0.2 (2)
	3秒	3.0 (31)	0.3 (3)
ResNeXt	1秒	14.9 (153)	0.6 (6)
	2秒	6.4 (66)	0.1 (1)
	3秒	4.0 (41)	0.1 (1)

1-3-14

1-3-14 話者依存度に応じた特徴抽出器によるdisentangleな声質変換

Disentangle voice conversion using feature extractors with multiple speaker-dependencies.

☆井路巧, 岸田拓也, 中鹿亘(電通大)

- ◆FaderNetVCと呼ばれる深層学習モデルは、Encoderにより発話から発話内容のみを出力し、one-hotな話者ラベルと共にDecoderに入力することで、声質変換を実現している。
- ◆先行研究では、FaderNetVCに話者情報を出力するEncoderを追加することで、変換時に話者ラベルを必要としないSplitterNetVCを提案した。
- ◆しかしSplitterNetVCでは、音声を発話内容と話者情報に完全に二分することが難しかったため、本研究では発話内容と話者情報が混在した情報を出力するEncoderをSplitterNetVCに追加し、話者情報、発話内容の分離を促すことを目指す。
- ◆Table.1より、音声品質において、提案手法は既存手法よりもよい結果が得られた。また、unseen時の変換音声類似性についても、他のモデルより良い結果が得られた。

quality	target	FDVC	SPVC	TCVC
seen	4.67	2.16	2.40	2.68
unseen	4.79	2.47	2.12	2.78
similarity	-	FDVC	SPVC	TDVC
seen	-	2.82	2.95	2.69
unseen	-	2.38	2.44	2.69

Table.1 Results of the subjective evaluation (MOS) FDVC: FaderNetVC, SPVC: SplitterNetVC, TCVC: Trienvc (Proposed)

1-3-15

1-3-15

L2音声を対象とした瞬時的了解度計測のための
シャドーイング音声コーパスの構築Development of a shadowing speech corpus for instantaneous
intelligibility measurement on L2 speech峯松信明, 朱伝博, 榎原卓也, 箱田峻, 齋藤大輔 (東大)
中西のりこ (神戸学院大)

- 学習者音声の評価軸は、母語話者らしさ？、伝わりやすさ？
- 後者をとった場合、学習者の音声はどの程度スラスラ相手に伝わっているのか？そして、それはどのように計測すべきなのか？
- 聞き手の頭の中を、簡便な方法で、覗き込むことはできないのか？

■ シャドーイング：

現在：外国語教室で学習者がモデル音声を聞きながら復唱する訓練
かつて：脳画像技術が一般化する前に、人間の音声知覚過程（心的辞書検索過程）を分析する目的で導入された知覚実験タスク

- 学習者音声を母語話者相当にシャドー&スクリプトシャドーさせ、両者を比較すれば、どこで、どのくらいシャドーが崩れたのか（＝どこでどのくらい聞き取りにつまずいたのか）が測定可能
→ 瞬時的了解度のラベリングが可能

- 言語背景の異なる母語話者（相当）に、中規模のL2音声サンプルをシャドー+スクリプトシャドーさせ、瞬時的了解度ラベル（単語単位、音素単位、フレーム単位）が付与されたL2コーパスを構築中



1-3-17

1-3-17 モデル音声への接近と逸脱及び
モデル音声の知覚に着目した
学習者シャドーイング音声の分析Analysis of learners' shadowing based on their convergence to and divergence
from model speech and their perception of model speech☆榎原卓弥, 朱伝博, 齋藤大輔, 峯松信明 (東大),
中西のりこ (神戸学院大)

- ◆ シャドーイングは、提示音声(モデル音声)に対する、1) 知覚、2) 理解、3) 生成(模倣)を同時に行う外国語学習の多目的型訓練法である。
- ◆ モデル音声(M)に対して、シャドーイング(S)を繰り返すことでMの聞き取りの様子がどのように向上するのか分かるようなタスクデザインを行った。
- ◆ 本実験では、Sを3回行い、その後テキストを見せながらシャドーイング(スクリプトシャドーイング, SS)、そして2週間後に音声を提示せずに読み上げ音声(R)を収録した。
- ◆ その結果、SがMへと接近する様子やRがMから逸脱する様子を分析することが可能となった。
- ◆ また、SとSSの違いはテキストの有無であるため、これらと比較することでリスニング力、すなわち1) 知覚、2) 理解を定量化することができる。

1-3-16

1-3-16 Automatic Prediction of Instantaneous
Intelligibility of L2 Speech in Sequence☆Chuanbo Zhu, Kunihara Takuya, Daisuke Saito,
Nobuaki Minematsu(The University of Tokyo),
Noriko Nakanishi(Kobe Gakuin University)

- ◆ L2 speech instantaneous intelligibility can be annotated in phoneme and word unit using the reverse shadowing technique.
- ◆ A corpus of L2 speech with its sequential annotation was collected.
- ◆ Making use of the corpus, we report our first attempt of automatic prediction whether a given word or phoneme is intelligible or not.
- ◆ Three model using acoustic and text features were tested out.
- ◆ Part of the classification results is reported in the tables below, which is promising.

word unit classification	Precision	Recall	F1 score
intelligible	0.734	0.632	0.679
unintelligible	0.645	0.744	0.691

phoneme unit classification	Precision	Recall	F1 score
intelligible	0.755	0.690	0.721
unintelligible	0.725	0.784	0.753

1-3-18

1-3-18 波形の相互相関に基づく基本周波数推定
における距離尺度による影響の調査Investigation of the effect of error function on fundamental frequency
estimation based on cross-correlation of waveforms

☆小口純矢, 森勢将雅(明治大)

君だけの誤差関数をカスタマイズして
最強の基本周波数推定器を作ろう！

$$d(n, \tau) = \sum_{k=n+1}^{n+W} ?$$

可能性は無限大！

- 絶対誤差
- 二乗誤差
- p 乗誤差 ($0 < p < 2$)
- Log-cosh

1-4-1

1-4-1 音声の時間情報に同期する $\theta \cdot \gamma$ 帯域神経オシレーションの帯域間相互作用とその障害*

The interaction of theta and gamma oscillations phase-locked to speech temporal information and its deficits in schizophrenia

©田村俊介, 光藤崇子, 平野羊嗣(九州大学大学院医学研究院精神神経医学)

- ◆連続的な音声に含まれる振幅包絡と時間微細構造の処理には、それらと同じ速さで生じる神経オシレーションである θ (4-8 Hz) $\cdot \gamma$ (30 Hz 以上) 帯域の神経オシレーションの関与が示唆されているが、両者の相互作用についてはほとんど知見がない。
- ◆健常者 (N=22) と γ 帯域神経オシレーションに特徴的な異常を有することが知られている統合失調症者 (N=14) を対象に、音声聴取時の $\theta \cdot \gamma$ 帯域の神経オシレーションを計測し、両者の間の相関を調べた。
- ◆刺激には基本周波数を 80 Hz で固定したモノトーン音声 (monotone speech, MS) と時間微細構造が含まれない雑音駆動音声 (noise-vocoded speech, NVS) を用いた。また、MS と似た時間構造を持つ非音声刺激として、80 Hz click train を 5 Hz で変調した刺激 (amplitude-modulated click train, AC) を用いた。
- ◆MS 条件では、 $\theta \cdot \gamma$ 帯域神経オシレーションが計測され、健常者では両者に有意な正の相関が見られた。一方、統合失調症者では有意な相関が見られなかった。
- ◆AC 条件でも $\theta \cdot \gamma$ 帯域神経オシレーションが計測されたが、健常者、統合失調症者ともに有意な相関が見られなかった。
- ◆健常者では、音声特有の θ 帯域 $\cdot \gamma$ 帯域神経オシレーションの相互作用が見られたが、言語性幻聴や連合弛緩などを特徴とする言語機能異常が見られる統合失調症者では相互作用が見られなかった。

1-4-3

1-4-3 発話時の骨導伝達特性の測定における気導音遮断の検討

Investigation of sidetone attenuation for measurements of bone-conduction transmission characteristics during vocalization

☆鳥谷輝樹(北陸先端大), △Peter Birkholz(TU Dresden), 鶴木祐史(北陸先端大)

- ◆背景: 骨導伝達特性測定における外耳道内放射音の観測では、外耳道内への気導音の混入を防ぐ必要あり。イヤーマフを利用した従来検討 (Reinfeldt *et al.*) では側頭部振動の同時観測が困難。前報 (鳥谷ら, 2021) での遮音壁の利用では、壁自体の振動が測定結果に影響した可能性あり。
- ◆目的: 遮音壁の振動を抑制した下での開放外耳道の伝達特性が、従来知見と類似するかどうかを検討すること。
- ◆方法: 無響箱に固定した遮音壁で顔を覆い、(1) 母音発話時の気導音声と外耳道内放射音を観測、(2) 気導音声を基準とした外耳道内放射音の相対振幅特性を従来検討・前報と比較。
- ◆結果: 前報に比べ、本稿での相対振幅特性は特に 1 kHz 以下でイヤーマフによる従来知見と類似傾向あり。
- ◆結論: 遮音壁の振動を抑制した下での開放外耳道の伝達特性は従来知見と類似することが分かった。

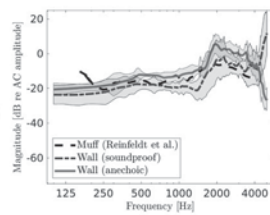


Figure 1: Ear canal sound pressure level relative to that of air-conducted speech for vowel /a/

1-4-2

1-4-2 宣伝音声の聴取によって生じられる感情に対する聴取者の年代及び性別の影響の分析

An analysis of the effect of listeners' gender and age on perceived emotions from advertising speech

©長野瑞生, 井島勇祐, 廣谷定男(NTT)

- ◆感情を媒介とする消費者行動モデル [R. Donovan +, 1982]
 - 消費者の購買行動を外部刺激→感情→購買意欲の3ステップで表すモデル
 - 音声刺激においても上記モデルが適用可能であることを示したが、聴取者の属性ごとの感情の媒介効果の違いは不明 [長野他, ASJ2021a]
- ◆本研究では聴取者の年代を統制して新たな主観評価実験を行い、感情の媒介効果の年代や性別による違いについて検証
- ◆平均 F0, 話速, F0 分散を変更した宣伝音声を用いて大規模な主観評価実験を実施し、年代・性別ごとに媒介分析
 - 年代・性別に依らず感情を媒介とする消費者行動モデルが有効
 - 感情の媒介効果は年代・性別によって異なる
 - ◇ 20代は他の年代に比べて感情そのものの媒介効果が小さい
 - ◇ 60代は他の年代に比べて覚醒の媒介効果が大きい
 - ◇ 女性は男性よりも快感情の媒介効果が大きい

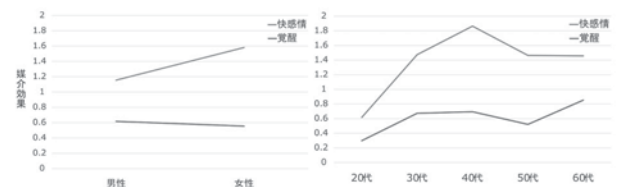


Fig.1: 性別ごとの媒介効果

Fig.2: 年代ごとの媒介効果

1-4-4

1-4-4 頭部運動が相対回転移動する音像の位置知覚に与える影響

The effect of head rotation on sound localization of relative rotated sound image

☆守部颯一郎, 坂本修一(東北大)

- ◆本研究では、耳入力信号の変化が等しくなる相対回転移動音像を用いて、音像の位置知覚に頭部回転の有無が及ぼす影響を検討した。
- ◆聴取者が静止し音源が回転する条件 (音源移動条件)、音源位置は変わらず聴取者が回転する条件 (頭部回転条件) を設定した。また基準となる条件として音源位置、聴取者ともに静止する条件 (静止条件) でも実験を行った。
- ◆知覚された音像位置のずれの大きさに着目すると、音源移動条件の方が頭部回転条件に比べずれが有意に大きかった (Fig. 1A)。頭部回転条件では自身の動きによって耳入力信号が変化することから、予測によってずれが小さくなったと考えられる。
- ◆知覚された音像位置の方向に着目すると、音源移動条件でのずれは音源の進行方向であったのに対し、頭部回転条件では頭部回転方向へ増加する傾向であった (Fig. 1B)。頭部回転条件の結果は、頭部が角速度回転運動を行う場合と同様の傾向であり、なぜ等速度回転中でも同様の傾向が見られたのかについて今後の検証を要する。

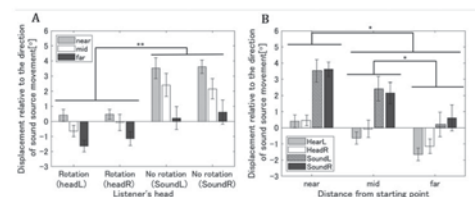


Fig. 1. Displacement relative to the direction of sound source movement. (A: The effect of the head rotation (**: $p < .01$). B. As the function of the distance from starting point (*: $p < .05$).

1-4-5

頭部側端の音像定位に
位置判断基準が与える影響

The effect of position criteria on sound image localization at the lateral position

☆ 加村響志朗, 森川大輔, モクタリパーハム, 平原達也(富山県立大)

- ◆ こめかみに当てた手を頭部側端位置の判断基準として用い、音像が頭部側端位置に定位する広帯域雑音の両耳間時間差 (ITD) 及び両耳間音圧差 (ILD) を測定した。
- ◆ 約半数の受聴者で判断基準がある方が測定値の分散が有意に小さく、側端位置の判断のふれを小さくするために有効であることがわかった。
- ◆ ただし、受聴者間の分散は判断基準がある場合も大きかった。

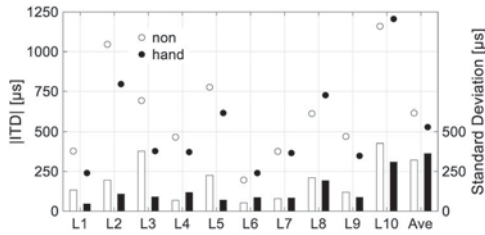


Fig.1 ITD at the lateral position

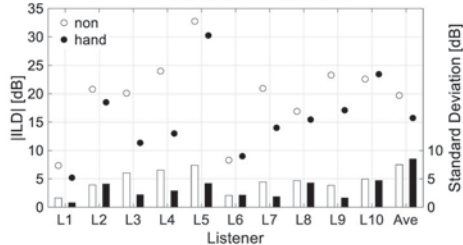


Fig.2 ILD at the lateral position

1-4-7

1-4-7 局所耳介表面反射が耳介伝達関数に
及ぼす影響の感度解析による検討

An investigation of the effect of local pinna-surface reflection on the pinna-related transfer function by sensitivity analysis

☆ 廣田裕太郎, モクタリ パーハム, 森川大輔 (富山県立大)

- ◆ 耳介伝達関数 (PRTF) のスペクトラルキューにおいて第 1 ノッチ (N1) が仰角方向の音像知覚に重要とされているが、その生成メカニズムは完全には明らかになっていない。
- ◆ 局所耳介表面での音の反射が、PRTF の N1 の振幅に与える影響を調べた。
- ◆ 基準となる耳介表面の反射率 0.99 に対し、局所耳介表面の反射率を 0.5 とし、N1 の振幅変化量 ΔA [dB] を算出した。
- ◆ 仰角 0° の N1 の深さ生成に重要な耳介箇所は舟状窩、三角窩、耳甲介舟、珠間切痕であり、それを阻害する箇所は耳輪、耳輪脚、耳甲介腔の後ろ壁であった。また、仰角 30° 以上では、その傾向が逆となった。

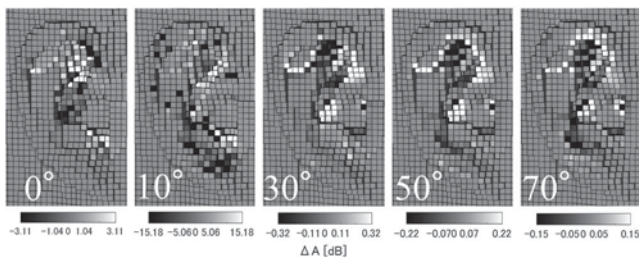


Fig.1 Amplitude sensitivity maps of N1 at 1.0 m in the median plane

1-4-6

1-4-6 合成バイノーラル信号の周波数帯域が
音像の分離知覚の検知限に与える影響

Effect of frequency band of synthesized binaural signal on detection limit of spatially split sound images

☆ 酒井翼, 森川大輔, モクタリパーハム (富山県立大)

- ◆ 両耳間時間差 (ITD) と両耳間音圧差 (ILD) の相互作用によって生じる音像の分離知覚の検知限に、刺激音の周波数帯域が与える影響を調べた。
- ◆ 2つの頭部インパルス応答を畳み込み、前面で分離する F 条件、背面で分離する B 条件について、それぞれ 3つの帯域フィルタを通した刺激音を提示する Low、High、Wide で分離知覚実験を行った。
- ◆ 検知限の角度差 DL_{θ} は、Low と Wide に差がなく、High との間に差があった。検知限の刺激音間の ITD の差 DL_{θ}^{ITD} は Low、Wide と ITD のみでの分離 (ITD-cond) と一致した。ILD の差 DL_{θ}^{ILD} は多くの条件や、F 条件と B 条件の組み合わせに差があった。
- ◆ ITD と ILD が同時に作用する場合、1 kHz 以下の帯域が含まれれば検知限が一致し、ITD によって音像が分離知覚されることがわかった。

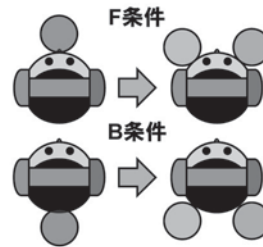


Fig.1 Sound images that listeners perceived

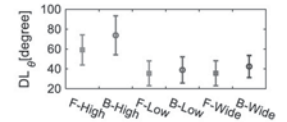


Fig.2 DL_{θ} in each condition

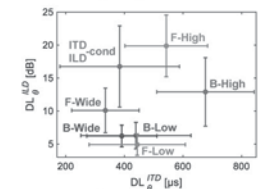


Fig.3 DL_{θ}^{ITD} and DL_{θ}^{ILD} in each condition

1-4-8

1-4-8 頭部伝達関数のノッチ周波数の変移が
音像の前後誤判定に及ぼす影響

Effects of shift in frequency of spectral notches of head-related transfer functions on the front-back error in sound image localization.

☆ 菊地勇成 (千葉工大・院), 飯田一博 (千葉工大・先進工)

- ◆ N1 および N2 周波数の変移が音像の前後誤判定および音像の上昇に及ぼす影響について検討した。その結果、以下の知見を得た。
- ◆ 目標方向が 0° では、N1 周波数の変移が音像の上昇に及ぼす影響が強く、180° では N2 周波数の変移が前後誤判定と音像の上昇の両方に及ぼす影響が強いこと示された (Fig.1)。
- ◆ また、0° で平均仰角誤差が顕著に増加しない N1 周波数の変移量は 0.16 oct であった。180° で前後誤判定率が顕著に増加しない N2 周波数の変移量は 0.16 oct であった。同様に 180° で平均仰角誤差が顕著に増加しない N2 周波数の範囲は 0.12 oct であった。

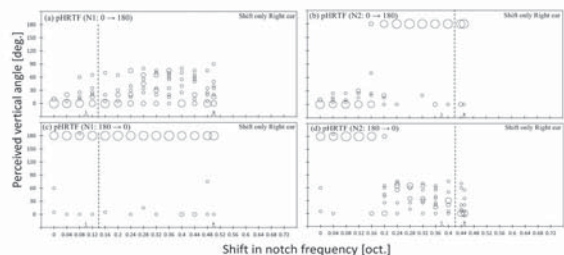


Fig.1 Responses to pHRTFs, of which notch frequency was shifted.

1-4-9

1-4-9 頭部伝達関数のノッチとピークの振幅レベルが正中面の音像定位精度の及ぼす影響

Effects of the amplitude of notches and peaks in head-related transfer functions on sound localization accuracy in the median plane

☆相崎翼(千葉工大・院) 飯田一博(千葉工大・先進工)

正中面0°および180°においてN1, N2, P1の振幅レベルが音像定位精度に及ぼす影響を検証した。その結果以下のことが明らかになった。

- 1) N1, N2の振幅レベルは0°, 180°どちらにおいても本人のノッチレベル、もしくはそれより深い場合は目標方向に知覚した。
- 2) P1の振幅レベルは、0°において本人のピークレベル±5 dBであれば目標方向付近に知覚する。また、180°においては本人のピークレベル、もしくはそれより低くすると目標方向付近に知覚した。

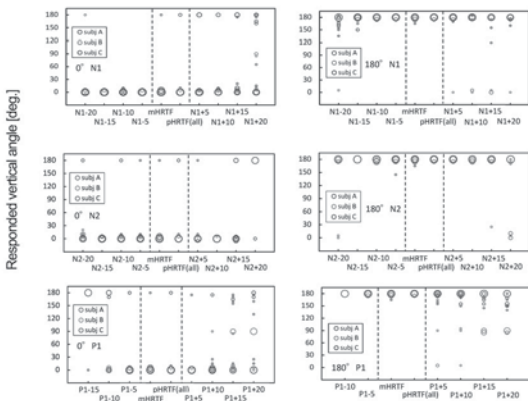


Fig.1 Responses to the measured HRTF, pHRTF(all), pHRTF_N1, pHRTF_N2, and pHRTF_P1 for the front and rear directions.

1-4-10

1-4-10 正面方向の個人化パラメトリック頭部伝達関数の生成 - ノッチとピークの振幅レベル・尖鋭度設定における固定値の適用性検証 -

Generation of individual parametric head-related transfer functions of the front direction

- Validation of the fixed values of levels and Q factors for the notches and peaks -

☆西山織絵(千葉工大・院) 飯田一博(千葉工大・先進工)

本稿では、個人化パラメトリックHRTF(pHRTF)の生成において、ノッチとピークのレベルと尖鋭度は聴取者に共通した固定値に設定できるのではないかと仮説を立て、正面方向を目標方向として音像定位実験を実施した。

その結果、Table 1に示す固定値を設定することにより、Fig.1に示す方向に音像を制御できることが示された。

Table 1 Amplitude level and Q factor of notches and peaks for pHRTF(A).

Stimulus	Level (dB)				Q factor			
	N1	N2	P1	P2	N1	N2	P1	P2
A1			4					
A2	-12	-15	7		20	25		
A3			10					
A7			4	6			3	7
A8	-18	-21	7		39	42		
A9			10					
A10			13					

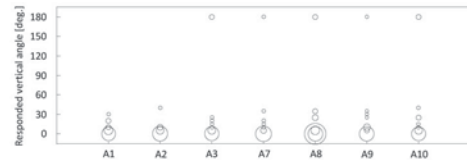


Fig.1 Responses to generated pHRTFs A1-A3 and A7-A10 for three subjects as the target vertical angle of 0°.

1-6-1

1-6-1 光学透過型ヘッドマウントディスプレイを用いた音圧分布可視化手法に関する研究

—その1 システム構成とデータ処理の概要—

Studies on visualization method of sound pressure distribution using optical see-through head-mounted display

- Part 1: Overview of system configuration and data processing -

◎井上敦登, 寺岡航, 及川靖広(早大理工/INSPIREI),

佐藤考浩, 岩根康之, 小林真人(飛島建設)

- ◆ビームフォーミングによって得られた音場情報を光学透過型ヘッドマウントディスプレイ (Microsoft HoloLens2) を用いて実空間上に拡張するシステムOTOMIRUを開発した。
- ◆ユーザは視点移動しながら観察することが可能。
- ◆ハンドトラッキングUIで操作。
- ◆複数ユーザが操作や情報を共有することが可能。
- ◆解析結果を最大25fpsで表示可能。
- ◆音源の移動や観測者の移動にもリアルタイムで追従。

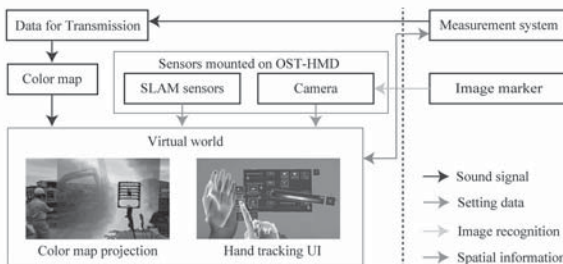


Fig.1: Process flow of the visualization system. The binary data of the sound pressure distribution are obtained from the measurement system. Various setting values are synchronized with the measurement system.

1-6-2

1-6-2 光学透過型ヘッドマウントディスプレイを用いた音圧分布可視化手法に関する研究

—その2 実験室および屋外におけるシステム有効性の検討—

Studies on visualization method of sound pressure distribution using optical see-through head-mounted display

- Part 2 Examination of system effectiveness in a room and outside -

◎佐藤考浩, 岩根康之, 小林真人(飛島建設),

及川靖広, 井上敦登, 寺岡航(早大理工/INSPIREI)

- ◆光学透過型ヘッドマウントディスプレイ (OST-HMD) を用いた音圧分布可視化手法について、実験室内および屋外での試験計測の結果から本システムの有効性や課題を検討した。
- ◆実験室内での検討では、計測断面が音源傍側に設定されていることが視覚的に把握でき、従来手法に比べカラーマップの奥行き方向の位置情報を把握しやすいシステムであることを確認した (Fig.1)。
- ◆屋外での計測では、建設工事現場での騒音管理や音源探査への適用可能性が確認された (Fig.2)。一方 OST-HMD での空間形状認識が機能しにくくなり、物体の形状に沿ったカラーマップの投影が行われぬ点やカラーマップに欠損が生じる点が課題として抽出された。

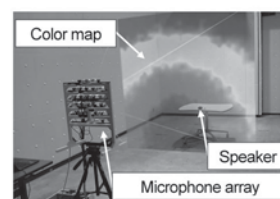


Fig.1 Visualization of sound pressure in a laboratory

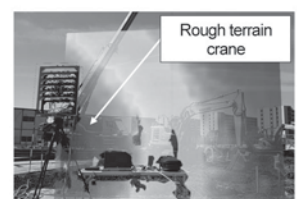


Fig.2 Visualization of sound pressure at construction site

1-6-3

1-6-3 複合現実技術と手持ちマイクロホンを用いた室内インパルス応答計測による音場の複数平面可視化

Visualization of room impulse responses on multiple planes with mixed reality and handy microphones

☆渡辺靖明, 池田雄介(東京電機大), 及川靖広(早大理工)

- ◆これまで我々は、複合現実技術と手持ちマイクロホンを用いた音場の可視化システムを提案してきた。
- ◆本発表では、より詳細な音場の可視化を目指し、2次元平面を複数枚同時に表示することで、音伝搬の様子を3次元的に可視化する。
- ◆実験では、シミュレーション実験により得られたグリッド上のインパルス応答データを、実際に Microsoft HoloLens を用いて可視化を行う

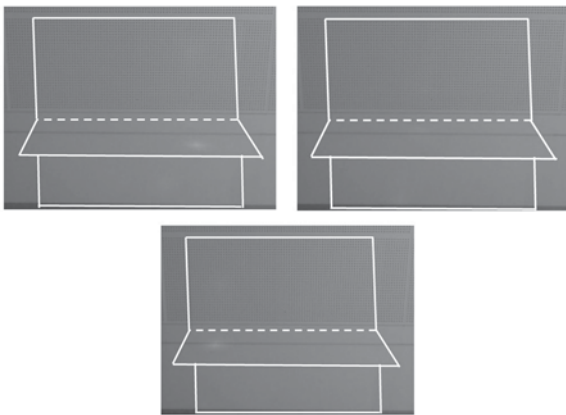


Fig. 1: Examples of MR visualization of impulse responses (Upper left part: 2.0ms, Upper right part: 3.0ms, Lower part: 4.0ms)

1-7-1

1-7-1 振幅拡大比が個々に設計可能な縦-ねじり振動源

Longitudinal-Torsional Vibration Source with Individually Designable Amplitude Amplification Factor

○浅見拓哉, 三浦 光(日大・理工)

- ◆筆者らは、縦-ねじり振動による面状振動軌跡を用いた超音波応用加工の提案をしている。本稿では、面状振動軌跡振を生じさせる振動源に用いるためのホーンとして、各幅拡大比が個々に設計できる Fig. 1 に示す様な中空部を有するステップホーンを提案し、検討を行った。検討は、汎用有限要素法解析ソフトを用い、中空部を有するステップホーンの振動モード、及び振動分布の検討を行った。
- ◆汎用有限要素法解析ソフトを用いて固有値解析を行った結果、Fig. 1 に示した中空部を有するステップホーンは、17.690 kHz でねじり振動の半波長共振となり、28.349 kHz で縦振動の半波長共振となることがわかった。また、ねじり振動の半波長共振の拡大比は、3.55 であり、縦振動の半波長共振の拡大比は、4.19 であった。これより、各振動の振幅拡大比の値を近づけることがわかった。

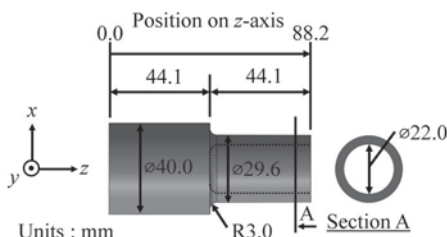


Fig. 1. Stepped horn with hollow part.

1-6-4

1-6-4 移動音源/受音点の3次元FDTD法シミュレーション

3-D FDTD simulation of moving sound source and receiver

○土屋 隆生(同志社大・理工), 手嶋 優風, 飛龍 志津子(同志社大・生命)

本稿では、任意軌道を伴う移動音源/受音点を3次元FDTD法に実装した。時間ステップごとに音源/受音点を移動させながら移動経路上のグリッド点から刻々の音源波形を放射あるいは音圧波形を受音する直接法についていくつかの数値実験の結果、ドップラー効果を組み入れた移動音源/受音点を精度良く表現できることが示された。

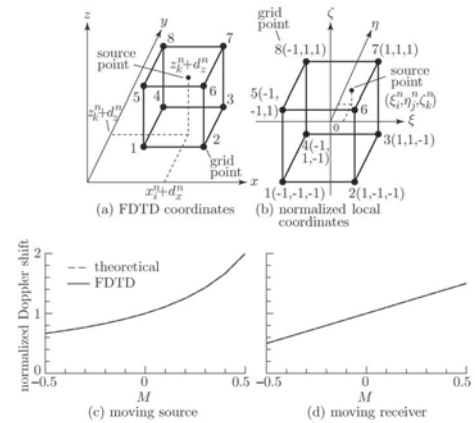


Fig: FDTD grid and local coordinates for 3-D field, and relationship between speed of moving source or receiver and normalized Doppler shift.

1-7-2

1-7-2 異なる断面積を有するヘリカルスリット付き伝送棒とホーンを用いた超音波振動源の開発

Development of ultrasonic vibration source using transmission rod with helical slits and horn having different cross section

☆宮田義大, 浅見拓哉, 三浦 光(日大・理工)

- ◆筆者らは、超音波接合の接合強度を向上する方法について検討している。従来の超音波接合は直線振動のみによる接合が多く、接合強度が低いなどの問題があった。そのため、筆者らは複合振動を生じる超音波複合振動源の開発を行っている。
- ◆本稿では複合振動を発生させるために、Fig. 1 に示すようなヘリカルスリットを施した伝送棒を用いた複合振動源の振動特性について検討した。

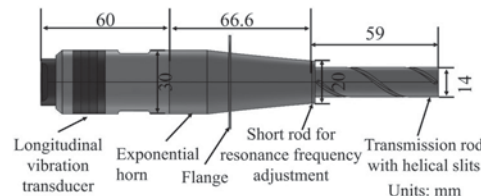


Fig. 1. Outline of ultrasonic vibration source.

- ◆共振周波数調整用短棒の長さを変化させた場合における超音波複合振動源の周波数差 (= |ねじり振動共振周波数 - 縦振動共振周波数|) の変化を検討した。
- ◆その結果、周波数差は共振周波数調整用短棒の長さ 0.8 mm において最も少なくなることが分かった。

1-7-3

1-7-3 2つの円筒形振動板型超音波音源を用いた煙霧質の凝集

Agglomeration of aerosol by two ultrasonic sound sources using cylinder type vibrating plate with rigid wall

☆宝田祐介, 浅見拓哉, 三浦 光(日大・理工)

◆本研究では、空中強力超音波による煙霧質凝集の向上を目的として、定在波音場を用いた凝集を行っている。本稿では、円筒形振動板型超音波音源を用いて、異なる入力電力における煙霧質の凝集について検討を行った。

◆Fig. 1 は本検討で行った実験結果である。煙霧質の質量濃度は超音波照射直後に急激に減少し、しばらくすると一定値に落ち着くことが分かった。また、凝集率は、音源への入力電力が1, 3, 5, 10, 20 W/unitの時、それぞれ32, 60, 62, 72, 85%となった。

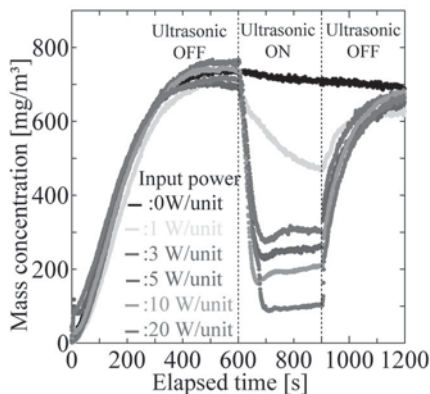


Fig. 1 Relationship between mass concentration and elapsed time.

1-7-5

1-7-5 端面反射型光ファイバマイクの空気中における指向性測定

Directivity of the optical fiber microphone based on boundary reflection

○大岡佳生, 星 貴之(ピクシーダストテクノロジーズ株式会社)

◆近年、強力超音波(40 kHz 周辺)の応用が盛んに試みられている。
➢ 超音波は波長が短いため、マイクが音場を乱して正確な値が得られなかったり、マイクの指向性の影響を受けたりする。
➢ 光ファイバを用いた音響センサは、音場を乱しにくく無指向性に近いことが期待される。

◆本報では光ファイバ先端の反射係数が外界の屈折率依存であることを利用したマイクを作製し、空気中において40 kHz 超音波に対する線形性と指向性を実測した結果について報告する。

➢ 測定地点の音圧を変えながら参照マイクと光ファイバマイクを比較したところ、線形性が確認された (Fig. 1)。
➢ 光ファイバマイクを回転させながら測定したところ、指向性が全方位にわたってほぼフラットであることが確認された。(Fig. 2)。

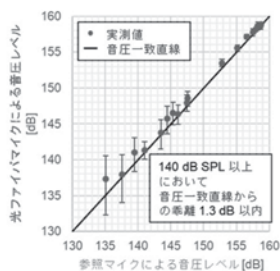


Fig. 1 : 線形性

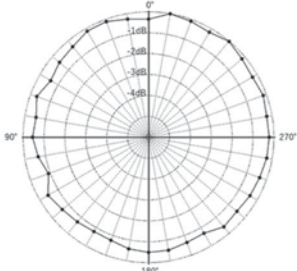


Fig. 2 : 指向性

1-7-4

1-7-4 超音波曝露の許容値に関する検討 第二報 非線形減衰の影響

Study on Tolerance of Ultrasound Exposure (2nd Report) Effect of Nonlinear Attenuation

○星貴之(ピクシーダストテクノロジーズ株式会社)

◆近年、パラメトリックスピーカの用途展開や超音波フェーズドアレイの登場により強力な超音波がより身近なものになってきている。

➢ 従来の安全基準は1900年代半ばの少数の論文に依拠しており、それらの論文は20 kHz 周辺の低い周波数帯に関するものが多い。
➢ 近年応用が広がっている周波数帯は40 kHz 周辺の高い周波数帯であるため、改めて検討し直す必要がある。

◆前報では、140 dB SPL 未満の40 kHz 超音波が危険であるというエビデンスは現在のところ存在しないことを指摘した。

◆本報では、強力超音波が非線形吸収により減衰してやがて飽和する点に着目し、平面波の場合は音源から1.2 m、球面波の場合には0.4 m以上離れれば初期音圧によらず140 dB SPL を下回ることを示した。

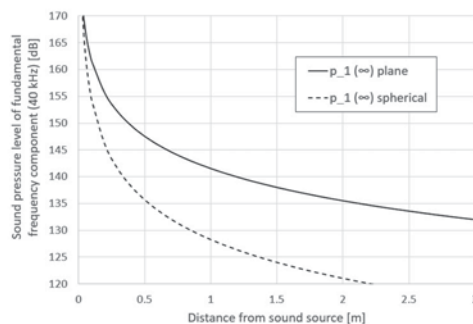


Fig. 1: Upper bound of plane and spherical waves of 40 kHz ultrasound.

1-7-6

1-7-6 超音波振動工具のトポロジー最適化

Topology optimization of large ultrasonic tools

○和田有司, 中村健太郎(東工大・未来研)

◆超音波接合等を目的とする振動工具の作用面を平坦に振動させることは産業応用上重要であるが、経験に基づくノウハウとなっている。

◆トポロジー最適化は設計空間内の部材配置を空隙を含めて決定する構造最適化手法であるが、超音波振動子への応用例は極めて少ない。

◆駆動面入力 $P=f.u$ 、制御対象面出力 $Q=g.u$ のとき最適化の目的関数を $\Theta=Re[P^*Q]$ としてこれを最大化するように材料配置を決定する。

◆モード重ね合わせ法を利用する際、上記の目的関数はモードごとに算出し、加算すると最適化が安定に進行する。

◆Fig. 1 が最適化の設計領域と最適な材料配置である。濃淡は要素内の材料の充填率を示している。

◆Fig. 2 が最適化前後の振動分布であり平坦な振動分布を得ている。

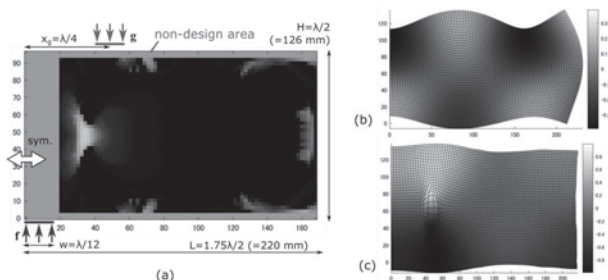


Fig. 1: (a) Design area and optimal material distribution, and vibration (b) before and (c) after optimization.

1-7-7

1-7-7 強力空中超音波照射によるソフトマテリアル表面温度の数値解析—空気流を考慮した数値解析—

Numerical simulation in temperature on soft material surface under high-intensity aerial ultrasonic wave

: Numerical simulation with air flow

☆菱沼真志, 大隅歩, 伊藤洋一(日大・理工)

- ◆強力超音波照射によるエネルギーを利用した技術に、HIFUによる癌治療やサーモソニック法を利用した非破壊検査がある。
- ◆これらの応用技術の一つとして、強力空中集束超音波(HIAFU)を利用した非接触加熱技術がある。一方、この技術を実用化するためには、音波照射時間と温度上昇、照射範囲と温度上昇分布等の関係を詳細に知ることが不可欠であり、そのための実験やシミュレーションによる検証が必要となる。
- ◆音響流を模擬し、空気流による放熱を考慮したシリコンゴムの表面温度上昇の数値シミュレーションの基礎検討を行い、実測との比較を行った。

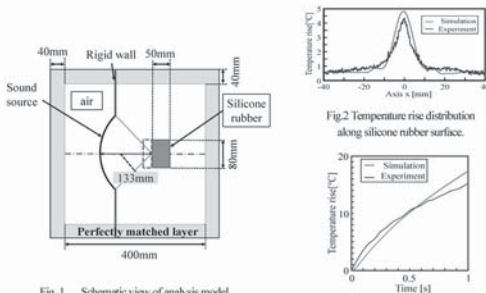


Fig.1 Schematic view of analysis model.

Fig.3 Temperature rise at sound wave focusing point

1-7-9

1-7-9 音響振動を用いた二成分流体の分離

Separation of two-components fluid using acoustic vibration

○渡邊海, △秋末薫, 上田祐樹(東京農工大)

- ◆音波による二成分流体の分離は、摺動部がスピーカーのみであることから構造が簡単で、製造、維持コストが深冷分離法や吸着分離法に比べ低く抑えられる利点がある。しかし研究報告が少なく、十分な知見が得られているとは言えない。
- ◆本発表では音波による二成分流体の分離を実験的に研究した結果について述べる。今回は円管内にヘリウムとアルゴンを1:1に混合させた気体を封入し、スピーカーから音波を入力して分離を行った。
- ◆経過時間に対するヘリウム濃度を Fig.1 に示す。時間が経過するにつれヘリウム濃度が変化していることから、分離現象が生じることが確認できた。

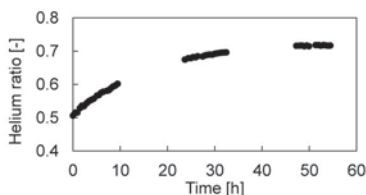


Fig.1 Changes in helium ratio over time

1-7-8

1-7-8 空中超音波フェーズドアレイ弾性波源走査法を用いたガイド波伝搬による金属薄板内欠陥の可視化

Visualization of defects in metal plate by guided wave propagation using airborne scanning phased array elastic wave source technique

☆清水鏡介, 大隅歩, 伊藤洋一(日大理工)

- ◆空中超音波フェーズドアレイ(Airborne Ultrasound Phased Array: AUPA)を用いた弾性波源走査法による非破壊検査について研究を行っている。
- ◆AUPAの放射音波に含まれるグレーティングローブは、AUPAを構築する超音波エミッタの素子間隔を変更することで抑制できる。
- ◆素子間隔を変更した AUPA を用いたガイド波伝搬による金属薄板内欠陥の可視化について検討した。
- ◆結果より、極めて精細な欠陥の可視化が可能であることを確認した。

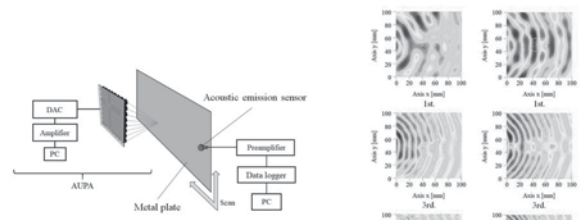


Fig.1: Measurement device.

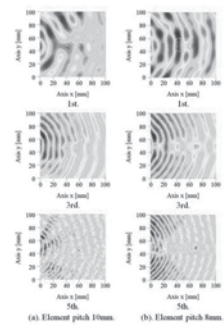


Fig.2: Guided wave propagation.

1-7-10

講演取消

1-7-11

1-7-11

空中超音波による触感の再現

Reproducing haptic feeling using airborne ultrasound

○篠田裕之(東大)

◆超音波による非接触での触覚再現技術が進展している。指先表面を計測しながら放射圧の場を生成する音響ホログラフィ技術、人間の触知覚特性利用し、圧覚を含む多様な触感を感じさせる技術、圧力だけでなく温冷覚をも提示する技術など、空中での触感再現の現状を紹介する。

◆Fig. 1 は、水のミスト雰囲気中で手掌部に集束超音波を照射したときのサーモグラフィである。超音波がミストの蒸発を促進することによってスポット状に皮膚を冷却し、冷覚を提示することができる。

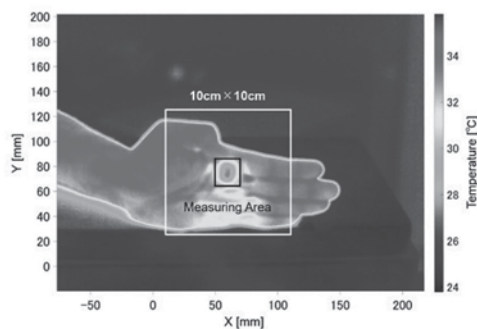


Fig.1: Producing cooling sensation by focused ultrasound.

1-7-13

1-7-13

強力集束超音波の治療応用と治療システム開発

Therapeutic application of high-intensity focused ultrasound (HIFU) and development of HIFU therapeutic system

○吉澤晋(東北大・工)

◆強力集束超音波 (HIFU: high-intensity focused ultrasound) を用いた治療では、体外で発生させた超音波を体内患部に集束させることで、非侵襲にがんなどを治療できる。近年では、数十 kW/cm² を超えるような非常に高強度な超音波パルスにより、瞬時に生体中に生成できるキャビテーション気泡群を利用した治療法の開発、臨床応用が進んでいる。このような治療法では、超音波イメージングが非侵襲な治療ガイドとして使われることが一般的である。気泡群は超音波イメージングにより可視化され、どこを治療しているかをリアルタイムに把握できる。キャビテーション気泡を利用した HIFU 治療の概念図を Fig. 1 に示す。このような治療システムについて紹介する。

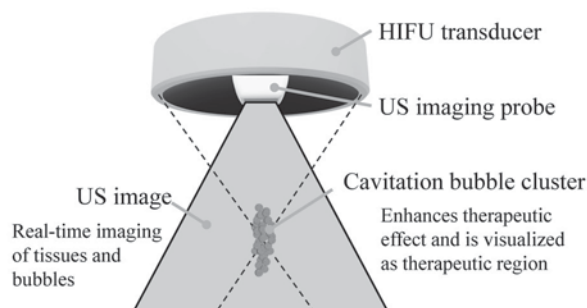


Fig.1: Schematic of bubble-enhanced HIFU treatment system.

1-7-12

1-7-12 音波照射加振とレーザドップラ振動計を用いた非接触音響探査法の最新動向

Latest trends in non-contact acoustic inspection method using acoustic irradiation induced vibration and laser Doppler vibrometer

○杉本恒美, 杉本和子, 中川 裕, 上地 樹 (桐蔭横浜大・院), 大歌川紀之, 黒田千歳 (佐藤工業㈱技術研究所)

◆音波照射による加振と高感度のレーザドップラ振動計 (LDV) を用いた非接触音響探査法により、遠距離から計測対象面の近傍に存在する空洞や剥離などを検出することが可能である。

◆本手法の利点は、音源とLDVの配置の自由度が高い点にあり、実際に、音源自体を無人航空機 (UAV) に搭載した計測も可能であることが明らかになっている。

◆今回は、NCAI 法の概要に加えて、橋梁下面 (高さ 30 m を超える高架橋)、地下空洞天井部における吹付コンクリート探査結果例 (離隔 25 m) および飛行中の UAV からの音波照射加振を用いた探査結果例など、NCAI 法の最新動向について紹介する。

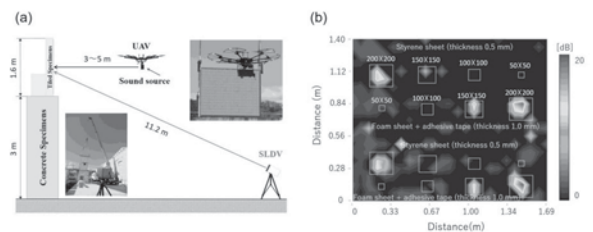


Fig.1: Acoustic irradiation induced vibration from UAV in flight. (a) Experimental setup, (b) Vibration energy ratio distribution.

1-7-14

1-7-14

改正 JIS C6790

ボルト締めランジュバン型超音波振動子の電力計法による負荷試験方法

Load test of a bolt-clamped Langevin vibrator using wattmeter method

○△高橋典久, △栗原利幸, △野本 智, 小長井孝 (超音波工業会)

◆JIS C6790 は、パワーアナライザなどの超音波電力測定器を用い、ボルト締めランジュバン型超音波振動子 (以下、BL振動子という) (Fig.1) の無負荷時と負荷時の入力電力を測定することによって、実用動作時におけるBL振動子の諸特性を算出する試験方法である。本規格の概要を紹介する。

◆この規格は、BL振動子の中心軸方向に振動する超音波縦振動子を利用する、共振周波数 10 kHz~100 kHz の振動子を対象としている。測定回路例を Fig.2 に示す。

◆算出する諸特性は次となる。
 ・制動損失 ・内部機械振動損失 ・音響出力 ・機械負荷抵抗
 ・電気音響変換効率



Fig.1: Bolt-clamped Langevin vibrator

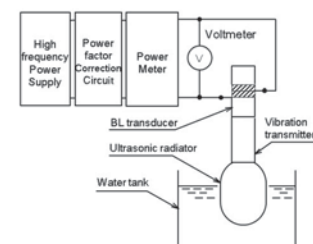


Fig.2: Example of input power measuring circuit of transducer

1-7-15

1-7-15 熱音響システムのエネルギーの有効利用に向けて

Research on thermoacoustic systems for effective use of energy.

○坂本真一(滋賀県立大), 渡辺好章(同志社大)

音響技術の応用は多方面にわたり、日常生活に必要な不可欠なものとなっている。様々なセンサ、デバイス等の進歩ならびに研究者らの地道な取り組みにより、音を力やパワーとして利用する強力超音波に注目が集まっている。強力超音波の応用は幅広く、その目的もさまざまである。

本報告では、熱音響現象を応用することによって、システム内に数kPaという非常に大きな音圧の音を発生させ、その大きな音エネルギーの利用を目指す熱音響システムについて記す。

熱音響システムの応用については、自然エネルギーや工場の廃熱などの未利用エネルギーを有効に活用し、新しい熱音響冷却システムや熱音響発電システムに取り組みされている。

熱音響システムの最大の長所は入力エネルギー源を選ばないことと考えられる。つまり、太陽熱エネルギーなどの自然エネルギー、内燃機関や工場などの廃熱を利用できる可能性を持っている。その他にも環境へ悪影響を与えるガスを用いる必要がないこと、可動部が無く構造が簡単なため信頼性が高いことなどが挙げられる。

1-7-17

1-7-17 集束型超音波音場における堅牢ハイドロホンの評価

Evaluation of Tough Hydrophone in Focused Ultrasound Fields

☆貝瀬不二丸(桐蔭横浜大), 椎葉倫久(日本医療科学大), 岡田長也(本多電子), 森下武志, 佐藤敏夫(桐蔭横浜大), 竹内真一(東京都市大)

- ◆私たちはこれまでに、平面振動子を用いた超音波洗浄器が形成する高強度超音波音場に 100 時間曝露しても受波感度が低下しない堅牢性を有したハイドロホンを開発した。
- ◆本研究では、HIFU を想定した凹面振動子が形成する集束型超音波音場に堅牢ハイドロホンを曝露し、堅牢ハイドロホンの評価を行った。
- ◆堅牢ハイドロホンを集束型超音波音場に1時間45分曝露しても受波感度は低下しなかった (Fig. 1)。また、音場に曝露している間に気泡がハイドロホンに向かって高速で移動の様子が確認できたが、堅牢ハイドロホンの前面板に傷は生じていなかった。

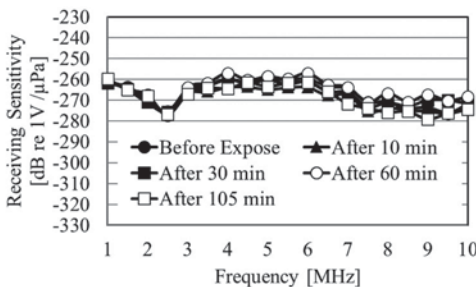


Fig. 1 Frequency characteristics of receiving sensitivity

1-7-16

1-7-16 超音波たわみ振動を用いた非接触型位置検出センサ

Ultrasound non-contact sensor using flexural vibration.

☆中岡夏海, 小山大介, 松川真美 (同志社大)

- ◆本研究では、超音波振動する板の電気インピーダンスを計測することで、直前の物体位置を検出する可能なセンサについて検討した。将来的には簡素な構成と高いロバスト性を活かし、非接触ディスプレイなどの空中操作技術への応用を目指している。
- ◆振動板近傍に物体(アルミニウム板)を設置し、物体位置の変化に伴う振動板の電流特性を測定した。
- ◆振動板・物体間の距離を変化した場合、空气中に定在波が発生する条件でインピーダンスが増加し、電流値は急激に減少した (Fig.1(a))。
- ◆物体を板長手方向に移動した場合、電流位相差が変化したことから、物体位置を一意的に決定できることがわかった (Fig.1(b))。
- ◆今後、異なる形状の物体位置検出について検討を行う。

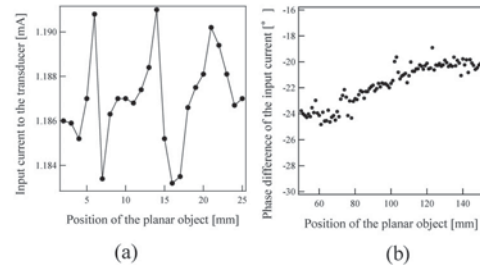


Fig.1: Changes (a) in the input current to the transducer and (b) in the phase difference of the input current with respect to the position of the planar object.

1-7-18

1-7-18 平行電界励振の高周波適用検討 - LiNbO₃ 回転 Y 板を用いた SMR 構造を有する厚みすべり振動子 -

Examination of high frequency application of parallel electric-field excitation —Thickness-shear resonators with an SMR structure using a LiNbO₃ rotated Y plate—

○長谷川洋, △尾崎卓美, △高橋芳美, △工藤泰之 (デクセリアルズ), △江口治, 中村信良

第5世代(5G)移動体通信規格の周波数フィルタでは、6GHz以下のSub 6(バンドn78, バンドn79)と呼ばれる周波数帯で広帯域素子が必要とされている。

LiNbO₃ 回転 Y 板を用いた平行電界励振による厚みすべり振動の基本波において、大きな電気機械結合係数が得られることを明らかにした。更に SMR (solidly mounted resonator) 型に適合させ、4GHzを超える共振特性を得ることができた。

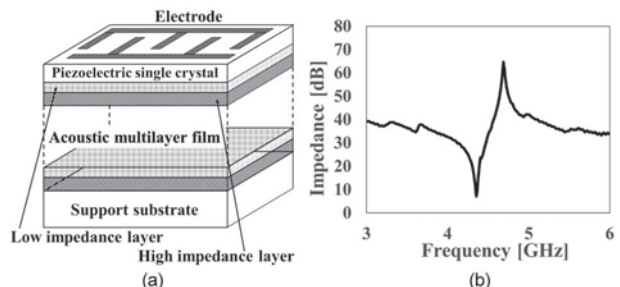


Fig.1 Parallel electric-field excitation thickness-shear vibration of an SMR type resonator using the fundamental wave. (a) Resonator structure. (b) Impedance response.

1-7-19

1-7-19 音響整合層を付加した 圧電高分子振動子の開発について

On the development of piezoelectric polymer [P(VDF/TrFE)] elements with an additional acoustic matching layer

○高橋貞幸(山形大学)

- ◆本研究では、PVDF系の圧電高分子材料である、P(VDF/TrFE)共重合体による振動子と、音響整合層としてポリエステルを利用した場合の効率の良い圧電振動子の開発を目的としている。
- ◆先行研究により、ポリエステル材が音響整合材として有望であることが判明している。しかし、整合層により空気中での超音波減衰は抑止するが、圧電振動子面とポリエステル間の接着層の影響のため、圧電振動子自体の性能は低下すると考えていた。最近の研究で、ポリエステルを付加したP(VDF/TrFE)圧電振動子は、接着剤による性能の低下が起こらないことが判明した。

Figure. 1に、P(VDF/TrFE)圧電振動子(厚み200 μ m)とポリエステル(接着層を含む厚さ95 μ m)の付加前後の動作特性を示した。

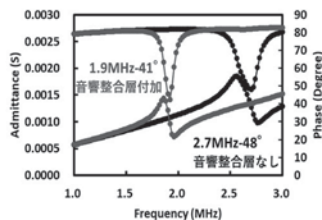


Figure 1

- ◆結果：圧電振動子の性能は接着層により低下すると思われていた。しかし、音響整合層の条件によっては、電気機械結合係数(k_t)値の変化(低下)がないことが分かった。

Fig.1 : P(VDF/TrFE) piezoelectric elements responses

2.7 MHz (normal) and 1.9 MHz(an acoustic matching layer (polyester film)).

1-8-2

1-8-2 Transformer エンコーダを用いたアクセント辞書 参照による L2 英単語発声の自動韻律評価

Automatic prosody evaluation of L2 English word utterances based on Transformer encoder referring to accent dictionary

☆鈴木 優, 加藤 恒夫, △田村 晃裕(同志社大院)

- ◆Transformer エンコーダを用いた英語母語話者の音素継続時間長、基本周波数、インテンシティの韻律特徴量の推定と、第二言語学習者による英単語発声の特徴量について、推定した母語話者の特徴量との差分をもとに、英語母語話者による韻律の主観評価値の推定を行った。
- ◆その結果、アクセント付き発音辞書から決定木により推定した母音継続時間長比を参照する従来手法に対して、母語話者による主観評価値との相関係数を大きく改善した。
- ◆音素継続時間長のみを韻律特徴量とする提案モデルに対して、母音継続時間長とインテンシティを韻律特徴量とする提案モデルではより高い相関係数を得た。一方、音素継続時間長のみを韻律特徴量とする提案モデルに対して、音素継続時間長に加えて基本周波数(F0)を韻律特徴量に組合せた提案モデルでは相関係数が低下した。

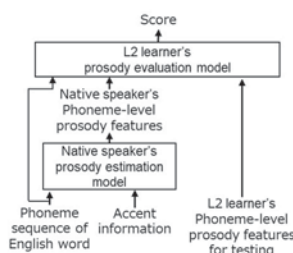


Fig. 1 proposed model

Table 1 the values of the correlation coefficient for each model

Method	Dur.	F0	Int.	ρ
Decision-tree	●	-	-	.198
Decision-tree	●	●	-	.324
Transformer	●	-	-	.519
Transformer	●	●	-	.476
Transformer	●	-	●	.536
Transformer	●	●	●	.488

1-8-1

1-8-1 文の発音難易度を用いた少数発話での 英語自動発音評価精度の向上

Improvement of Automatic English Pronunciation Assessment with Small Number of Utterances Using Sentence Pronunciation Difficulty

☆内城 沙月, 伊藤 彰則, 能勢 隆(東北大)

- ◆本研究では、短い発話での英語自動発音評価精度の向上を目指す
- ◆従来の DNN-HMM ベースの発音評価システムに文の発音難易度を考慮することで評価精度の向上を行った。
- ◆従来法から得られる機械スコアと文の発音難易度を表した文スコアを合わせて最終的な機械スコアを算出することで、文の発音難易度を考慮した自動発音評価システムの構築を行った。
- ◆結果、文の発音難易度を考慮した提案法では、短い発話において相関係数が最大 0.11 ポイント向上し、また従来の機械スコアに見られた文スコアの偏りを軽減させることがわかった。

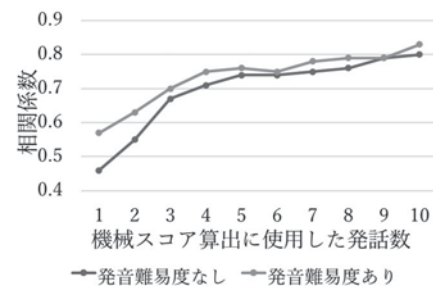


Fig.1: Correlation between the machine score and the human score with respect to the number of utterances for evaluation

1-8-3

1-8-3 防災放送のための 発話トレーニング法の比較

Comparison of voice training for broadcasts for disaster prevention

○高野佐代子, 土田義郎(金沢工大)

- ◆防災放送を行う一般職員の方に対し、短時間かつ効果的なトレーニング法を模索している。本研究では韻運動制限法(割り箸法)と日本語子音を欠落させ母音のみで発話させる母音法の比較を行う。
- ◆繰り返し練習→1回目録音→「トレーニング」→2回目録音→繰り返し練習→3回目録音の順で行う。「繰り返し練習」では1分間の繰り返し読みを行い、ある程度慣れることを目指し、その人の思う話し方に近い状態になることを意図する。
- ◆3週間に渡るトレーニングは音響的に一貫した傾向は確認できなかった。聴覚的には、ばらつきが出る結果となった。
- ◆短期のトレーニングでは、割り箸法は Fo の変化が顕著であるが、Loudness の変化は小さい。母音法は Fo の変化は小さいが、Loudness の変化がみられる。さらに3回目の録音(2回目の繰り返し練習)の後、さらに Loudness が上昇する傾向が見られた。

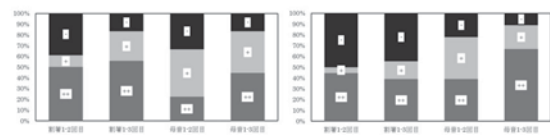


Fig. 1. The chopstick vs vowel method (Left) Fo differences, (Right) Loudness differences

1-8-4

1-8-4 Preliminary approach to the method of accent and intonation correction for learners of Japanese

Anna Sheldeshova and Sumio Ohno (Tokyo University of Technology)

Shadowing has approved its efficacy for self-studying among advanced learners of foreign languages, but remains difficult for beginners. This study aims to give a preliminary approach to the method of accent intonation correction for learners of Japanese in order to make it as a basis for the future development of an automatic feedback system.

Results of the experiment with learners of different levels of language proficiency showed common mistakes that are significant to the native speakers. Such mistakes can be tracked and commented automatically in order to make them clear for self-correction by learners.

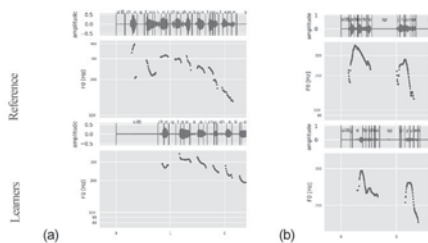


Fig.1: Examples of mistakes at shadowing exercise: read (a) at normal speed and (b) at slow speed

1-8-6

1-8-6 日本語とイタリア語における重子音に先行する母音の音響的特徴について

Acoustic properties of vowels preceding geminate consonants in Japanese and Italian

☆角田正路(上智大院), 守本真帆(国語研), 岩上恵梨, 荒井隆行(上智大)

- ◆日本語母語話者による日本語促音の発話では、促音に先行する母音の末尾にフォルマント遷移(FT)が観察されている(柳澤ほか2015)。
- ◆イタリア語母語話者による重子音の発話でも、重子音の先行母音末尾にFTを観察した(角田ほか2020)。
- ◆本研究では、角田ほか(2020)よりも多くの発話データを根拠としてFTが存在することを観察し、さらに日本語母語話者とイタリア語母語話者のL2におけるFTがL1から転移したものである可能性を、他の音響的特徴量と関連付けながら論じることを目的とした。
- ◆角田ほか(2020)と異なるイタリア語(L1)の発話でも、FTを確認した。
- ◆先行母音の音響的特徴量のうち、フォルマントそのものの値は日本語とイタリア語で違いがあるもののL1、L2では有意差がなかった。
- ◆日本語母語話者とイタリア語母語話者のL2でも観察されたFTは、L1から転移したものである可能性が考えられる。

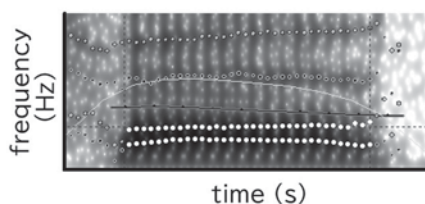


Fig.1: Formant transition at the end of vowel /a/ in the Italian word /fat: a/.

1-8-5

1-8-5 中国語学習者の複母音発音矯正のための音声分析

Analysis of acoustic features of compound vowel pronunciations for automatic quality evaluation of Chinese learners

☆齊藤裕仁, △陳淑梅, 大野澄雄(東京工科大)

- ◆本研究では学習者の発話した複母音の発音、強さ、発話時間について分析、評価し、学習者へのフィードバック方法の検討を行なう。
- ◆音声分析の手法としてフォルマント周波数を用いた分析を行ない、第三フォルマントまでを観察した。また、図1のような母音表形式へのプロットも行い、口の開き方や舌の位置の関係についても観察する。
- ◆音声データとして男性のネイティブ話者2名、女性のネイティブ話者2名の音声を録音し、学習者のデータとしては中国語の講義受講経験のある2名の男性の音声を録音した。
- ◆結果としてネイティブ話者による母音の発音に関しては丁寧な発音と自然な発音に大きな差はないと考えられる。ネイティブ話者と学習者発音の比較ではネイティブ話者は主母音を意識して発音している可能性が挙げられた。

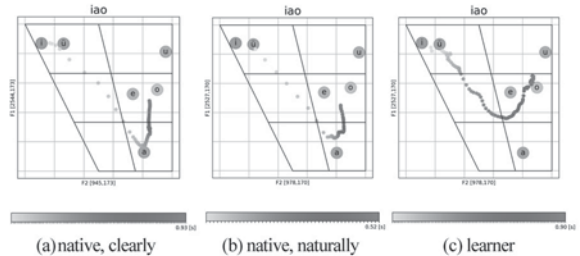


Fig. 1 Formant frequency movements on vowel diagram in utterance /iao/

1-8-7

1-8-7 会話がCOVID-19の感染リスクに及ぼす影響

Effect of Conversation on the Infection Risk of COVID-19 Transmission

○飯田明由(豊橋技科大), 吉永司(豊橋技科大)

SIR 数理モデルと Wells-Riley モデルを用いて COVID-19 の感染リスクを推定し、会話・発声が感染リスクに及ぼす影響について明らかにする。感染確率を示す Well-Riley モデルにおいて、エアロゾル排出量を表す変数φが感染リスクを考える上での主要因子となっている。会話や歌唱におけるφは、呼吸と比較して数十倍から数百倍大きいことから、会話が感染を広めていることを事例をまじえて紹介する。また、人工声帯を用いた実験により、発話時におけるエアロゾルの生成について示す。

$$P = 1 - \exp \left\{ -\frac{BT}{N_0} \frac{\rho\phi}{\lambda V} \left[1 - \frac{1}{\lambda T} (1 - e^{-\lambda T}) \right] \right\}$$

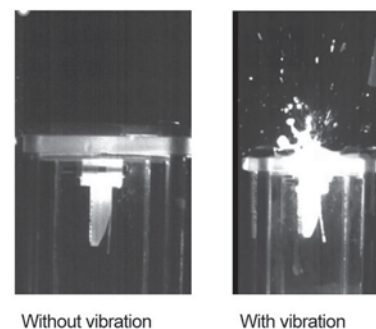


Fig.1: Aerosol distribution near artificial vocal fold

1-8-8

1-8-8 声道模型を用いた音声生成時における飛沫の可視化

Visualizing droplets using vocal-tract models of human speech production

○荒井隆行(上智大・理工)

- ◆新型コロナウイルス感染拡大に伴い、会話や歌唱時の飛沫の問題が注目されている。
 - ◆レーザー光を照射することで飛沫の可視化は実現されるが、人体への影響あり。
- ↓
- ◆声道模型を用いて音声の生成過程を模擬した上で可視化を試みた。
 - ◆BMW-RL モデル (Fig. 1 左) を用い、人工的に作られた模擬唾液を用いて口唇から飛沫が飛散される様子を可視化。
 - ◆口腔内圧が高くなる/b/において、飛沫が多く飛散することが確認された (Fig. 1 右)。

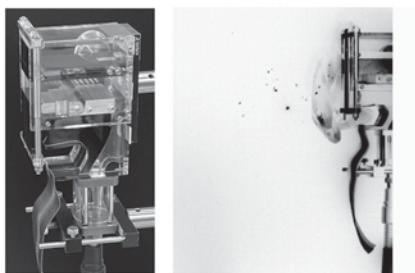


Fig. 1 : Experimental settings using the BMW-RL Model (left) and a snapshot after releasing its lip closure (right).

1-8-10

1-8-10 遠隔による第二言語コミュニケーション訓練のためのロボットアバター言語学習支援システム

Robot-Avatar-Assisted Language Learning System for Remote Training of Speaking in Second Language

☆名島大生(同志社大院), △村本大樹, 加藤恒夫, △田村晃裕, 山本誠一(同志社大)

- ◆新型コロナウイルス感染拡大により遠隔学習の重要性が増している。遠隔での英会話訓練向けに、ロボットアバター言語学習支援システム (Fig. 1) を開発した。
- ◆遠隔による訓練と対面による訓練の学習効果の違いを調査するため、対面でのロボット言語学習支援システム (Fig. 2) との 6 日間の比較実験を行った。遠隔での言語学習において、ロボットアバターを用いた場合とロボットを用いた場合の違いを調査するため、Zoom を用いた遠隔でのロボット言語学習支援システムとの比較実験を行った。
- ◆結果、ロボットアバター言語学習支援システム、対面でのロボット言語学習支援システム、遠隔でのロボット言語学習支援システムの順に学習効果が高いことを確認した。



Fig. 1: Screenshot of robot-avatar-assisted language learning system



Fig. 2: Screenshot of robot-assisted language learning system

1-8-9

1-8-9 知覚スコアに基づく遠隔操作コミュニケーションロボットの存在感コントロール

Presence control of remote-controlled communication robots based on perceptual score visualization

○荒川隆行, 高瀬遥矢, 杉本武士, 堀井省次, 蒲池恒彦(京セラ)

- ◆ウィズコロナ・ポストコロナのコミュニケーションにおいて、今後ローカルとリモートが混在する新しいスタイルが模索されていくと考えられる。
- ◆遠隔操作ロボットは、リモートコミュニケーションで失われやすい傾きやジェスチャと言った非言語情報を補う効果があり、リモートからローカルコミュニケーションに参加する場合に有効である。
- ◆ロボットの活用が期待される一方で、存在感を測定し、その量をコントロールするといった取り組みは少ない。これは存在感の定義が曖昧であることが一つの理由であると考えられる。そこで我々は、存在感を直接測定する代わりに、相手が自分をどれだけ知覚しているのかを知覚スコアとして計測、可視化することで、この知覚スコアを元にロボットを操作する方法を提案する。
- ◆本稿では、知覚スコアとして音に関する量に着目し、2つの可視化手法の検討を行った。1つは、マイクロホンアレイを用いてローカル話者の声の方向と音量を抽出し、可視化する手法。もう1つは、スピーカから拡声されたリモート話者の声が、ローカルにいる相手にどれだけ伝達されているのかを可視化する手法である。遠隔操作ロボットを介したコミュニケーションを実施し、アンケート評価を用いて本手法の効果を検証した。

1-8-11

1-8-11 隣接子音が母音声道形状に及ぼす影響の分析

Analysis of effects of adjacent consonants on vowel vocal tract shape

☆天野沢海, 後藤翼, 竹本浩典(千葉工大), 北村達也(甲南大), 能田由紀子, 前川喜久雄(国語研)

- ◆われわれはリアルタイム MRI 動画による日本語調音運動データベースの構築を行ってきた。現在、20 名の動画の任意のフレームから、発話器管 5 部位の輪郭を等間隔の点群として抽出することが可能となった。
- ◆本研究では、異なるフレームにおける発話器管の輪郭線を構成する点同士の平均距離を形態距離と定義し、形態距離がより小さい隣接子音が母音への影響が大きいと仮定して、隣接子音/k/と/m/が母音/a/, /i/, /u/の声道形状に及ぼす影響を男性 M1, 女性 F1 で定量的に評価した。
- ◆その結果、M1 では先行子音の影響が大きく、F1 では先行・後続にかかわらず/k/の影響が大きいことが明らかになった。



Fig1: Speech organ contours for /m/, /a/, and /k/ in /maka/ (left panel: M1, right panel: F1).

1-8-12

1-8-12 IPA を介した音素—調音データ変換のための IPA 継続長推定手法の検討

IPA duration estimation for phoneme-to-articulatory mapping through IPA

☆飯山智晴, 澤田隼, 大村英史, 桂田浩一(東京理科大)

調音運動は音声の生成過程に深く関与するため、音声合成のための中間情報として有用性が高い。Wei らは、言語特徴量と呼ばれる特徴量を介して音素列から調音運動のデータ(調音データ)を生成していた。しかし Wei らは前後の音素を考慮せずに言語特徴量を持つ調音情報を生成しているため、調音情報は十分に正確とは言えない。そこで我々は、前後音素から実際に発話される口形を表す国際音声記号(IPA)を推定し、IPA 系列から調音データを生成することを目指している。IPA 系列から調音データを生成するには、IPA と調音データの時間的な対応関係をモデル化する必要がある。すなわち、IPA の継続長モデルを構築する必要がある。そこで本研究ではBLSTMを利用してIPA継続長を推定する手法を検討する。また、IPA 全体を部分集合に分割し、集合ごとにネットワークを構築することを提案する。実験の結果を表1に示す。最も良好なRMSEを記録したのは分割方法2であった。このことから、提案手法によりIPA継続長推定精度が改善されることを確認できた。

Table 1 RMSE of estimated duration in each IPA subset

分割方法	部分集合	RMSE [ms]
1	IPA 全体 $\{ipa \mid ipa \in IPA\}$	47.5
2	母音, 子音, 半母音, 無音	31.9
3	各 IPA $\{\{ipa\} \mid ipa \in IPA\}$	36.6

1-8-14

ネコの鳴き声の言語表現について ノドを鳴らす声は「ゴロゴロ」か? -

How we express the throat sound cats make: Is it "gorogoro"?

○横山安紀子(日本大学)

異なる言語において、動物の鳴き声の表現についてはある程度の類似性が見られ、典型的な「鳴き声」として定着していると言える。例えば日本語では犬の吠え声は「ワンワン」、ネコの鳴き声は「ニャー」といわれている。ネコがノドを鳴らす声は辞典などで一般的に「ゴロゴロ」(/gorogoro/)ということばで表現されている。しかし、実際に発せられるノドの声は音声学的にかなり異なるものになる。本発表では実際のネコの声がどのように知覚されるのかということについて日本語母語話者を対象に2つの実験を行い、その結果を考察する。

実験では、はじめに実際のネコから採録した音声データを聞かせ(実験1)、次に同じ音声データを聞かせる際にネコの画像を提示した(実験2)。被験者には音刺激を聞いて、ことばとしてどのように表記をするか、また音の発生源はどこであるか、について回答してもらった。その結果、被験者は画像の提示があることにより、音の発生源が「ネコ(の声・音)」であるとする回答を多く示し、視覚情報が「ことば」の表現を判断することに影響を与えていることがわかった。これらの回答において表記を分析したところ、語頭子音 /g/ の増加が実験1から実験2にかけて見られ、第2音節子音では /r/ という回答が多く見られた。母音については /o/、/u/ とする回答が多数であった。形態については反復形以外のものも見られた。

このような結果から、聴者が与えられた聴覚刺激をことばとして表現する場合、付随する視覚情報があれば視覚情報による辞書的知識を活用する言語処理過程をとることが考えられた。それに加え、母音や形態において多様性が見られたことは、被験者各人がもつ知識の活用程度が一定ではなく、言語話者個人の言語知識や経験が関与することを示唆するものとなった。

1-8-13

1-8-13 電磁制御されたリード式人工喉頭を用いた短時間母音生成

Short-time vowel generation using an electromagnetically controlled reed-type artificial larynx

☆小泉正樹, 吉永司(豊橋技大), 荒井 隆行(上智大), △飯田明由(豊橋技大)

- ◆リード式人工喉頭はクラリネット等に用いられるシングルリードと類似した構造をもつ装置であり、送気することによりリードが自励振動し、喉頭原音を発する。人間の声帯の動作原理を模擬するという点において優れた装置として知られている。
- ◆本研究ではリード式人工喉頭に改良を加え、強磁性体製のリードを電磁ホルダにより任意のタイミングで吸着・固定することで、リード振動の電気的なON/OFF制御を実現した。
- ◆リードを固定した状態から短時間の間だけ開放し、その後再び吸着・固定する制御を行うことにより、リードを任意の回数振動させることに成功した。
- ◆声道モデルと組み合わせることで短時間の母音生成が可能となる。

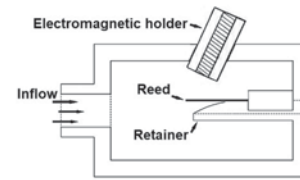


Fig.1 Electromagnetically controlled reed-type artificial larynx

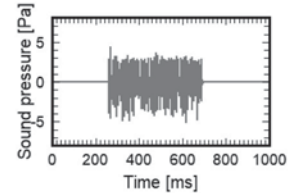


Fig.2 Short-time control response of artificial larynx for 50 vibrations

1-8-15

1-8-15 局部時間反転キメラ音声の了解度に影響を及ぼす周波数帯: 実験および誤答分析

Intelligibility of chimeric locally time-reversed speech: Relative contribution of four frequency bands revealed with confusion analyses

☆棟近光太郎(九大・芸工), 上田和夫(九大・芸工院), 松尾行雄(東北学院大・教養), △竹市博臣(理研・情報統合本部), Gerard B. Remijn(九大・芸工院)

- ◆4つの標的帯域 ($n=1, 2, 3, 4$; それぞれ 50-540, 540-1700, 1700-3300, 3300-7000 Hz) のうち、一部を局部時間反転(区間長 75, 150, 300 ms)により劣化したキメラ刺激を用いて音声の了解度を調べた。標的帯域のみを局部時間反転した条件(LTR- n)または、それ以外の帯域を局部時間反転した条件(ORG- n)の2条件を設定した。
- ◆区間長にかかわらず、第2帯域が保存されたときに音声の了解度が最も高くなり、劣化された時に最も低下した(Fig.)。
- ◆さらに母音・子音ごとに誤答分析を行ったところ、母音の了解度のみ帯域による特異性がみられた。

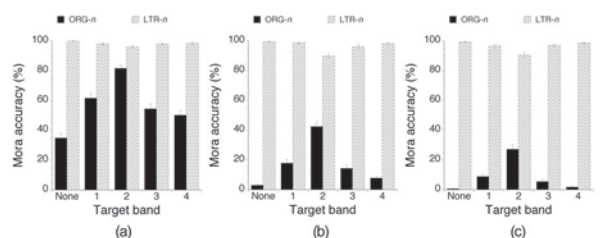


Fig. Mean percentage of mora accuracy ($n=18$) as a function of segment duration, target band, and stimulus types. Segment duration: (a) 75 ms, (b) 150 ms, and (c) 300 ms. The numbers 1-4 on the horizontal axis identify the four frequency bands from the lowest to the highest. The target band "none" refers to the conditions in which no band was preserved or degraded. Error bars, standard error of the mean (SEM). From Ueda and Matsuo (2021).

1-8-16

1-8-16 市松音声の知覚:聴覚体制化における断続音声との違い

Perception of checkerboard speech: Limitations in auditory organization contrasted with the organization for interrupted speech

◎川上 里以菜, 上田 和夫(九大), 竹市 博臣(理研)

- ◆時間・周波数平面に点在する音声の知覚手がかりがどのように体制化されるのかを調べるため、原音声の50%を削除した断続音声(時間軸で断続)と市松音声(時間軸および周波数軸で断続)の了解度を区間長と帯域数を変えて調べた(Ueda, Kawakami, and Takeichi, 2021)。
- ◆断続音声の了解度は、約100%から50%近くまで低下した(Fig.)。
- ◆市松音声の了解度は、20帯域では常にほぼ100%、2,4帯域では区間長160msで40%以下となり、聴覚体制化の限界を示した(Fig.)。

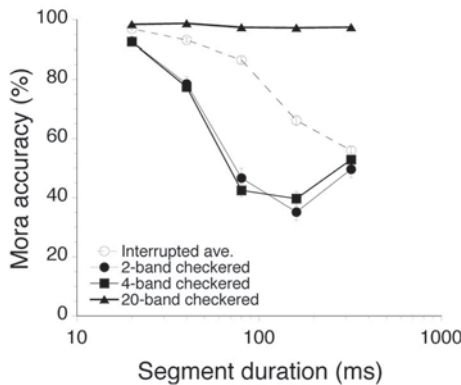


Fig. Mean percentages of mora accuracy (n = 20) for the interrupted speech and checkerboard speech stimuli as a function of segment duration and number of frequency bands. Error bars: standard error of the mean. From Ueda, Kawakami, and Takeichi (2021).

1-11-2

1-11-2 未知なるサイン音のデザイン方法に関する探索的検討 —設計過程の分析—

An Exploratory Research on Design of Unexperienced Auditory Signals —Analysis of the Design Process—

☆村木亜未香, △小越安奈, 山内勝也(九州大芸工)

- ◆サイン音を新たに設計する際の方法について検討するため、実験参加者にサイン音を設計させる実験を行った。
- ◆既存の音の模倣やアレンジになりにくいサイン音として、誰もが既存の音を知らない高度自動運転システムの権限委譲要請(TOR)に用いるサイン音を設計対象とした。
- ◆設計の際には、加齢性難聴を考慮するという要件を設け、模擬難聴システムを使用できるようにした。また、サイン音の秒数に関する制約も設けた。作業の様子は映像で記録し、レポートやインタビューと合わせて設計過程の分析を行った。
- ◆設計過程の分析をもとに作業フロー図を作成した。各設計者の設計過程には相違点があった。設計者の意図とそれに対応する作業についてフロー図より読み取ることができる(Fig. 1)。



Fig.1: Example of a design workflow

1-11-1

1-11-1 UIサウンドデザインのための9つのヒューリスティクス

9 UI sound Heuristics

和氣早苗(同志社女子大)

- ◆【目的】製品のユーザインタフェースに利用される音であるUIサウンドの評価の観点となる「ヒューリスティクス」を提案する。
- ◆【背景と概要】ユーザビリティ評価手法の一つであるヒューリスティック評価では、専門家が経験則(ヒューリスティクス)からの観点に基づいて、対象物のユーザビリティを評価する。ヒューリスティクスとしては「ヤコブ・ニールセンのユーザビリティ10原則」有名だが、今回これにない、UIサウンド評価のための「UIサウンド9原則」を提案する。
- ◆【活用】これを利用して、製品企画者、UIデザイナー、サウンドデザイナーなどの専門家が評価を実施できる。この9つの観点は、開発したUIサウンドの評価に使えるだけでなく、初期の問題発見、あるいは発音設計/サウンドデザイン時のガイドラインとしても活用できる。

UIサウンド9原則

1. UIサウンド(報知音、操作反応音)がそれを必要とする人に「聞こえる」。
2. 報知音が適切に「メッセージを伝える」。
3. 操作反応音が「操作の確実性と快適性を向上」させる。
4. 「一貫性を保ち標準に倣う」。
5. 「他メディアとの併用と補完」が行われている。
6. サウンドの「快適性が確保」されている。
7. サウンドで「感性価値」が付与されている。
8. サウンド利用の「柔軟性」がある。
9. ヘルプとドキュメントが準備されている。

1-11-3

1-11-3 未知なるサイン音のデザイン方法に関する探索的検討 —評価と統合的分析—

An Exploratory Research on Design of Unexperienced Auditory Signals —Evaluation and Integrated Analysis—

☆村木亜未香, △小越安奈, 山内勝也(九州大芸工)

- ◆サイン音を新たに設計する際の方法について検討するため、実験参加者にサイン音を設計させる実験を行い、設計過程の分析を行った(音講論集(秋), 1-11-2, 2021)。本稿では、実験で作成されたサイン音の評価実験、及び、設計過程と評価結果の統合的分析を示す。
- ◆評価実験では、設計者のレポートやインタビューから抽出した言葉に基づいて作成した評価尺度を含めることで、設計者の意図が反映されているか検証した。
- ◆加齢性難聴を考慮するという設計要件が満たされているか検証するため、模擬難聴処理を行った音も評価対象とした。
- ◆緊急感や覚醒度が高い刺激、「鋭い」「固い」「すんだ」「きれいな」印象を持つ刺激は、設計対象であったTOR場面に用いるサイン音としての適切性の評価が高いことが示唆された。
- ◆設計過程の分析をもとに作成した作業フロー図と評価結果を統合的に分析することで、各設計者が作成したサイン音が高評価あるいは低評価となった要因が推察された。また、同様の意識で設計されたサイン音でも評価はさまざまであったこと、設計者の意図が伝わったサイン音が必ずしも高評価を受けたわけではなかったことから、より具体的な設計ガイドラインやフィードバックの重要性が示唆された。

1-11-4

1-11-4 サイン音の継続聴取による
わずらわしさへの影響

Effects of continuous listening of auditory signs on annoyance

☆角野美帆, 山内勝也 (九州大・芸工),

△城戸恵美子, △山崎友賀, △山田亘 (三菱電機・デザイン研)

- ◆サイン音は、便益が少なく継続的に鳴っている場合に不快感やわずらわしさが増加することがある。
→繰り返し聴き続けてもわずらわしさを感じないような音にデザインすることが重要である。
- ◆サイン音は擬音語表現しやすいと覚えやすい[1]とされている。継続聴取するサイン音の場合は、覚えやすいデザインだと聴取するたびに印象が蓄積され、わずらわしさが増大する可能性がある。
- ◆以下の仮説を検証するため、主観評価実験を実施した。
 1. サイン音を継続聴取するとわずらわしさが増大する。
 2. 擬音語表現の一致度が高いほど、その変化量が大きくなる。
- ◆継続聴取実験の結果、いずれの刺激でもわずらわしさの有意な変化は観察されなかった。わずらわしさの評価値は音の特徴によって異なるが、最初に聴いた際の印象は継続聴取で変化しなかった。
- ◆サイン音を繰り返し聴くとわずらわしくなることが本研究で観察できなかった要因として、被験者が継続聴取という文脈を考慮せずに評価している可能性が考えられる。

[1] 岩宮, サイン音の科学, コロナ社, 2012.

1-11-6

1-11-6 報知音の協和感が警告感に及ぼす影響の
検討

Effects of degree of consonance on the auditory impression of warning of alarms

☆阿部理花子, 大塚翔, 中川誠司(千葉大)

- ◆警告効果の大きい報知音の設計を目指して、三和音の協和感が警告感に及ぼす影響を調査した。
- ◆構成音の周波数比を段階的に変化させることで協和感の異なる刺激音を作成し(Table 1), 16名の聴覚健常者を対象とした一対比較にて評価した
- ◆先行研究等から得られた予想に反し、協和感が低いほど警告感スコアは有意に小さくなった (Fig. 1).

sound	周波数 [Hz]			周波数比
A	420	630	840	2 : 3 : 4
B	420	525	630	4 : 5 : 6
C	420	490	560	6 : 7 : 8
D	420	472.5	525	8 : 9 : 10
E	420	462	504	10 : 11 : 12
F	420	455	490	12 : 13 : 14
G	420	450	480	14 : 15 : 16

Table 1: Triads used in the experiment

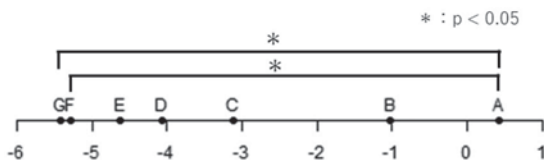


Fig.1: Averaged scale value of warning sensation

1-11-5

1-11-5 アンケート調査による静音車接近通報音の
現状と求められる要素の把握

Questionnaire survey on the current status and required elements of approach informing sound of quiet vehicles

○鶴田真理子, △辻永智也, △駒場雅貴, △野竹惟仁,
長谷川光司(宇都宮大)

- ◆電気自動車やハイブリッド車などの静音車に搭載されている接近通報音に関して、大学生を対象にしたアンケート調査を行った。
- ◆接近通報音の認知度は43.4%で、2016年の調査(39.7%)からわずかに上昇したが、未だ認知度が低い現状が明らかになった。
- ◆接近通報音には「聞き取りやすさ」「警戒感」が「自動車らしさ」「耳にした際の快さ」に比べ優先的に求められる傾向にあった (Fig. 1)。
- ◆自動車への興味や運転頻度の高さが接近通報音に求める要素に影響を与える可能性が示された。
- ◆自由記述では「老若男女に分かりやすい」「加齢による聴力低下や障害があっても伝わりやすい」などが接近通報音に求める要素として挙げられた。

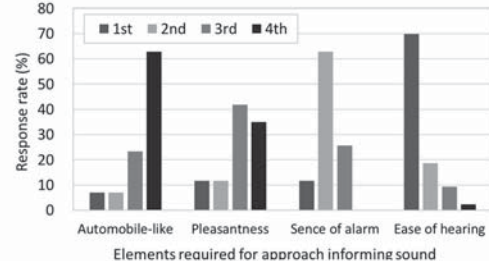


Fig.1: Rank of required elements for approach informing sound (Respondents who did not know the sound)

1-11-7

1-11-7 セカンダリアクティビティと警告音の差異
による自動運転引継時の行動評価

Behavior evaluation based on the difference of secondary activity and warning sound at takeover request

☆五十部健太(中央大院), 戸井武司(中央大)

自動運転中は運転動作とは別の作業、すなわちセカンダリアクティビティ(SA)を行うことが可能になる。本研究では、SAの種類と警告音の緊急感の差異による自動運転引継時の行動評価を行う。

まず、周波数、吹鳴周期、音圧の3つのパラメータを変更した警告音について、どのパラメータが寄与しやすいか評価する。吹鳴周期の差異により、評価点が大きく変化するため、緊急感には吹鳴周期の寄与が大きいことを把握した。

次に、評価点の大きく異なる吹鳴周期2Hzと6Hzを警告音に選定し、シミュレータによる引継時の行動評価を行う。Fig. 1に自動運転中のSA無を基準とした課題ごとのローレンツプロットの面積比を示す。計算課題は集中力が必要となり、交感神経が優位になり面積比が小さく、動画課題はリラックスしやすいため副交感神経が優位になり、面積比が大きくなったと考えられる。Fig. 2に警告音提示から手動運転への移行時間を示す。SA無が遅く、計算課題が早くなることから集中力を維持することで、手動運転へスムーズに移行が可能になると推察される。

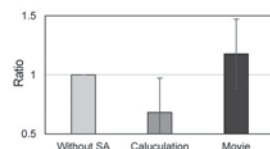


Fig. 1 Ratio of LPS

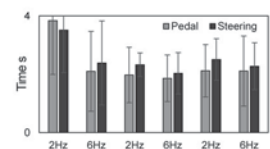


Fig. 2 Response time

1-11-8

感触と操作音の複合刺激によるロータリースイッチ操作感覚のオノマトペによる評価

Onomatopoeic expression for rotary switch operating feelings composed of tactile sensation and operation sound

☆宮入徹(中央大院/都産技研), △下村尚登, △白坂剛(アルプスアルパイン), 戸井武司(中央大)

- ◆ 感触と操作音の複合刺激から構成されるロータリースイッチ操作時の印象に対して、自由回答によるオノマトペ評価を行った。
- ◆ 実験結果に対して計量テキスト分析を行い、操作時の印象から連想されやすいオノマトペの傾向について調査した。
- ◆ 対応分析の結果から、操作音の音質およびクリックトルクの大きさによって、用いられる語の傾向が異なることがわかった (Fig.1)。

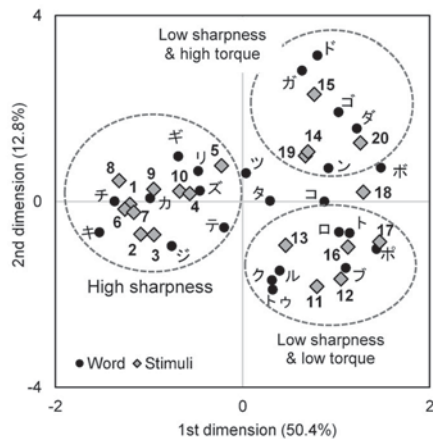


Fig.1: Correspondence analysis results

1-11-9

1-11-9 操作者の嗜好を考慮した快適なキーボード操作感覚の創生

Creation of the comfortable keyboard operation feeling in consideration of the preference of the operator

☆藤倉圭(中央大学院), 戸井武司(中央大学)

先行研究では使用状況に応じた操作感覚の提示により操作時の快適感の向上が示された。しかし、操作者の嗜好に応じた操作感覚は明確になっていない。そこで本研究では、操作者の嗜好を分類し、嗜好に適した操作感覚を提示することで更なる快適感の向上を目指す。

Fig. 1に嗜好を考慮した複合刺激時の主観評価を示す。各クラスターA、Bで好まれる操作音を選定し実験を実施した。結果より、各クラスター用に選定した操作音の快適感が最も高く、他の操作音では快適感が低下した。また操作音1は重厚感が高く、操作音2は明瞭感が高いことからクラスターAでは重厚感の高い音が好みの傾向があり、クラスターBでは明瞭感の高い音が好みとわかる。

Fig. 2に感触を考慮した複合刺激実験の主観評価を示す。動作圧と押下圧の異なる3種類の感触F1~F3を用いて同様の実験を実施した。結果より動作圧の大きさが押下圧の大きさを下回るごとに重厚感が低下することがわかる。

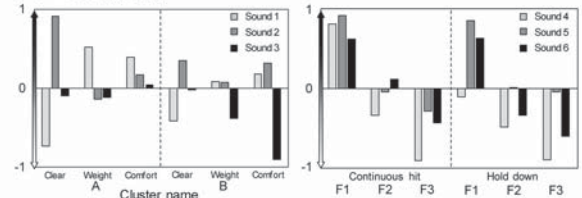


Fig. 1 Evaluation results of complex stimuli by SD method

Fig. 2 Evaluation results of complex stimuli by SD method

1-11-10

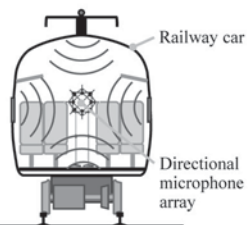
1-11-10 高臨場な音響再生システムを活用した聴感に基づく音響設計手法の検討

Study on sound design method based on auditory sensation utilizing high-spatial sound field reproduction system

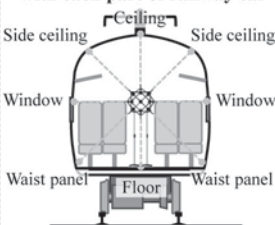
◎富田齊央, 田部洋祐, 山内源太(日立)

- ◆ 本報告では、開発上流で聴感に基づく騒音評価を実施することによって音響設計を高度化させることを目的に考案した、高臨場な音響再生システムを利用した聴感に基づく音響設計手法について述べる。

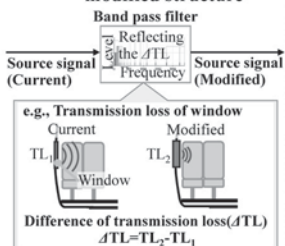
STEP1: Record the sound of current railway car



STEP2: Associate the recorded sound with each part of railway car



STEP3: Generate the sound of modified structure



STEP4: Evaluate the sound of modified railway car

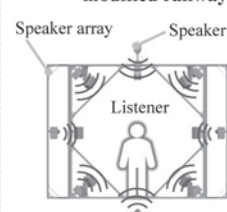


Fig. 1 Flow of the sound design method based on auditory sensation

1-11-11

1-11-11 Musical Chills の生体反応への影響と聴感印象に関する検討

Effect of pleasurable musical chills on driver's physiological response

○亀山勇希, 谷本典之, 村尾天地, 須藤季砂, 谷田部笙太, 石光俊介(広島市立大), △仙井浩史, △菊地敏博, △高橋尚子, △大津千佳(マツダ株式会社)

- ◆ 近年、自動車室内の快適性、静粛性の向上、カーオーディオの高音質化が進んでいる。それより、運転中に高音質な音楽を楽しむことが可能となった。

◆ 本稿では、生演奏のような臨場感のある音楽、高音質なオーディオでの音楽聴取により感じる可能性があるゾクゾク感(Pleasurable Musical Chills)に着目した。そして、音質の違い及びゾクゾク感が生体反応にどのような影響を与えるかを調査した。

◆ 生体反応として、脳波、心拍及び呼吸について解析した。また、主観評価として、アンケート及びゾクゾク感を感じた際のボタン操作を用いて評価を行った。

◆ 結果として、運転中の音楽聴取によって交感神経系が活性化することが示唆された。また、音源位置が明確である「一般的なスピーカ」での音楽聴取時に最も交感神経系が活性化することが分かった。

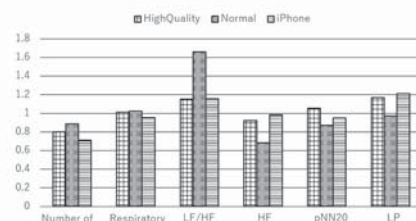


Fig. 1 Results of number of pressing button, respiratory analysis and HRV analysis

1-11-12

1-11-12 音の大きさの評価における男女差に評価条件が与える影響

Effect of evaluation conditions on the gender difference in loudness perception

☆菊地 凌雅, 鶴田(濱村) 真理子, 長谷川 光司 (宇都宮大学)

- ◆本研究では、音の大きさの評価における男女差に実験条件が与える影響について詳細に検討する。
- ◆先行研究とは異なる評価条件を設けたマグニチュード推定法とマグニチュード産出法を用いた音の大きさの評価実験を行い、ステープンスのべき法則の α の値における男女差の有無を検討した。
 - マグニチュード推定法実験: 評価値の範囲を0より大きく20以下に限定した。
 - マグニチュード産出法実験: 先行研究の基準音の設定方法を参考に、その基準音の大きさの ± 10 dBの2条件を加えた計3種類の基準音を用いた。
- ◆先行研究と同様にマグニチュード推定法 (Fig. 1) とマグニチュード産出法 (Fig. 2) とともに明確な男女差は認められず、評価条件は音の大きさの評価における男女差に影響しないことが示唆された。

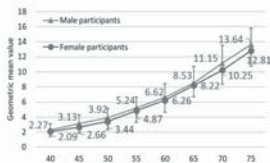


Fig. 1: Geometric means and standard deviations of loudness values assigned by male and female participants for pink noise in the experiment using magnitude estimation method

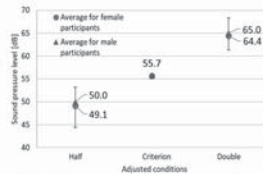


Fig. 2: Average sound pressure levels for men and women's adjustments to halve or double the loudness of the reference sound (55.7 dB) in the experiment using magnitude production method

1-11-14

1-11-14 せせらぎの主観的好ましさに及ぼすシャープネスの影響: 複数音源間での比較

Effects of sharpness on the subjective preference of murmuring of a stream: Comparison among multiple sources

☆高橋貴行, 大塚翔, 中川誠司 (千葉大)

- ◆シャープネスが聴感印象に与える影響を複数の音源 (3種類のせせらぎ, 波が打ち寄せる音, コップに炭酸ジュースを注ぐ音) において評価した。
- ◆再生速度を1/3~3倍に変化させることで、シャープネスを段階的に変化させた7種類の刺激音を各音源について作成した (Table 1)。また、Scheffeの一对比較法を用いて各刺激音の“好ましき”を評価した。
- ◆シャープネスの変動に伴い、刺激音の“好ましき”も変化した。多くの先行研究と異なり、シャープネスが一定値より小さい場合には、却って好ましきが低下する傾向がすべての刺激音に対して認められた。
- ◆シャープネスが小さい刺激音で好ましき低下した理由として、現実的なシャープネスを逸脱したために不自然さが増してしまっ可能性や、ラウドネスの影響が考えられる。

Table 1: Stimuli used in the experiment

Stimulus	Playing speed [times]	Sharpness value (Murmuring of a stream) [scum]	Sharpness value (The sound of waves crashing into the shore) [scum]	Sharpness value (The sound of pouring carbonated juice into a cup) [scum]
A	1/3	1.05	1.00	0.98
B	1/2	1.29	1.31	1.14
C	2/3	1.46	1.60	1.27
D	1	1.68	1.90	1.47
E	3/2	1.89	2.03	1.68
F	2	2.02	2.14	1.89
G	3	2.23	2.56	2.27

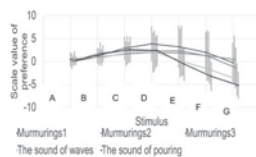


Fig. 1 Averaged scale value of the preference for the stimuli.

1-11-13

1-11-13 BGMが学校教育環境に及ぼす効果 -複数中学校における現場比較-

Effects of background music on the school educational environment - Field comparison in multiple middle schools -

○小松正史 (京都精華大学メディア表現学部)

- ◆教育機関 (中学校) の環境改善を目的とした背景音乐 (BGM) の活用の方を模索するため、複数中学校の現場調査を報告する。
- ◆SD法の採用人数は、K中学校が191名、S中学校が145名、T中学校が284名、I中学校が399名であった (いずれも福岡県内)。
- ◆3校に共通した評価傾向として「気分を変えるスイッチ」の効果がBGMに備わっている可能性が考えられる。
- ◆4つの中学校において、共通したBGMの評価傾向がみられた。
- ◆今回の調査結果を反映させたBGMの楽曲制作をはじめ、活用方法を具体化していくことが必要。

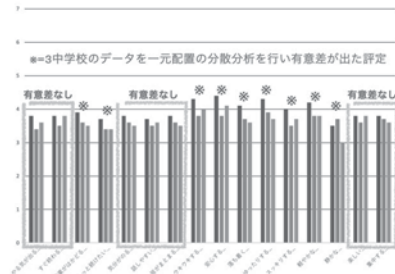


Fig. Grade Point Average Comparison

1-11-14

1-11-15

1-11-15 音環境のアノニマス性その2: カフェの文脈に関する検討

"Anonymity" in the sound environment part 2: Study on café context.

☆大庭駿平, 小橋侑生 (神奈川工科大), 三浦貴大 (産総研), 土田義郎 (金沢工業大), 田中哲雄, 上田麻理 (神奈川工科大)

前報では、大学図書館の音環境のアノニマス性について、構成要因や現況を検討した。一方で、このようなアノニマス性は、場所によって異なると予想される。

そこで本研究の大目的を、図書館のみならず、様々な空間を対象とした音環境のアノニマス性の構成要因を明らかにすることとする。本稿ではカフェを対象として、サウンドスケープ調査及び、アノニマス性の構成要因の詳細を質問票調査により探った。



Fig.1 Café scenery

1-11-16

1-11-16 食品売り場での宣伝コンテンツにおける音の印象評価方法の一考察

An impression evaluation method for sound of promotion content set at food corner in supermarket

☆於勢真紀子, 脇田由美(大阪工大)

◆近年、食品売り場などで頻繁に見かけるサイネージでは、食品動画とともに食品に関連する音(食関連音)が流れているが、これらの音が宣伝効果に貢献しているかは疑問である。

◆本実験では、スーパーマーケットの売り場騒音下で、食関連音がどのような印象で聞かれているのかを、特に同定と印象との関係に注目し、騒音下で3種の食関連音にて聴感実験を行った。

「何の音に聞こえるか」の回答が、一部でも正解(同定あり)だった場合と全く不正解(同定なし)だった場合の、「どのくらいおいしそうか」という問いについて、Fig.1に示す。図のバーは標準偏差で、缶で同定なしの回答は数が足りず標準偏差が出せなかった。焼く音と缶の音で、同定ありの方が評価平均値が高い結果になり、スーパーマーケットの売り場環境では、食品を同定した方がおいしそうに聞こえることが確認できた。

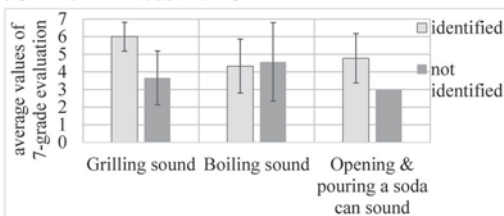


Fig.1: Comparison of the 7-grade impression evaluation for between identified food sound and not identified food sound

1-3P-1

1-3P-1 複数話者音声認識における強調信号の選択的利用に関する検討

Switching Between Enhanced and Observed Signals for Overlapping Speech Recognition

◎佐藤宏, 落合翼, デルクロアマーク, 木下慶介, 森谷崇史, 加茂直之(NTT)

◆概要

・複数話者の混合音声の音声認識に対して、目的話者抽出などの音声強調は有効であるが、抽出処理で生じる非線形歪みがかえって音声認識に悪影響を与える可能性もある。

・強調音声と、元の観測信号と、どちらが音声認識にとって有利かをSNR(signal to noise ration) およびSIR(signal to interference ratio)ごとに調査したところ、観測信号の方が性能がよい場合があることが分かった。(下記は文字誤り率での評価結果)

(a) 強調信号				(b) 観測信号			
SIR [dB]	SNR [dB]			SIR [dB]	SNR [dB]		
[dB]	20	10	0	[dB]	20	10	0
10	5.8	7.8	24.9	10	16.5	13.8	20.4
15	5.4	7.6	23.6	15	7.9	7.5	15.9
20	6.3	8.9	23.2	20	5.3	6.0	14.8

・推定したSNRとSIRを元に強調音声と観測信号どちらを用いるかを選択し、必要な時だけ音声強調する手法を提案することで、性能向上を実現した

(c) 提案法				(d) (a)からの相対改善率			
SIR [dB]	SNR [dB]			SIR [dB]	SNR [dB]		
[dB]	20	10	0	[dB]	20	10	0
10	5.8	7.8	21.1	10	0%	0%	15%
15	5.4	7.3	17.3	15	0%	4%	27%
20	6.1	7.3	17.0	20	2%	18%	27%

1-11-17

1-11-17

運転パフォーマンスに影響を与える音環境の検討

Study on sound environment that affects driving performance

◎須藤季砂, 村尾天地, 谷本典之, 亀山勇希, 石光俊介(広島市立大) 仙井浩史, 菊地敏博, 高橋尚子(マツダ株式会社)

◆音環境

➢ マインドフルネスに着目したフィードバック(FB)音やBGMを準備

◆トラッキングタスク実験

➢ 音環境が集中力を過渡に必要なタスクに与える影響調査

◆運転タスク実験

➢ 音環境が運転を想定したタスクに与える影響調査

◆回復タスク実験

➢ マインドフルネスがタスク後に与える影響調査

◆評価方法

➢ 脳波解析, 呼吸解析, HRV解析による生体反応の評価

➢ タスクスコアの評価

◆実験結果

➢ 操作に対するFB音 ⇒ タスクに意識を向けさせる効果

➢ 呼吸FB音 ⇒ タスク後のパフォーマンス向上



Fig.1: トラッキングタスク



Fig.2: 運転及び回復タスクの実験イメージ

1-3P-2

1-3P-2 環境の違いに頑健な話者年齢推定のためのドメイン適応法の検討

Study of unsupervised domain adaptation for robust speaker-age estimation under mismatched recording conditions

◎依直弘, 小川厚徳, 北岸佑樹, 神山歩相名, 井島勇祐(NTT)

◆学習時と運用時で録音環境が異なる場合に生じる Deep neural network (DNN) 話者年齢推定モデルの性能低下を解決するため、新たな教師なしドメイン適応を提案

◆提案手法では年齢ラベルつきソースデータと、年齢ラベルなしターゲットデータを用いて以下の新たな目的関数を導入することで、入力音声の分布の違いに頑健なDNN 音声年齢推定モデルを実現

➢ Contrastive loss により同一話者内の変動を抑制

➢ Local maximum mean discrepancy 基準で中間特徴量の分布間距離を最小化することにより、入力音声のドメインの違いを吸収

◆In-house データを用いた年齢推定実験により、雑音、残響、収録機器の違いに対し、提案手法は性能悪化を抑制することを確認

Table 1 Mean absolute errors under mismatched conditions

Method	Clean→Noise	Reverb→Noise	Clean→Reverb	Noise→Reverb
Source only	10.58	10.23	21.60	20.07
Rev-Grad	10.65	10.12	16.33	17.86
MMD	9.92	10.19	10.11	10.50
LMMD (proposed)	9.19	9.97	9.92	9.84
Target	7.86		9.06	

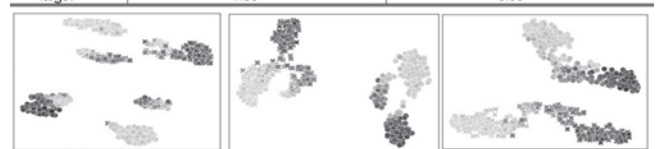


Figure 1 Visualization of the output from the models of utterances of Clean (blue) and Reverb (orange) datasets

1-3P-3

1-3P-3 入力音声データの拡張と認識結果の統合による音声認識率改善の検討

Improving the accuracy of speech recognition by augmentation of input speech data and integrating acoustic model outputs

◎ 早川大智(東芝 研究開発センター)

- ◆目的: Connectionist Temporal Classification を用いた音響モデルの半教師あり学習に向けて、音響モデルを変えなくても、より高品質な疑似ラベルを生成する技術を開発。
- ◆ポイント: 入力音声とともに、入力音声をデータ拡張して生成した複数の音声を単一の音響モデルに入力し、それぞれの認識結果を統合。
- ◆方法: 上記音声からそれぞれ音響モデルの出力(以下、音響スコア)を取り出し、音響スコアから貪欲法で得られたラベル列をマージするよう、音響スコアの統合を実施。
- ◆結果: CSJの評価セットで文字誤り率(Character Error Rate, CER)を評価。元の音声やデータ拡張で生成された音声を入力したときの認識性能と比較して、提案手法では優れた認識性能を達成。

	CER		
	Eval 1	Eval 2	Eval 3
A: Original	14.30	11.54	7.71
B: Speed x0.9	14.41	11.75	7.80
C: Speed x1.1	14.86	12.06	8.04
D: Pitch -100 cent	14.82	12.15	8.10
E: Pitch 100 cent	14.49	11.92	8.07
[Proposed Method]			
F: Combine(A+B+C)	13.89	11.32	7.43
G: Combine(A+D+E)	13.91	11.38	7.53
H: Combine(F+G)	13.83	11.32	7.41

Table 1: CERs on three CSJ evaluation sets.

1-3P-5

1-3P-5 音声仮名化のための加工音声重畳と非可逆性評価

Irreversibility evaluation of speech superposition for speaker pseudonymization

☆甲斐優人(都立大), 高道慎之介(東大), 塩田さやか, 貴家仁志(都立大)

- ◆本研究では、匿名化領域においても話者照合が可能な音声データの仮名化に着目し、仮名化された音声同士を重畳することで非可逆性の特徴を持つ加工音声重畳法を提案する。
- ◆複数のタスクに基づき提案法の仮名化性能を評価した結果、全てのタスクにおいて従来の信号処理に基づくベースラインを上回った。
- ◆非可逆性の評価において、敵対者は仮名化音声を持っており、かつ音声加工法が既知である条件下で、Fig. 1のフロー図に従って仮名化音声から原音声の復元できるかを調査した。
- ◆実験結果から復元したスペクトログラムは原音声とはかけ離れており、話者照合率の悪化がかなり低下したため、提案法は高い非可逆性を持つことが確認できた。

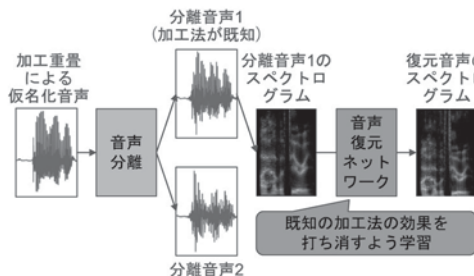


Fig.1: Reconstruction of spectrogram of original speech from superposed speech by using DNN based source separation and reconstruction network

1-3P-4

1-3P-4 相補的な複数のニューラル言語モデルと文脈情報を用いた繰り返しラティススコアリング

Iterative lattice rescoring using complementary neural language models with contextual information.

○小川厚徳, 浅見太一, 俵直弘, 荒木章子 (NTT)

- ◆前向き(forward)または後向き(backward)の、LSTMまたはTransformerに基づく言語モデルを用いた、繰り返しラティス生成によるリスコアリング(繰り返しラティススコアリング)手法を提案する。
- ◆これらのニューラル言語モデルは互いに相補的であるため、これらを順に適用していくことで、認識誤りを徐々に削減することができる。
- ◆講義音声などの一連の発話に対しては、以前の発話に対するリスコアリング結果を文脈情報として引継ぎ、現在の発話のリスコアリングに用いることで、リスコアリングの精度を更に向上させることができる。
- ◆CSJ講義音声コーパスを用いた実験で、最大で8個のニューラル言語モデルを前後方向に文脈情報を用いつつ順に適用することで、元の音声認識結果から24.4%の誤り削減を達成した。

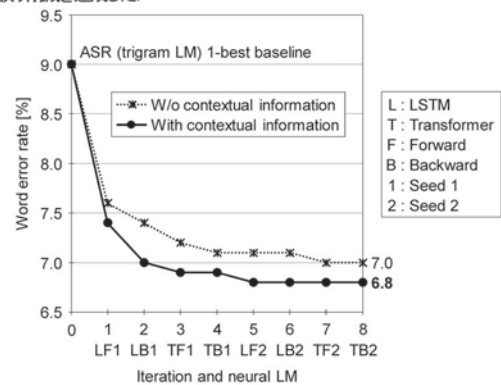


Fig. Eight iterative lattice rescoring with and without using contextual information.

1-3P-6

1-3P-6 非定常音のための基底分解型非負値行列因子分解に関する基礎検討

Basic study on basis-decomposed nonnegative matrix factorization for nonstationary source

☆大島風雅, 中山仁史(広島市大)

非負値行列因子分解(NMF)は非負値行列を対象とした行列分解の一種であり、定期的なスペクトルパターンを持つ楽器信号を低ランク近似することができる。その一方で、音声のような非定期的かつ連続的にスペクトルパターンが変化する信号に対しては課題が残る。

本稿では、このような信号に対しても低ランク近似可能な基底分解型非負値行列因子分解(BD-NMF)の提案と検討を行った。そのために、基底行列をケプストラム分析し、再合成することでFig.1のように音声における調音情報と音源情報に分解できることを確認した。

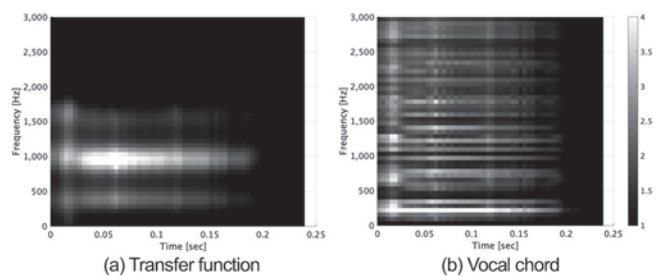


Fig. 1 Spectrogram that each quefrency components of the basis matrix multiplied activation matrix.

1-3P-7

1-3P-7 VocalTurk: クラウドソーシングを用いた話者照合の性能調査

VocalTurk: Exploring Feasibility of Crowdsourced Speaker Identification

©斎藤 1,2, 井手悠太 1, 中野鐵兵 1,2, 小川哲司 1 (1早大, 2ifLab Inc.)

◆クラウドソーシングを用いた音声データラベリングの実用可能性を検討した。

実験環境: Amazon Mechanical Turk
実験タスク: CMU-ARCTIC Dataset 発話データを対象とした話者照合

RQ1: 高品質な回答を行えるワーカーの割合はどの程度で、どう抽出可能か?

- 1vs1 話者照合タスク x24 問 (Fig. 1) において、ワーカー3,854 人中 34.5% が全問正解、87.6% が 7.5 割以上正解
- 正答数が少ないワーカーほど、自明な (指示を理解すれば誰でも分かるような) 照合ペアでさえも自信を示さない傾向が見られた
→ 自明なペア + 確信度を問うことで、早期に低品質ワーカーを選別可能

RQ2: 期待できるデータセット品質と、適切なタスク UI はどのようなものか?

- 上記実験で全問正解したワーカー351 人の、より難しいタスクへの回答精度は全タスク平均 $\geq 95\%$ (最低 68.1%)、多数決 3、7 人の場合 $\geq 97.5\%$ 、 $\geq 99\%$
- 話者照合タスクの UI において提示する比較音声数を 1vs1、1vs3、1vs5 と増やしていくと、回答精度は保ちながら最大 33% のタスク完了時間短縮

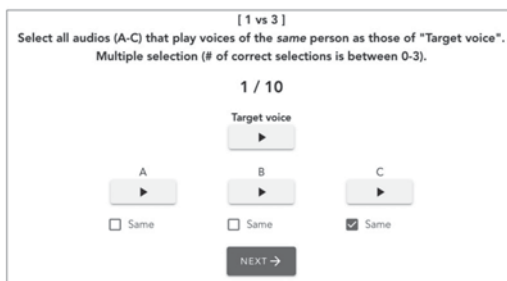


Fig. 1: Microtask UI example used for experiment (1 vs 3 voice comparison)

1-3P-9

1-3P-9 雑音環境下音声に対して頑健な Zero-Shot 声質変換

Noise Robust Voice Conversion for Target Speakers

☆坂本 瞭(静大・院・総合科技研),

立蔵 洋介(静大・院・総合科技研)

- ◆背景: 新しく音声データの獲得が難しい話者の音声データに雑音が含まれている場合、変換対象として利用できない
- ◆目的: 雑音環境下音声を用いた声質変換の性能向上
- ◆提案法: 学習時にクリーン音声のみでなく雑音混合音声を用いてモデルの学習。学習の収束性を高めるために二段階に学習
- ◆結果: 既知話者の雑音の含まれている音声を用いた声質変換において、変換性能の向上を確認

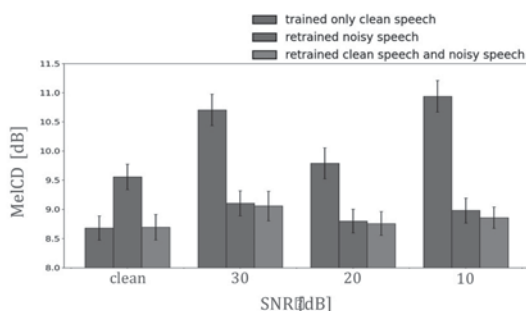


Fig 1: Evaluation result of the Mel-CD for the conversion from seen male to seen female.

1-3P-8

1-3P-8 強制アライメント版Parallel Tacotron 2 とHiFi-GANを用いたCPU型リアルタイム日本語ニューラルテキスト音声合成システムの実装

Implementing real-time neural text-to-speech system for Japanese on CPUs using Parallel Tacotron 2 with forced alignment and HiFi-GAN

○岡本拓磨¹, 戸田智基^{2,1}, 河井恒¹

(¹情報通信研究機構, ²名古屋大学)

- ・高速・高品質ニューラルボコーダ: HiFi-GAN
- ・CPUでの高速生成が可能
- ・転置畳み込み型アップサンプリングを採用
→本検討では補完型およびサブピクセル型アップサンプリングと比較
- ・複数話者・高速・高品質音響モデル: Parallel Tacotron 2
- ・予備検討より単一話者でのSoft-DTWでの実現は課題
→本検討では強制アライメントを用いた単一話者モデルを検討
- ・実験
 - ・HiFi-GANの3種類のアップサンプリング方式を比較
→従来法としてMulti-band MelGANとSub-DiffWaveを追加
 - ・単一話者Parallel Tacotron 2とTransformer型モデルの比較
- ・結果
 - ・Parallel Tacotron 2とHiFi-GAN V1で16CPUコアを用いたRTF約0.1の高速生成を実現(音質はTransformerと同等)

1-3P-10

1-3P-10 音声匿名化変換に用いる音響特徴構成の評価

Evaluation of Acoustic Feature Composition for Speech Anonymization Transformation

☆近藤伊佐直, 西村竜一(和歌山大)

- ◆WORLD[森勢+, 2016]特徴量の交換による音声の匿名加工システム[近藤+, 2020]において、各特徴量が話者特定に与える影響を100話者の発話データを用いて評価した。
- ◆10話者ずつ10グループに分け、10話者のデータでCNNによる10クラス分類器を十分学習させ、識別正解率を求める。この正解率を標準化したスペクトル包絡 (SP Norm) と標準化した非周期性指標 (AP Norm)、標準化を行っていないスペクトル包絡 (SP) と非周期性指標 (AP) とで比較した。
- ◆結果、スペクトル包絡に比べ非周期性指標が話者特定に寄与しないこと、標準化処理を行うと話者識別が容易になることが示された。

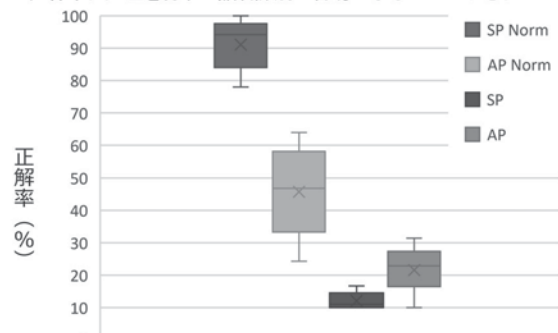


Fig.1: 100話者の音声を用いた自動話者識別の結果(集計)

1-3P-11

1-3P-11 歌声の基本周波数変化に伴う 音色変化に対応する動的変動量の解析

Analysis of dynamic variation in timbre
related to the change of frequency in singing

☆金井郁也, 森勢将雅(明治大)

- ◆目的: 人間歌唱音声と F0 変換音声における
基本周波数変化に伴うスペクトルの時間的変動の違いの調査
 - 大幅な音高補正で生じる個人性の破綻の改善
 - 対数パワースペクトルの時間方向の標準偏差の抽出
- ◆先行研究: 基本周波数変化に伴うスペクトル包絡の変化
 - スペクトルの形状に関する分析が主であった
 - 形状だけでなく、時間的変動にも差があるか?
- ◆解析結果: 人間歌唱音声において、発声周波数の上昇に伴う、
対数パワースペクトルの時間方向の標準偏差の減少傾向を確認

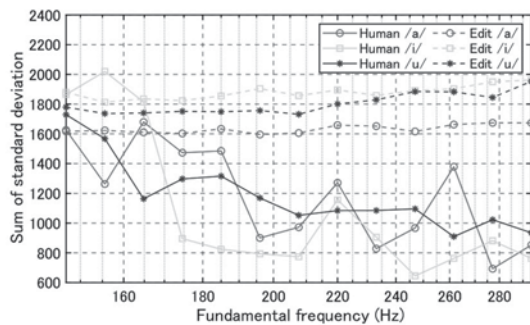


Fig. 1: Difference of temporal variation between human singing and F0-shifted singing

1-3P-13

1-3P-13 統計的音声合成のための 非負値自己符号化器を用いた 音響モデリングの検討

A study of acoustic modeling using non-negative autoencoder
for statistical speech synthesis

☆五来丈瑠, 齋藤大輔, 峯松信明(東大)

- ◆背景: 統計的音声合成において、高次元なスペクトル包絡を低次元な音響特徴量に圧縮することで効率的に推定できることが知られている
- ◆Non-negative Autoencoder (NAE)
 - Non-negative Matrix Factorization(NMF)を Autoencoder として解釈し直すことで NN に拡張したモデル
- ◆提案手法
 - NAE による音響特徴量の次元削減を用いた TTS
 - NAE の再構成と DNN 音響モデルをマルチタスク学習させることで言語情報を考慮した、より音声合成に適した次元削減を行う
 - 音響モデルは話者非依存に学習し、NAE は話者ごとに分ける複数話者モデルについても検討する
- ◆主観評価実験
 - NAE と音響モデルを同時学習することの有効性が示された
 - Deep Autoencoder を用いた場合と比較して高い自然性を示した
 - 複数話者モデルは、単一話者のみで学習したモデルと比較して話者性を保ったまま自然性を向上させることに成功した

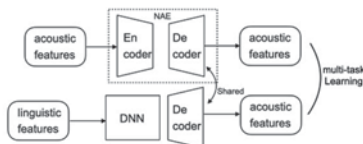


Fig. 1: Overview of proposed method

1-3P-12

1-3P-12 正則化 TV-CAR 分析を用いた音声の F0 推定に関する一検討

A study on F0 estimation based on regularized TV-CAR speech analysis

舟木慶一(琉球大)

解析信号に対する時変複素 AR (TV-CAR) 音声分析として、時間軸でも周波数軸でもスペクトルの急激な変化を抑える l_2 ノルム正則化 TV-CAR 分析の提案を行った。これは、隣接フレーム間、すなわち、時間軸でスペクトルが急激に変化しない罰則項と、周波数軸上でスペクトルが急激に変化しない罰則項の和を正則化項とする l_2 ノルム正則化アルゴリズムである。音声分析法の性能を評価するため、分析して得られる複素残差信号を用いて IRAPT による F0 推定を行い、性能比較によりその有効性を明らかにしている。本稿では、前処理として、骨導音の特性を有する 1 次の AR フィルタを導入し、どの程度性能が向上するか
の検討を行った。検討の結果、女性音声のノイズレベルが高い場合に性能向上が見られた。今後、骨導音を模擬するより精巧なフィルタを検討し導入する予定である。

1-3Q-1

1-3Q-1 単母音発声を対象とした 深層話者照合手法の検討

Speaker Verification based on Deep Similarity Learning
for Single-Vowel Sounds

☆高見澤真央, 黒岩真吾, 堀内靖雄(千葉大), 柘植寛(大同大)

- ◆2 つの音声入力に対して、2 つの音声は同一話者によるものであるかを判定する同一話者判定 DNN を提案する。(Fig.1)
- ◆科学警察研究所のデータベースに収録されている単母音音声 (0.12 秒) を用いて話者照合実験を行い、同一話者判定 DNN を評価するとともに、マルチタスク学習の導入を検討した。
- ◆ユークリッド距離に比べ、同一話者判定 DNN の EER は低くなった。マルチタスク学習による EER の変化は見られなかった。(Fig.2)
- ◆母音毎の DNN 出力の平均を用いて、照合実験を行ったところ EER が減少した。5 母音 (0.12 秒×5) の結果を統合することで、提案手法が話者照合に応用可能であることが示された。(Table1)

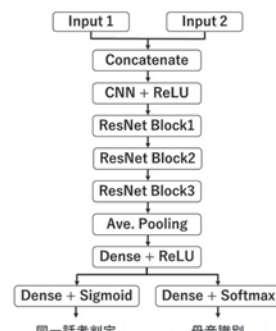


Fig.1 同一話者判定DNN

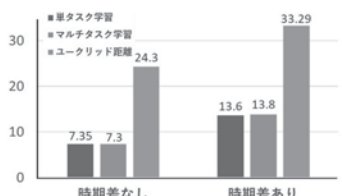


Fig.2 単母音の話者照合結果 (EER[%])

Table1 5母音を用いた話者照合結果 (EER[%])

Time Delay	Single-task Learning		Multi-task Learning	
	Time delay none	Time delay present	Time delay none	Time delay present
同一話者判定	1.59	6.90	1.72	5.41

1-3Q-2

1-3Q-2 音声強調のための深層学習に基づくクリッピングノイズ抑圧の検討

A study on clipping noise suppression based on deep learning for speech enhancement

○福森 隆寛(立命館大)

◆ 研究背景

音声を収録するときに音声の音圧に対して録音レベルが必要以上に高い場合、収録音声がクリッピングすることがある。音声の再収録が困難な状況においては、クリッピングによって生じたノイズを収録音声から除去する必要がある。

◆ 研究概要

3種類の深層ニューラルネットワーク(Bidirectional LSTM, Denoising Convolutional Autoencoder, U-Net)を用いて収録音声に含まれるクリッピングノイズを抑圧し、これらのネットワークが音声強調に与える効果を包括的に評価した。

◆ 評価実験

クリッピングさせた LibriSpeech ASR corpus の音声データセットを用いてネットワークを学習・評価した結果、Fig.1のU-Netでクリッピングノイズを抑圧させることで、PESQが0.8ポイント改善した。

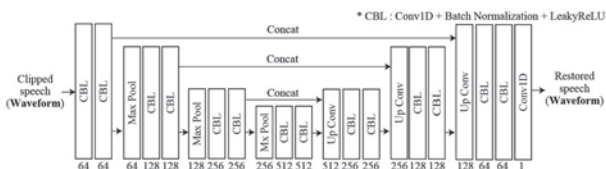


Fig.1: Overview of a U-Net structure for clipping noise suppression.

1-3Q-3

講演取消

1-3Q-4

1-3Q-4 Kaldi ベースのストリーミング音声認識システムの開発

Development of a Streaming ASR System based on Kaldi

☆レオ テーシャン, 王宇, (山梨大院・工), 小林彰夫(筑波技術大・産業技術), 宇津呂武仁(筑波大・システム情報), 西崎博光(山梨大院・工)

◆ 研究内容:

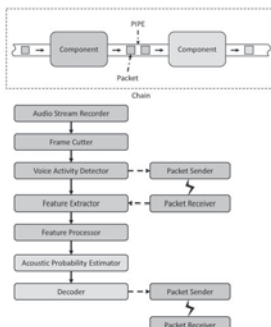
Kaldi ベースのオンラインストリーミング音声認識システムを構築・評価

◆ システム構成

```

1 # Read stream from microphone and cut frames.
2 reader=StreamReader(WebrtcvadDetector())
3 cutter=ElementFrameCutter()
4
5 # Online MFCC extraction and processing.
6 extractor=MfccExtractor()
7 processor=FeatureProcessor(FrameSlideOvnNormalizer())
8
9 # Apply original DNN acoustic model.
10 estimator=AcousticEstimator()
11 estimator.acoustic_function>Your_Function
12
13 # Define an online decoder.
14 decoder=HstDecoder(symbolTable="words.txt",
15                   silencePhones="12:3:4:5",
16                   frameShiftSec=0.01,
17                   tmodel="final.ndi",
18                   graph="HCLG.fst")
19
20 # Link all components.
21 chain = Chain()
22 chain.add(reader)
23 chain.add(cutter)
24 chain.add(extractor)
25 chain.add(processor)
26 chain.add(estimator)
27 chain.add(decoder)
28
29 # Run and display the results dynamically.
30 dynamic_run(chain)

```



- ▶ 独立性の高いモジュールをパイプラインで結合することで音声認識システムを実現 Python で短いコードで実装可能(上図参照)
- ▶ 雑音除去やVADなどの処理を容易に実装でき、評価可能
- ▶ オリジナルのKaldiと同程度の性能を引き出すことが可能

◆ 評価

Table1: Online と Offline 処理時の WER [%]

	Offline	Online
Baseline	18.58	—
DNN	19.81	19.97

Table2 音声分離適用時の WER[%]と RTF

	WER	RTF
Clean	22.66	0.60
Clean + noise w/o separation	47.58	0.67
Clean + noise w/ separation	38.09	0.86

1-3Q-5

1-3Q-5 音響タグ付けのためのイベントクラスを考慮した音響合成に基づくデータ拡張

Audio Synthesis-based Data Augmentation Considering Event Class for Audio Tagging

☆杉浦東風(山梨大・院), 小林彰夫(筑波技術大・産業技術), 宇津呂武仁(筑波大・システム情報), 西崎博光(山梨大・院)

◆ 研究目的:

野球中継放送の音響データに対する音響タグ付けタスクにおいて,

「イベントクラスに類似した音データの合成に基づくデータ拡張方法」を提案

◆ 使用する音響データ:

- ▶ 野球中継放送音響タグ付けデータセット
- ▶ データ拡張用データセット
- ◇ 使用するタスクに類似しているが、タスクで設定されているラベルとは異なる種類の音を使用

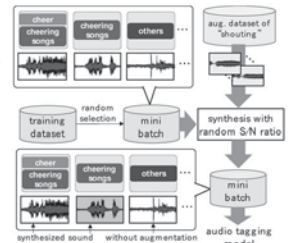


Fig. 1 Proposed method

◆ 提案するデータ拡張方法:

データ拡張を行うクラスを持つデータに対し、データ拡張用データセットから適切なクラスのデータをランダムに合成

◆ 実験結果

ベースライン(データ拡張無し)や従来法の mixup を適用した場合と比べて、音響タグ付けタスクでの F1 スコア向上を確認

Table 1 Audio tagging performance

データ拡張方法	F1 スコア (マクロ平均) [%]
ベースライン	55.7
mixup	56.8
提案手法	69.7

1-3Q-6

1-3Q-6 妨害発話に基づく補助損失を用いたマルチモーダル目的話者音声認識

Interference speaker loss for audio-visual target speaker speech recognition

☆角田遼太(神戸大), 相原龍(三菱電機)

高島遼一, 滝口哲也(神戸大), △今井良枝(三菱電機)

- ◆マルチモーダル音声認識モデルとして、音響特徴量をクエリとして画像特徴量に Attention をかける Cross-modal attention 機構により音声と画像の特徴量を統合する AV Align が提案されている。
- ◆AV Align は背景雑音環境下において認識精度の改善を示しているが、妨害発話環境下では認識精度が大きく劣化するという課題がある。
- ◆本研究では、モデルの学習時に妨害発話に基づく補助損失関数と口唇動画画像から発話内容の認識を行うための補助損失関数を併用する手法を提案する。
- ◆実験結果から、妨害話者の発話認識を行うための損失関数を用いることは、目的話者の発話認識精度を向上させるために有効であることが分かった。

Table 1 Character Error Rates (CER) of each experimental condition

Model	CER[%]				
	10 dB	5 dB	0 dB	-5 dB	-10 dB
Audio only	29.7	38.9	49.6	57.2	60.5
Baseline	18.4	21.1	23.9	27.5	30.6
+visual loss	21.6	24.3	27.2	30.5	33.3
+interference loss	15.7	17.8	20.3	23.1	26.4
+visual loss	17.4	19.4	22.0	24.5	27.9

1-3Q-8

1-3Q-8 ドメイン適応と話者一致損失を用いた話者適応によるクロスリンガル音声合成

Cross-lingual Speaker Adaptation using Domain Adaptation and Speaker Consistency Loss for Text-To-Speech Synthesis

☆辛 徳泰, 齋藤 佑樹, 高道 慎之介,

郡山 知樹, 猿渡 洋(東大院・情報理工)

- ◆ドメイン適応と話者一致損失を用いた話者適応によるクロスリンガル音声合成方法を提案する。
- ◆提案法はまずドメイン適応を使い、言語非依存のマルチリンガル話者照合モデルを学習する (Fig. 1)。
- ◆更に話者照合モデルが生成した話者埋め込みを用い、モノリンガル音声合成モデルを学習する (Fig. 2 Normal Training)。
- ◆提案法は最後に合成された音声から生成した話者埋め込みと自然音声から生成した話者埋め込みのコサイン類似度を最小化するという話者一致損失を使い、他言語話者の音声でモデルを fine-tuning する (Fig. 2 Speaker Consistency Training)。
- ◆実験結果は提案法がより少ないデータ量で高い自然度を持つクロスリンガル音声合成できることを示す。

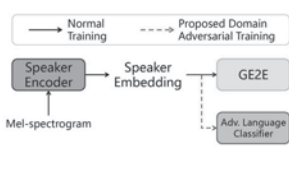


Fig. 1: Architecture of the proposed speaker encoder.

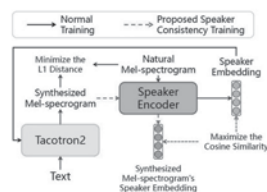


Fig. 2: Architecture of the proposed TTS model.

1-3Q-7

1-3Q-7 音声意味理解への応用を指向した非自己回帰型 End-to-end 音声認識

Non-autoregressive speech recognition for spoken language understanding

◎大町基, 藤田悠哉(ヤフー株式会社), 渡部晋治(CMU), Tianzi Wang (JHU)

背景

- ◆筆者らは処理時間が短く高精度な音声意味理解システムの実現を目指し、Transformer を用いて発話内容に含まれる個々の単語の表記と素性を並べた単一の系列を出力する方式を提案した
- ◆しかし、Transformer は音声は left-to-right に生成されるという仮定に基づくため、推定したい系列の長さに比例して実行時間が長くなるという課題があった

提案法と特長

- ◆非自己回帰型モデルのひとつである、Mask-CTC の枠組みを用いて単語の表記と素性を同時に推定する方式を提案
 - Mask-CTC は left-to-right の仮定が不要なため、Transformer よりも高速な推論が可能
 - mask-predict を CTC (Connectionist temporal classification) に基づく非自己回帰モデルに適用することで、CTC よりも高い認識性能を実現
- ◆Mask-predict を複数種類のトークンが含まれる系列に適用するために、各反復で更新するトークンを制限する Type-wise mask-predict を提案 (Fig.1 に例を示す)

結果

- ◆CSJ を用いた音声認識実験： CTC よりも高い認識性能が実現でき、Transformer よりも低い実時間比 (RTF) を達成
- ◆音声意味理解タスク： 品詞推定、言い淀み・固有名詞抽出のタスクで CTC よりも良い性能が得られることを確認

```
CTC : 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 静帯 <PHON:sa> <PHON:to> <PHON:i> <POS:ID> ...
Itr-0: 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 声帯 <PHON:sa> <PHON:to> * * <POS:ID> ...
Itr-1: 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 声帯 <PHON:sa> <PHON:to> * * <POS:ID> ...
Itr-2: 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 声帯 <PHON:sa> <PHON:to> * * <POS:ID> ...
Itr-3: 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 声帯 <PHON:sa> <PHON:to> <PHON:i> <POS:ID> ...
REF : 提示 <PHON:ta> <PHON:to> <PHON:ji> <POS:ID> が <PHON:ga> <POS:PP> 声帯 <PHON:sa> <PHON:to> <PHON:i> <POS:ID> ...
```

Fig.1: Example of the mask-predict. Itr-n and REF denote the sequence refined by mask-predict of the n-th iteration and reference sequence, respectively. "*" denote the mask token

1-3Q-9

1-3Q-9 音響モデルの話者適応に基づく脊髄性筋萎縮症者の音声明瞭化の検討

Speech Clarification in Persons with Spinal Muscular Atrophy Based on Speaker Adaptation of Acoustic Model

☆吉本拓真, 高島遼一(神戸大),

△佐々木千穂(熊本保健科学大), 滝口哲也(神戸大)

- ◆I型の脊髄性筋萎縮症(SMA)者の音声は発話が不明瞭となるため、コミュニケーションが困難である。そこで本研究では、本人らしい声(話者性)を維持しつつ、かつ明瞭性の高い音声を合成するシステムの実現を目的とする。
- ◆先行研究では、健康者データで学習した音響モデルをSMA者へ話者適応する手法を検討したが、この手法では適応の度合いによって話者性と明瞭性がトレードオフの関係になっていた。
- ◆そこで提案手法では、話者適応時に明瞭性が失われないようにするため、音響モデルの出力を健康者音声認識モデルに入力し、正しく音声認識できることを担保する補助損失関数を導入する。(Fig. 1)
- ◆音声認識モデルによる損失関数を用いない従来法と比べて、提案手法は話者性と明瞭性を両立した音声合成できる可能性が示唆された。

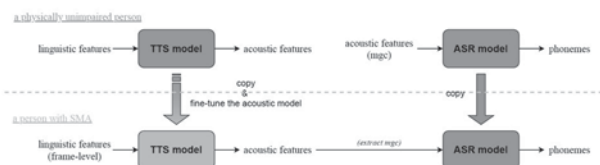


Fig. 1 Training procedure of the proposed TTS system.

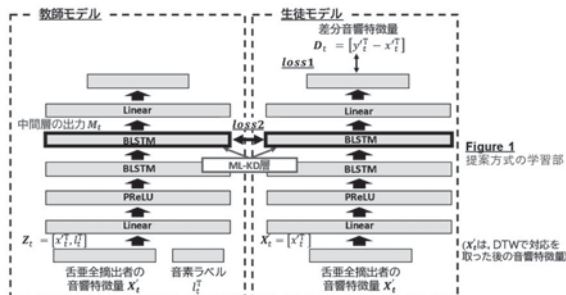
1-3Q-10

1-3Q-10 音素情報を知識蒸留する舌歪全摘出者の音韻明瞭度改善法

Method for Improving the Speech Intelligibility of Glossectomy Patients for Knowledge Distillation of Phonemic Information

☆高島和嗣, 原直, 阿部匡伸(岡山大・HS 統合科学研)

- ◆音素ラベルを用いた声質変換により, 舌歪全摘出者の音韻明瞭度の改善が報告されている。本報告では, 知識蒸留を用いて, 変換時に音素ラベルを不要とする方式を提案する。
- ◆Figure 1に示すように補助情報として音素ラベルを用いる方式の変換モデルを教師モデル, 合成時に音素ラベルを必要としない方式の変換モデルを生徒モデルとし, 知識蒸留をおこなう。
 - 教師モデルと生徒モデルの中間層の出力の誤差を学習に用いる。
- ◆評価実験により, 知識蒸留をおこなうことによって, 音素ラベルを用いず音声情報のみの入力での音韻明瞭度の改善を示した。
 - 生徒モデルの出力の誤差 loss1 と中間層の出力の誤差 loss2 を合計した値を用いて学習する方法が最も良い結果を示した。



1-3Q-12

1-3Q-12 Wav2Vec2.0 による音声認識手法を用いた音声合成モデル学習用入カラベル推定手法の検討

Using Speech Recognition by Wav2Vec 2.0 to Estimate Input Labels for Learning a Speech Synthesis Model

©栗原清(NHK メディア開発企画センター)

- ◆音声認識手法である **Wav2Vec 2.0** を用い, 音声から音声合成モデル学習用入カラベルである「読み仮名と韻律記号」を推定する手法
- ◆Wav2Vec 2.0 を用いた音声認識手法を応用する事で, **5時間程度**の少量の正解ラベル(読み仮名と韻律記号)を用いて音声のみから学習用ラベルを推定できる
- ◆読み仮名(片仮名)だけでなく, アクセント上昇・下降, アクセント句区切り, 文末等の**韻律記号**も推定可能な事を確認できた
- ◆提案法の評価結果を Table1 に示す。実験では, 従来法(Conventional)として, 既存の公開済み学習モデルと OpenJTalk の組み合わせを使用した。提案法(Proposal ①)として, Wav2Vec 2.0 のみを用いた手法, 提案法(Proposal ②)として, Wav2Vec 2.0+音素誤り訂正用 Transformer を用いた手法について, CER を算出し提案法の有効性を確認した

Table 1 Effectiveness of Wav2Vec 2.0

	Method	CER %
Conventional	EspNet ASR + OpenJTalk	22.6
Proposal ①	Wav2Vec 2.0	8.5
Proposal ②	Wav2Vec 2.0 + Transformer	4.7

1-3Q-11

1-3Q-11 喉頭の高速度デジタル撮像を用いたロック歌唱におけるフォールスコードスクリームの検討

A study of false-cord scream in rock singing using high-speed digital imaging of the glottis

☆田中誠人, 李庸學, 錦木時彦(九州大)

- ◆ フォールスコードスクリームは歪み発声の1つであり, ロックやメタルの歌唱で用いられる声帯-仮声帯発声である。本研究では, フォールスコードスクリームの特徴と発声メカニズムを, 音声データと喉頭動態の高速度デジタル撮像(HSDI)から明らかにすることを目的とした。
- ◆ 音声と HSDI の比較は, 共に長時間平均スペクトラム(Long Time Average Spectrum, LTAS)を求めることで行った。HSDI から LTAS を求めるため, 仮声帯を含む画像中の特定の線分上でキモグラフと平均画素値の時間変化を求めた。
- ◆ 分析により, フォールスコードスクリームは, 仮声帯が声帯に対して整数倍の周期で振動する発声であることがわかった。音声データからは基本周波数 F0 の他に, およそ F0/n (n: 2,3,4,5) のピッチ(サブハーモニクス周波数)を聴取できた。サブハーモニクス周波数の整数倍の周波数成分は, 音声と HSDI 双方の LTAS から明確に確認できたため, サブハーモニクス周波数での仮声帯振動が低いピッチを生成する直接的な原因であることを示すことができた。

1-3Q-13

1-3Q-13 音声分析合成システムの無声音生成に用いる励起信号の違いが分析合成音の品質劣化に与える影響の調査

Influence of difference of residual signal used for generating unvoiced sound on sound quality in synthesized speech

☆田鎖佑弥, 森勢将雅(明治大)

- ◆ 目標: 無声音生成に用いる励起信号の違いが与える影響の調査
 - 異なる励起信号で再合成音声を生成
 - DMOS 評価を行い音声の劣化度合いを調査
- ◆従来法: Gaussian white noise (GWN)
 - 既存の高品質ボコーダの励起信号として用いられている
 - 短時間パワーにばらつきがある
- ◆提案法: Modified velvet noise (MVN)
 - ベルベットノイズを音声合成向けに改良
 - 短時間パワーの安定性に優れる
- ◆解析結果
 - 有意差は認められなかった
 - 個別の音声について定性的な解析を行った
 - MVN は摩擦音生成に強い可能性が示された

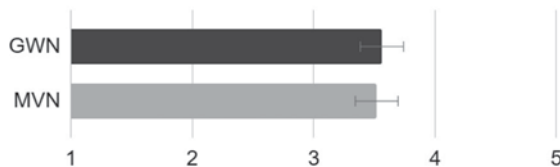


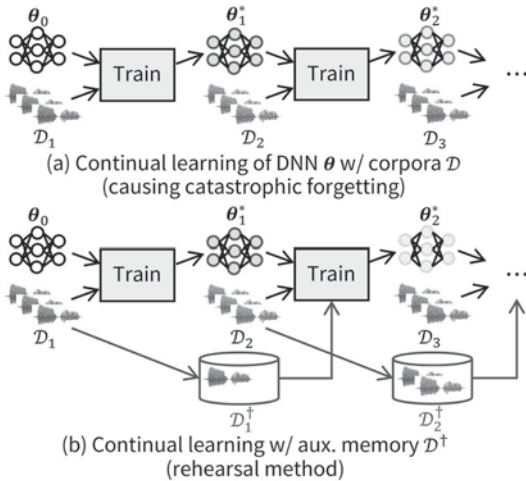
Fig.1: Results of DMOS evaluation

1-3Q-14

1-3Q-14 End-to-End 音声合成の Continual Learning における 破滅的忘却の影響の調査

Investigation of effects caused by catastrophic forgetting in continual learning of end-to-end text-to-speech synthesis

○齋藤 佑樹, 猿渡 洋(東大院・情報理工)



過去に学んだ知識を忘れずに、
継続的に学習可能な

End-to-End 音声合成を実現するための初期検討です

1-4P-2

1-4P-2 モーフィング音声を用いた 話者の類似性判断と同定判断の比較

A comparison of speaker similarity judgement and speaker identification judgement using morphed speech sounds

☆内堀颯太, 松井淑恵(豊橋技科大院)

- ◆原音声間の類似度が高い組み合わせと低い組み合わせのモーフィング音声を用いて、原音声の類似度が話者類似性判断(2択)と話者同定判断(3択)にどのように影響するかを調べた。
- ◆類似性判断では、モーフィング割合の異なる音声2つのどちらが、ターゲット音声に似ているかを判断させた。音声の片方はターゲット音声の割合を50%とした。同定判断でも同様に、ターゲット音声と同一話者なのはどちらかを判断させた。
- ◆話者類似性判断より話者同定判断のほうが丁度可差異(JND)が小さかった。また、原音声同士が似ていない組み合わせのモーフィング音声を用いると、似ている組み合わせに比べ、JNDが小さかった。

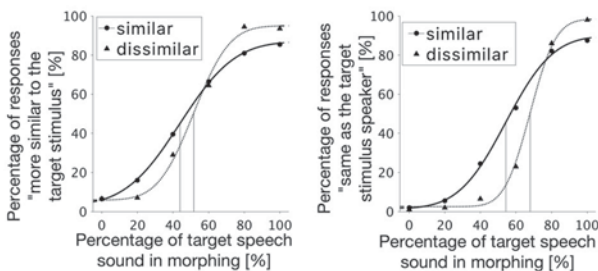


Fig. 1: The results of two speaker recognition judgements using morphed speech sounds. Each marker means whether the two original sounds combination used for morphing are the most similar or the most dissimilar.

1-4P-1

1-4P-1 飽和を含んだ1自由度モデルによる耳音響放射の抑圧シミュレーション

Suppression Simulation of Otoacoustic Emissions Using a One-Degree-of-Freedom Model with Saturation

☆藤匠海(大島商船高専), 村上泰樹(九大芸工), △浅川貴史(大島商船高専)

- ◆OAEの抑圧現象から得られる蝸牛の周波数選択性は、実際の蝸牛の周波数選択性よりも低下するとされる。この原因として、抑圧現象が蝸牛内部の基底膜の広い領域で作用するためと考えられている。
- ◆本シミュレーションでは、抑圧現象の基底膜上の広がり影響を検証するため、空間的広がりを排除した1自由度振動系のモデルを用いて、実際の周波数選択性と抑圧による推定値を比較する。
- ◆モデルの入出力特性から求めた実際の周波数選択性(実線)と、OAEの抑圧現象から推定した周波数選択性(破線)を図1に示す。
- ◆シミュレーションの結果、実際の周波数選択性よりも、抑圧を用いた推定において周波数選択性の低下が観測された。
- ◆このことから、基底膜の局所的な範囲で、OAEの抑圧現象から得られる周波数選択性が低下することが示唆される。

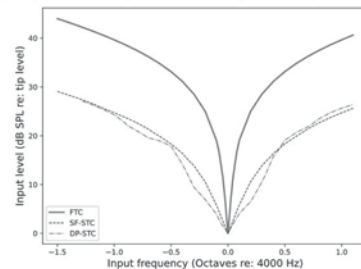


Fig.1: Frequency Tuning Curve and Suppression Tuning Curve

1-4P-3

1-4P-3 音像の水平移動が ノイズネスに及ぼす影響に関する一考察

A consideration on the effects of horizontal movement of sound image on the noisiness

○石井要次, 星加 慧(小林理研)

- ◆音像の水平移動とノイズネスの関係について考察するため、スピーカを用いた聴感実験とヘッドホンを用いた聴感実験を実施した。
- ◆スピーカによる合成音像を水平移動させた場合、移動速度が速くなるにつれてノイズネスが大きくなること、移動速度によらず静止状態よりもノイズネスが大きくなる可能性が示された(Fig.1:実線)。
- ◆ヘッドホン実験により両耳間時間差(ITD)や両耳間レベル差(ILD)のみを変化させた聴感実験を行った結果、ITDよりもILDの方がノイズネスに及ぼす影響が大きい可能性が示された(Fig.1:破線と点線)。
- ◆また、ヘッドホン実験でノイズネスを検証する場合、頭部伝達関数などにより音像方向を制御する必要性が示唆された。

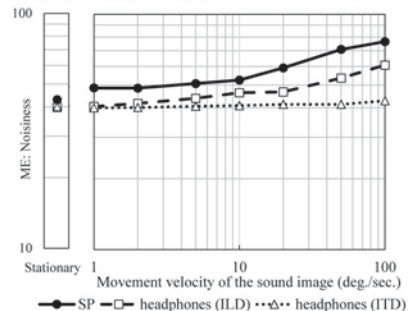


Fig.1: The result of the experiments by the loudspeakers and the headphones.

1-4P-4

聴覚的注意の全方位における空間特性の検討

Spatial characteristics of auditory attention for full azimuthal directions.

☆富松智志, 坂本修一(東北大), 川瀬哲明(東北大)

- ◆本研究では、正面以外への方向へ向けられた聴覚的注意が、他方向から到来する音の聴取に与える効果について検討した。
- ◆周囲 30°おきに配置されたスピーカを中心に聴取者をおき、いずれかのスピーカの方向に注意を向けた状態で、提示された競合音中の標的音・偽標的音を検出させた。
- ◆実験の結果、注意を向けた方向の検出性能は概ね高いものであった。また、注意を向けた方向と耳軸上で対称な方向での検出性能も高くなった。
- ◆前後方向 (Fig. 1 (a), (c)) と横方向 (Fig. 1 (b)) では、グラフの概形が大きく異なることから、注意の空間的な特性が異なることを示唆する結果が得られた。

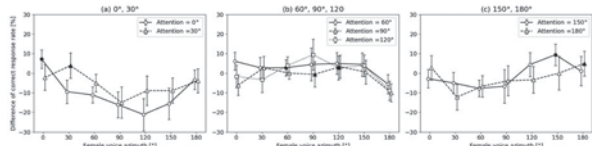


Fig. 1 Difference of correct response rate (CRR) between attention condition and random condition. (a), (b), and (c) show the CRR differences for 0° and 30°, 60°, 90°, and 120°; and 150° and 180°, respectively. Filled markers denote the direction of attention. Error bars denote the standard error of the mean.

1-4P-6

1-4P-6 妨害課題に対する注意資源の配分がオリブ蝸牛束反射に与える影響

Effect of allocation of attentional resources to distractor tasks on medial olivocochlear reflex

☆鈴木麻里子, 石坂勇毅, 大塚翔, 中川誠司(千葉大)

- ◆脳幹から外有毛細胞へは、オリブ蝸牛束と呼ばれる遠心性神経の当社がある。オリブ蝸牛束は音響刺激に反応し、外有毛細胞の増幅を抑制する働きがある。この抑制反応は、オリブ蝸牛束反射 (MOCR) と呼ばれ、雑音下での聞き取りを改善する役割 (アンチマスキング効果) を果たすと考えられている。
- ◆MOCR が視覚的・聴覚的注意の影響を受けることが報告されているものの、注意状態に応じた MOCR の変動が聴取に与える影響を調べた研究はなかった。
- ◆MOCR のアンチマスキング効果を考えると、注意状態の変化に伴う MOCR の変化は雑音下での聴取に影響を与えると推察される。
- ◆本研究では、妨害課題を加えることで聴取課題への注意資源の配分を調整し、それに伴った雑音下での聴取成績の変化と MOCR の変動の関係を調べた。
- ◆被験者 4 名を対象とした予備的な検討を行った結果、妨害課題の負荷によって、雑音下の聴取課題の正答率が低下する傾向が見られた。一方で MOCR 強度には、妨害課題の負荷に応じた変化は見られず、本検討からは仮説を支持する結果は得られなかった。

1-4P-5

1-4P-5 振動子呈示部位が軟骨伝導知覚に及ぼす影響 -外耳道注水条件下における検討-

Influence of transducer placements on cartilage conduction perception -An assessment under the water-injected condition into the auditory canal-

☆菅原優聖, 大塚 翔, 中川誠司(千葉大)

- ◆骨伝導の装用にかかる問題点を解決する方法として“軟骨伝導”が提案され、外耳道閉鎖症用の補聴器やスマートホンに应用されている。
- ◆軟骨伝導では振動子の呈示部位に依存して知覚特性が変化するが、伝搬経路の変化には不明な点が残る。前報では、伝搬経路の解明を目的として、耳栓で気導成分を、外耳道への注水で外耳道内放射成分を遮断した際の聴覚閾を計測した。しかし、外耳道への注水により本来の伝搬経路とは異なる経路が生じた可能性も否定できない。
- ◆本研究では外耳道への注水の影響をより詳細に検討するため、注水量を段階的に変化させて聴覚閾計測を行った。
- ◆500 Hz までの低周波域では、注水量が 60% を超えると聴覚閾の低下が観察された (Fig. 1)。水位が外耳道軟骨部に到達することで、外耳道放射成分が水を介して鼓膜に到達したためであると考えられる。
- ◆1000 ~ 4000 Hz では、注水量にかかわらず乳様突起に比べ耳介、耳前部で有意に大きな聴覚閾上昇が観察された (Fig. 2, 3)。注水により気導・外耳道内放射成分が遮断されたためであると考えられる。

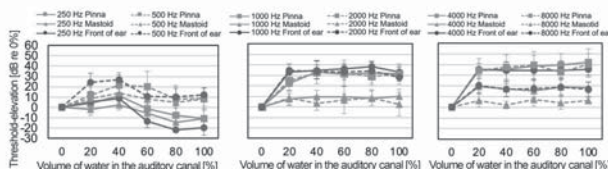


Fig. 1 Threshold-shifts from the 0%-water injection at 250 and 500 Hz. Fig. 2 Threshold-shifts from the 0%-water injection at 1000 and 2000 Hz. Fig. 3 Threshold-shifts from the 0%-water injection at 4000 and 8000 Hz.

1-4P-7

1-4P-7 顔面頭蓋に呈示された骨伝導音の検出閾計測

Threshold measurements of bone-conducted stimulation presented to the human facial cranium

☆上村 昂, 大塚 翔, 中川誠司(千葉大)

- ◆一般に骨伝導振動子は側頭骨の乳様突起や下顎骨の顎状突起等に呈示されるが、鼻根部等の顔面に呈示する方法も検討されている。
- ◆顔面には副鼻腔等の空洞構造が存在する。また、一部の顔面部位は聴覚器に近接し、従来部位とは異なる知覚特性を示す可能性がある。
- ◆顔面頭蓋における骨伝導知覚特性の解明を目的とし、鼻骨、眼窩下部 (上顎骨)、頬骨、下顎角 (下顎骨)、および顎 (下顎骨) に骨伝導呈示した際の検出閾計測を行った。
- ◆顎を除く顔面各部位の 500-2000 Hz における検出閾は、従来部位である乳様突起、前額部と遜色ない値を示した。
- ◆顔面部位間で比較すると、500-8000 Hz においては顎を除く各部位間の検出閾の差異は小さかったが、125-250 Hz では一部の部位間の差異が増大した。外耳道内放射成分の多寡が影響している可能性がある。

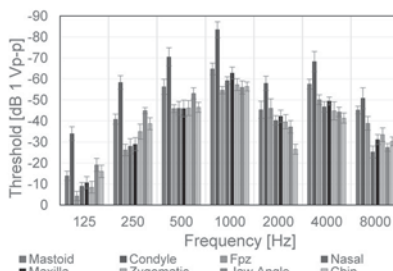


Fig. 1 Hearing threshold for each placements

1-4P-8

1-4P-8 Deep Decoder を用いた聴覚誘発反応の電流源推定

Current source estimation of auditory evoked response using deep decoder

☆山名莉央, 矢野肇, 高島遼一, 滝口哲也(神戸大), 中川誠司(千葉大)

- ◆脳機能計測手法である脳磁図, 脳波の計測結果から脳内の電流分布の推定は劣決定問題であり, 解くためには電流源の分布を明示的に与える必要があるがその分布は不明である.
- ◆本研究では明示的に事前分布を与えないようにするため, Deep Decoder を用いて電流源推定を行った.
- ◆畳み込みからなるネットワークが事前分布の役割を果たすが, パラメータが多くノイズにオーバーフィッティングしやすい.
- ◆そこで畳み込みをしないネットワーク(Deep Decoder)を用いることでパラメータ数を減らし電流源を推定した.

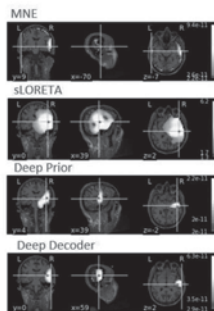


Fig.1: Current source densities estimated from simulated MEG data in the primary auditory cortex

1-4P-9

1-4P-9 音源近傍の反射壁による反射波が音像定位に与える影響

Effect of reflected waves from a reflective wall near a sound source for sound localization

○森川大輔, △河内悠希(富山県立大)

- ◆反射壁により音像が音源方向より後方ずれて知覚されることがある.
- ◆反射壁と対側の耳に届く反射波が成因であるかを検証するため, 反射波の有無を変更した合成バイノーラル信号を用い, 聴取実験を行った.
- ◆反射を含まない S^0 と左チャンネルのみ反射を含む S_{left}^r に対して, 両チャンネルに反射を含む S_{bi}^r と右チャンネルのみ反射を含む S_{right}^r は後方に定位された.
- ◆反射壁と対側の耳に届く反射波を含む場合に音像が後方に定位されることがわかった.

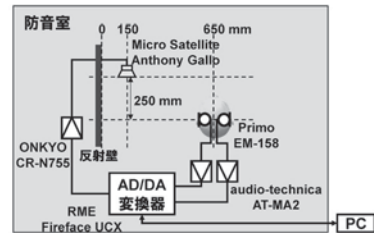


Fig.1: Impulse response measurement system with an arrangement whereby the direction of the sound image shifts due to the reflection wall.



Fig.2: Relationship between the direction of each stimulus.