

2-A-1

2-A-1 【招待講演】聴覚神経系の理解と人工知能

Understanding of the auditory neural system and Artificial Intelligence

○寺島裕貴(NTT CS 研)

- ◆一見縁遠いと思われる人工知能と聴覚の研究が、ある意味で深く結びついていることを示す。
 - ◆特に深層学習の登場後、その結びつきは「最適化を通じたブラックボックスの理解」という観点からさらに強まっている。
 - ◆深層学習を含む機械学習技術は分野間の壁を乗り越える「共通言語」としても有用であり、その普及は異分野交流と理解を加速する。
-
- ◆聴覚神経科学における機械学習応用は大きく二分できる。
 - 神経データ解析
 - 神経系の機能を理解するための計算モデル
 本講演では主に後者を紹介する。
 - ◆聴覚研究だけではなく視覚研究とのかかわりにも触れながら、深層学習「前」、そして深層学習「後」と区分して流れを紹介する。

2-A-3

2-A-3 人工知能研究の動向と今後の展望

Trend and future prospects of artificial intelligence research

○古井貞照(国立情報学研究所/TTIC)

AI 研究が進歩を続けており、近年の進展に大きな貢献をした技術に、GAN (敵対的生成ネットワーク、Generative Adversarial Networks)、Transformer、BERT がある。音声認識でも、単一あるいは複数のマイクロホンを用いて収録した音声入力に対して、重畳雑音、複数話者の音声、部屋の残響を分離する技術などで、大きな性能向上が達成されている。人の能力をすでに凌駕している領域もあるが、まだ現在の AI 技術ではできないことが沢山あり、これから技術が劇的に変化する可能性もある。

今後の進歩においては、社会現象を含む、幅広い基礎科学と AI の融合が、鍵になる。それによって、AI そのものだけでなく、基礎科学も大きな進歩を遂げ、人々の生活を変える。音声を含む音響学が対象としている分野は、人の発声器官や音響空間における物理現象が背景にある。まだ解決していない問題に対処するためには、単に観測されたデータや波形を、既存の AI 技術で処理するだけでなく、物理現象と、その背景にある因果関係のモデルを含む AI 技術を構築していく必要がある。

AI の進歩が加速度的に進んでいった先で、人間の理解が追いつかなくなるかもしれない。兵器や詐欺などに、AI が容易に悪用されるようになる可能性もある。公正性、正しい行動規範(倫理)をどのように堅持するか、人間の尊厳をどのように守っていくかが、新たな時代の幕開けへの重要な課題である。その時代に向かって、分野横断的にデータサイエンスの教育を行う仕組みなど、短期的・長期的な教育改革も必要となってくるであろう。

2-A-2

2-A-2 【招待講演】機械学習による音響イベント検出および音響シーン分類の研究動向

Recent research trend of detection and classification of acoustic scenes and events based on machine learning

○原田登(NTT)

- ◆本報告では、音響イベント検出、音響シーン分類を扱った国際会議である DCASE (Detection and Classification of Acoustic Scenes and Events) の最新のチャレンジタスクを中心に、どのような課題設定がなされているかという観点から、音響イベント検出および音響シーン分類に関わる研究動向について紹介する。
- ◆従来からの継続課題に加え、異常音検知 (Anomaly detection) や音の説明文生成 (Audio Captioning) など、新たな切り口で意欲的な課題が設定されており、参加者も積極的に工夫を凝らして精度向上に取り組んだ。このような競争型の取り組みにより、一層の技術発展が期待される。(DCASE 2020 Workshop は 2020 年 11 月 2 日～3 日に完全 Virtual 会議として開催予定です。積極的なご参加をお願いします。)

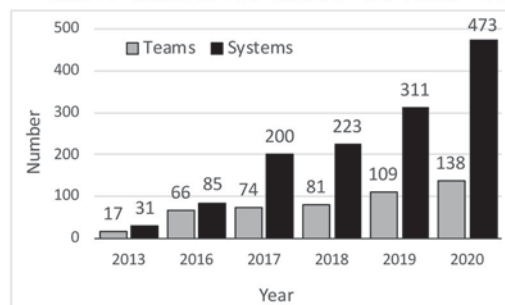


Fig.1: Number of contribution to DCASE Challenge

2-1-1

2-1-1 実環境におけるマイクロホンの移動に対する空間ケプストラムの頑健性の調査

Robustness of spatial cepstrum to microphone movement in a real environment

☆河村泰雅、宮崎亮一(徳山高専)、井本桂右(同志社大学)

- ◆ケプストラムやメル周波数ケプストラムは音響信号処理において広く使用されるが、これらの特徴量は空間情報を利用することは難しい。
- ◆マイクロホンの位置や正確な時間同期を必要とせず、各音源の空間情報を抽出する空間ケプストラムと呼ばれる特徴量が提案されている。
- ◆空間ケプストラムは録音環境の変化に対して頑健な特徴量であると主張されているが、最小限の条件でのみの実験しか行われていない。
- ◆マイクロホンの移動による空間ケプストラムの頑健性を調査するために、シミュレーションと実環境での評価実験を行う。

(i) シミュレーションでマイクロホンを移動させ、低次元の頑健性と高次元の録音環境の変化の影響を確認した (Fig.1)。

(ii) 実環境の音源を使用し、シミュレーションと同様の結果が得られ、実環境でも実用に耐え得ることが確認できた (Table 1)。

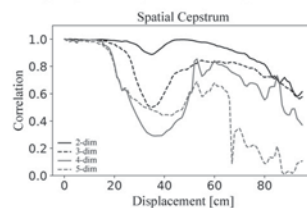


Table 1: Absolute value of the correlation coefficient between spatial cepstra obtained by true and mistaken microphones

Dimensions	2	3	4	5	6	7	8
Node 1	1.00	1.00	1.00	0.99	0.84	0.81	0.22
Node 2	1.00	1.00	1.00	1.00	0.99	0.68	0.65
Node 3	0.90	0.87	0.84	0.97	0.97	0.66	0.64
Node 4	1.00	1.00	0.98	0.99	0.68	0.59	0.67
Node 6	1.00	0.95	0.89	1.00	0.98	0.63	0.69
Node 7	1.00	0.98	0.95	0.89	0.95	0.74	0.79

Fig.1: Absolute value of the correlation coefficient between spatial cepstra obtained by initial and change positions.

2-1-2

2-1-2 弱ラベルで示される特定の共起関係に基づいたクロスモーダル音検索

Crossmodal Sound Retrieval based on Specific Target Co-occurrence Denoted with Weak Labels

©安田昌弘, 大石康智, 小泉悠馬, 原田登 (NTT)

【クロスモーダル音検索】

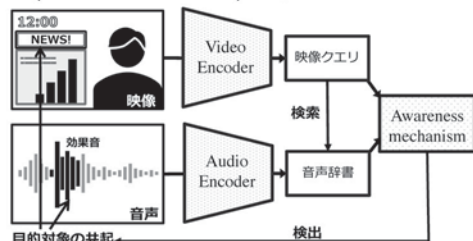
- 特定の音と映像が同時に現れやすい関係 (共起) を利用して、与えられた映像から音を検索する課題

【従来手法とその制限】

- 画像と説明音声 (言語的対応) や、物体を叩く映像と音 (物理的対応) といった単一の観点の共起に基づいた検索
 - ➡ 実世界データへの応用には制限がある

【提案法の特徴】

- 実世界データでの音検索タスクとしてテレビ放送における効果音推定問題を取り上げた
- 背景音楽や話声といった他の要素の中から、効果音に関連する特定の共起関係に注視するため、弱ラベルに基づく効果音検出器 (Awareness mechanism) を導入



2-1-4

2-1-4 物理的な対称性を保証したアンビソニクス領域 DNN による音響イベント検知・方向推定

Ambisonic domain DNN preserving physical symmetry and its application to sound event detection and direction of arrival estimation

©佐藤遼太郎, 丹羽健太, 小林和則 (NTT メディアインテリジェンス研究所)

課題

- 多チャネル信号の物理的(空間的)性質は人間には既知の知識
例) アンビソニクス形式の立体音響信号は線形変換で回転可能
- しかし現状の DNN はこの知識を活用できておらず
全方向からの多様な音データから空間的な性質を習得している
- この知識を DNN に「ハードコーディング」するのが自然では?

提案法

- ✓ 回転・スケール変換・時間並進の変換(群)に同変(equivariant)なアンビソニクス信号処理 DNN 構造を提案

$$f_{\text{DNN}}(\varphi_g(\mathbf{x})) = \psi_g(f_{\text{DNN}}(\mathbf{x})) \quad \forall (g, \mathbf{x})$$

音響信号 $\mathbf{x}$ を変換 φ_g してから DNN に通しても DNN に通してから変換 ψ_g しても出力結果は常に同じ

- ✓ あらゆる地点からの到来音を同等に扱うことが保証される

適用例

- ✓ 音響イベント検知・音源方向推定タスクに本手法を適用
- ✓ 音源方向が偏った学習データから全方向検知可能モデルを得た

	SED ER	DOA [°]
Conv.	0.320	13.9
Conv. + Rotation DA	0.233	8.5
Proposed	0.181	10.4

2-1-3

2-1-3 決定木に基づく軽量な音響イベント検知の検討

A study of lightweight sound event detection method based on decision tree

©村田伸, 齊藤翔一郎, 小林和則, 中川朗 (NTT)

- ◆ 音響信号からの周囲環境のイベント検知は、車両内の赤ん坊の置き去り検出や家庭内の見守りなど様々なアプリケーションが期待される。
- ◆ イベント検知においては、エッジ端末のような演算性能が限られた状況であることも多く、低演算量で検知可能な技術が必要となる。低演算で高精度な手法として経験的には決定木に基づく手法が知られている。本発表では、決定木を用いた分類手法を音響特徴量の時系列に適用可能に拡張し、AudioSet データを用いて模擬的な赤ちゃんの泣き声検知の評価を行った。

- ◆ Fig. 1 に学習データ量と検知性能の関係を示す。比較手法として、時系列特徴量をそのまま決定木に入力した場合と、CNN を用いた手法を示している。提案手法が少量の学習データでも精度よく学習できることがわかる。また、検知にかかる時間も CNN 比で 100 倍以上拘束であり、軽量な音響イベント検知手法として期待できる。

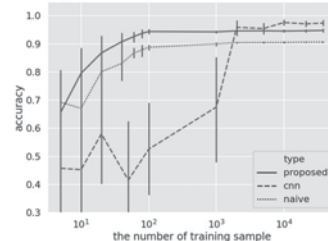


Fig. 1 Dependency of accuracy on the number of training examples.

2-1-5

2-1-5 音源数と音源位置を同時推定する視聴覚統合 DNN の自己教師あり学習

Audio-visual self-supervised learning for estimating number and locations of sound sources

©升山義紀(早大理工/産総研), 坂東宜昭, 佐々木洋子, 大西正輝(産総研), 矢田部浩平, 及川靖広(早大理工)

- ◆ 目的: 音源数推定する DNN と音源定位する DNN の自己教師あり学習
- ◆ 従来手法: 画像 + 単チャネル音響信号
 - 画像と音の共起関係に基づいて DNN を学習
 - 音源になりうる物体が複数存在する場合、どれが実際の音源か識別困難
- ◆ 提案手法: 全方位画像 + 多チャネル音響信号 + ベイズ推論
 - 音の空間情報を利用し、複数の音源になりうる物体を空間的に識別
 - 音源方向と混合比を潜在変数に持つ混合複素ガウスモデルを定式化
 - 潜在変数の事後分布を近似するように 2 つの DNN を学習
 - ◇ 音源定位 DNN: 音源になりうる物体 (例えば人物) の方向を推定
 - ◇ 混合比推定 DNN: 各物体から実際に音が発せられているかを検証

◆実験結果

- 提案手法により教師データを用いずに学習行なった DNN は、MUSIC 法を含む従来手法を上回る音源定位、音源数推定性能を実現

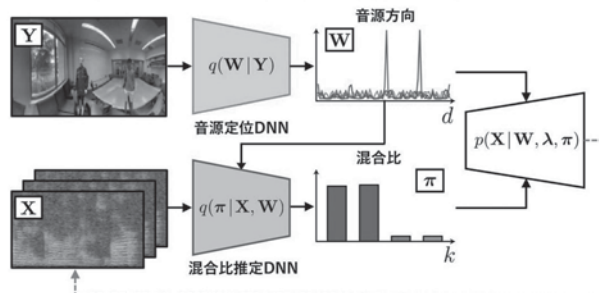


Fig. 1 Illustration of proposed self-supervised learning.

2-1-6

2-1-6 瞬時周波数計算のための
微分窓のサイドローブ最小化Sidelobe minimization of differential window
for instantaneous frequency computation

☆草野翼, 矢田部浩平, 及川靖広(早大理工)

- ◆ 近年、音響信号処理において短時間 Fourier 変換 (STFT) の瞬時周波数が利用されている。
- ◆ STFT の瞬時周波数計算では窓関数とその微分窓が用いられ、推定結果はそれらの特性に大きく影響を受ける。
- ◆ 本研究では、微分窓の周波数応答のサイドローブエネルギー (SE) と最大サイドローブレベル (HSL) の最小化に基づいて設計された、瞬時周波数の計算に適した二種類の窓関数を提案する。
- ◆ 微分窓のサイドローブを最小化することにより (Fig. 1), 瞬時周波数計算に対するサイドローブの影響を低減し、瞬時周波数の応用である synchrosqueezing の性能向上を確認した。

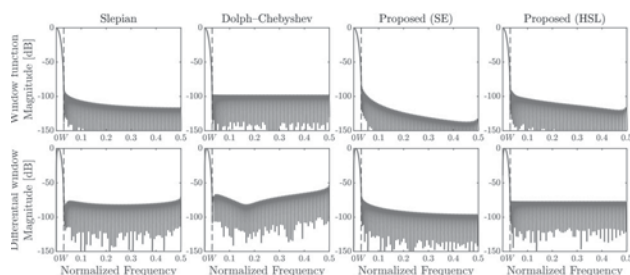


Fig. 1 Frequency responses of designed windows. The top and bottom rows show frequency responses of designed windows and their differential windows, respectively. Their frequency responses were normalized so that the maxima are 0 dB.

2-1-8

2-1-8 複素 DNN を用いた
深層 Griffin-Lim 位相復元

Deep Griffin-Lim iteration using complex DNN

◎ 升山義紀, 矢田部浩平 (早大理工), 小泉悠馬, 原田登 (NTT), 及川靖広 (早大理工)

- ◆ Griffin-Lim algorithm (GLA)
 - 短時間フーリエ変換の冗長性を利用した反復位相復元アルゴリズム
 - 再合成信号に人工的な雑音が発生
- ◆ 深層 Griffin-Lim 位相復元 (DeGLI)
 - GLA の各反復における出力に対し、DNN を用いた雑音除去を適用
 - DNN は、GLA の中間変数にあたる複数の複素スペクトログラムから、雑音成分の複素スペクトログラムを推定
 - ⇒ 複素数としての特徴を考慮できる DNN の設計が必要
- ◆ DeGLI のための複素 DNN: ゲート付き複素畳み込み
 - 複素スペクトログラムの隣接関係をとらえることのできる 複素畳み込み
 - 復元対象の振幅と 複素特徴量の振幅に基づき情報を抽出するゲート構造

$$\text{AGC}_C(C) = \text{AGate}(C) \odot \text{Conv}_C(C)$$

$$\text{AGate}(C) = \text{Sigmoid}(\text{Conv}_R([A, |C|]))$$

- ゲート付き実畳み込みなどを用いた場合と比較 PESQ, STOI を改善
- GLA, WaveNet (WN), WaveGlow (WG) と比較し、有効性を確認

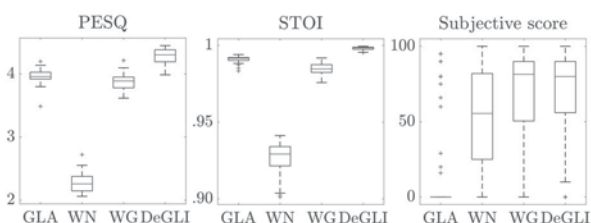


Fig.1 PESQ, STOI and score of subjective test.

2-1-7

2-1-7 非凸スパース最適化による
Audio Declipping

Audio declipping by non-convex sparse optimization

☆田中僚郎, 矢田部浩平, 及川靖広 (早大理工)

- ◆ 目的: Hard-Clipping が生じた信号から、元の信号を推定する
- ◆ 従来手法: フィデリティとスパース正則化による最適化
- ◆ スパース正則化には **L0 ノルム** や **L1 ノルム** が用いられる
 - L0 ノルム: 理想的であるが、組み合わせ最適化になる
 - L1 ノルム: 解きやすいが、**バイアス**を持つ
- ◆ 提案手法: **スパース性を誘導する非凸関数**を用いる
 - L0 正則化より解きやすく、L1 正則化よりバイアスが小さい
- ◆ 実験結果: 短時間で従来手法と同程度の性能を確認

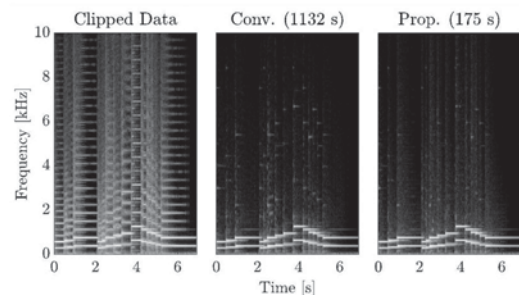


Fig.1 The Spectrogram of clipped data and declipped data

2-1-9

2-1-9 回転移動に頑健なアレイ信号処理
のための音場の補間に関する一検討A study of sound field interpolation
for rotation-movement-robust array signal processing

◎ 若林佑幸, 小野順貴 (東京都立大)

- ◆ 多くのアレイ信号処理は音源とマイクの位置 (系) の時不変性を仮定しているが、実応用ではこの仮定は常に満たされず、系が変化する度にフィルタ更新などの複雑な演算を繰り返し行う必要がある。
- ◆ 我々は円状アレイの利用を想定し、アレイが回転移動しても、円周上の音場を補間することで、アレイ処理の演算コスト低下を可能にするような枠組みを提案する。
- ◆ 本研究ではまず、アレイの回転移動と円周上の連続音場のサブサンプルシフトとの関係を示し、フーリエ変換の時間シフト定理によって、この関係の定式化を行った。結果として単純な線形変換でアレイが回転したときの観測スペクトルが推定できる。
- ◆ 本実験では、推定性能確認のための収録シミュレーションを行った。任意の角度 Δ 回転した M チャネルアレイにより収録されるスペクトルを推定し、実際の収録スペクトルとの SN 比で評価した。結果として、低帯域において推定できることを確認した (Fig. 1 参照)。

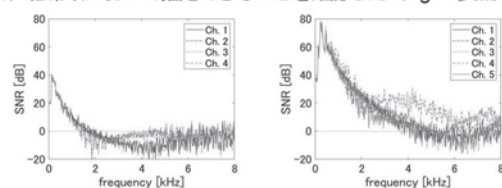


Fig. 1: Channel-wise SNR, $M=4$ (left), $M=5$ (right), rotation angle $\Delta=20$ deg.

2-1-10

2-1-10 音源の特徴に適応する
スパースネス統合 IVA

Sparseness-integrated IVA adaptive to source characteristics.

☆牛島隆裕(大分大), 太刀岡勇気(デンソーITラボ), 上ノ原進吾,
古家賢一(大分大)

- ◆近接分離最適化を用いたブラインド音源分離は、音源項を容易に交換可能である利点を持つ。音源分離に必須の白色化の影響を抑制するような音源項を用いたスパース独立成分分析(IVA)が提案されている。
- ◆本報告では、種々の音源での分離性能の向上のため、従来のスパース IVA で考えられている周波数方向のスパースネスの他、時間方向と到来方向のスパースネスの複数のスパースネス組み合わせたスパースネス統合 IVA に対し、分離の途中で音源の特性に適応することが可能なパラメータ設定方法を提案した。
- ◆実験では音声データのほかに音楽データに対し検討を行い提案法の有効性を示す。

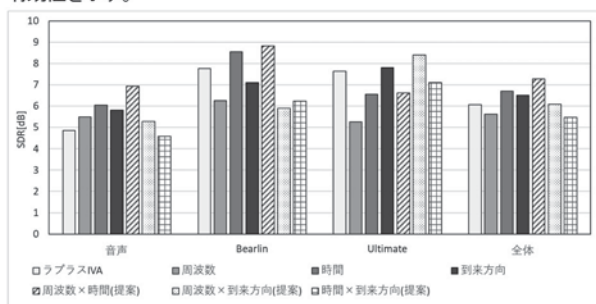


Fig.1: 音源に適応したスパースネス統合 IVA の分離結果

2-1-12

2-1-12 多重極再展開法を用いた
簡易音場再生手法のシミュレーション

Simulation of Tottedashi Using Multipole Reexpansion.

◎岩見貴弘, 柏崎紘, 尾本章(九大・芸工)

- ◆音場再現シミュレーションにおいて収録・再生素子は単純な指向性を仮定する場合が多い
- ◆本稿では多重極再展開法を用いることで複雑な指向性の素子に対して準理論解析的に音場再現シミュレーションする方法を検討する
- ◆提案手法を用いて簡易音場再生手法の単一周波数平面波の音場再現性能を評価したところ、高周波数域において素子配置及び指向性によっては高次アンビソニクスより優れた性能を示すことを確認した

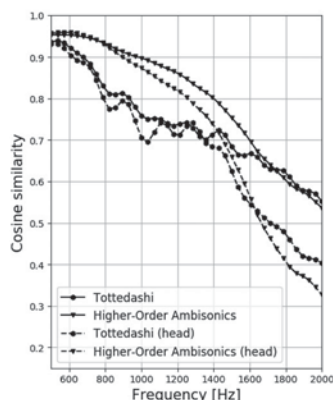


Fig.1 Cosine similarity of Tottedashi and Higher-Order Ambisonics

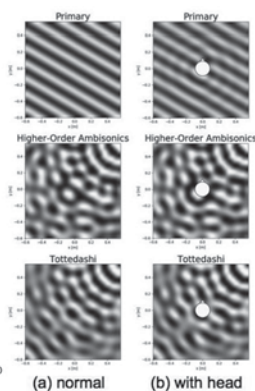


Fig.2 Sound field of each method (2 kHz)

2-1-11

2-1-11 A Generalized Minimal Distortion Principle to
Solve the Scale Ambiguity in Blind Source
Separation

© Robin Scheibler (LINE 株式会社)

- ◆We consider the scale ambiguity problem and the estimation of source images in determined blind source separation.
- ◆We propose the **generalized minimal distortion principle (GMDP)** that maximizes the likelihood with respect to a model.
- ◆We demonstrate the efficiency of the **mixed-norm** model that allows to tune sparsity in both time and frequency.
- ◆We propose to use an **MM algorithm** for the computation of the source images that is both simple and easy to implement.
- ◆We evaluate the performance of GMDP on popular BSS and joint BSS/dereverberation algorithms (AuxIVA, ILRMA, and ILRMA-T).
- ◆We find that GMDP allows to improve the SIR by up to 2 dB without any degradation of the SDR and at a small computational cost.

Algo.	Mics	PB		MDP		GMDP/SDR		GMDP/SIR		GMDP/SIR-10	
		SDR	SIR	SDR	SIR	SDR	SIR	SDR	SIR	SDR	SIR
AuxIVA [6]	2	9.72	22.78	9.50	22.16	9.99	23.94	9.51	25.04	9.51	25.04
	3	6.01	17.92	5.85	17.53	6.43	19.34	5.92	20.55	5.92	20.55
	4	-0.71	9.96	-0.72	9.78	-0.62	10.14	-0.72	10.45	-0.72	10.45
	2	7.06	21.55	7.36	20.19	7.91	20.94	7.46	21.96	7.46	21.96
ILRMA [10]	3	5.14	17.91	5.25	16.80	5.74	18.08	5.27	18.99	5.27	18.99
	4	-4.05	8.44	-3.24	7.96	-3.22	8.06	-3.22	8.11	-3.22	8.11
	2	5.99	10.02	5.87	9.84	6.23	10.26	6.09	10.39	6.14	10.38
ILRMA-T [11]	3	4.44	8.00	4.35	7.84	4.94	8.62	4.66	8.81	4.81	8.77
	4	0.32	3.37	0.07	3.32	1.26	4.19	0.19	4.56	1.23	4.22

Table 1: Projection back (PB), minimal distortion principle (MDP), and GMDP for best SDR, best SIR, and best SIR under 10 iterations.

2-1-13

2-1-13 剛楕円スピーカアレイを用いた
焦点音源生成についての検討

On Virtual Source Reproduction using Rigid Elliptical Loudspeaker Array

☆任逸, 羽田陽一(電通大)

- ◆高臨場感再生技術のひとつとして、音像がスピーカ面より前に飛び出しているような焦点音源(仮想音源)生成技術が発展している。筆者らはこれまで2つの剛楕円形スピーカアレイのモデルを提案し、焦点音源生成技術において、低域における放射効率の向上に成功した。
- ◆本研究では、これまでの結果を踏まえて2円アレイの非対称性に注目し、剛楕円スピーカアレイを用いて焦点音源を生成することを試みる。筆者たちが提案したマッシュ関数展開に基づく内側音場再現法をベースとし、剛楕円の反射や外側音場再現に適応した手法を提案する。焦点音源生成技術において、楕円の軸長や扁平率などによる再現性能への影響についても検討する。
- ◆計算機シミュレーションを行い、剛楕円スピーカアレイを用いた焦点音源の生成が確認できた。また、楕円の片方の軸長を固定した場合、扁平率が高いほど放射効率が高くなることや、面積および円周の長さを固定した場合、扁平率が低いほど放射効率が高くなることなどがわかった。

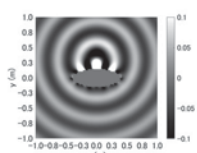


Fig.1: Wave front of reproduced virtual source at 1000 Hz

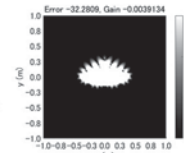


Fig.2: Reproduction errors at 1000 Hz

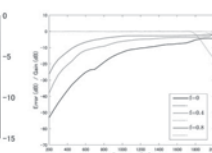


Fig.3: Error and gain for arrays with different flattening

2-1-14

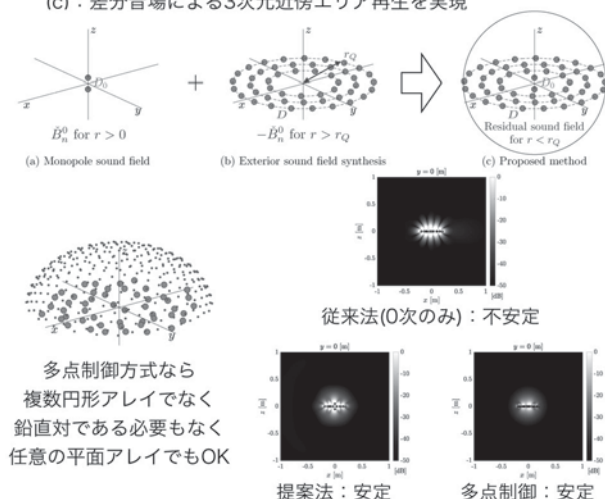
2-1-14 平面アレイと鉛直スピーカ対を用いた
3次元近傍エリア再生

Generating 3D localized sound zone using a planar array combined with a vertical loudspeaker pair

○岡本拓磨 (NICT)

Q: 「スピーカアレイの近くでのみ音が聞こえる近傍エリア再生の再生エリア形状を安定させたいのですが」
オカモト: 「それならこれで!!」

- (a): z軸上に配置した鉛直スピーカ対の放射音場を
(b): 複数の円形アレイで相殺
(c): 差分音場による3次元近傍エリア再生を実現



2-1-16

2-1-16 ガウス過程に基づく音場計測のための
センサ配置最適化

Optimization of sensor placement for sound field measurement based on Gaussian process

☆西田智哉, 植野夏樹(東京大), 小山翔一(東京大/JST さきがけ), 猿渡洋(東京大)

◆背景:

- 音場計測では複数のセンサを用いて音空間を推定・補間する
- 音場計測の精度はセンサ(マイクロフォン)配置に依存

◆目的: 推定対象領域とセンサ配置候補領域を独立に設定可能なセンサ配置

- 従来は推定対象領域=センサ配置候補領域を仮定
- 実用上, 配置候補領域が制限される場合が多い

◆提案手法:

- 推定の二乗誤差期待値・エントロピーの2種類の規準によるコスト関数を導出
- 貪欲法あるいは凸緩和法による近似解法を定式化

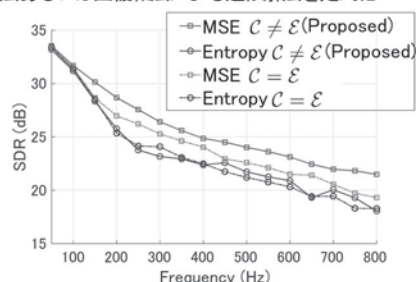


Fig.1: SDR of sound field reconstruction with respect to its frequency

2-1-15

2-1-15 音場制御におけるスピーカ・制御点配置
最適化法の比較評価

Comparison of source and sensor placement methods in sound field control

○小山翔一(東大/JST さきがけ), Gilles Chardon (CentraleSupélec),
Laurent Daudet (Institut Langevin)

- ◆音場制御の問題において, 二次音源(スピーカ)とセンサ(制御点)の配置は, 再現精度や逆フィルタの安定性に大きく影響する。
- ◆近年, 機械学習・センサネットワークの分野等で提案されている, 様々な二次音源・センサ配置最適化法を多点音圧制御法による音場制御の問題に適用し, 比較評価を行った。
- ◆制御対象領域の境界上に等間隔に配置する方法では, 禁止周波数の影響によって大きく精度が劣化する周波数が存在する。
- ◆比較を行った手法のうちのいくつかは, 禁止周波数の問題を回避しつつ, 高い再現精度を達成。
- ◆特に, 二次音源・センサ配置の同時最適化を可能とする, empirical interpolation method (EIM)が高い性能を示した。

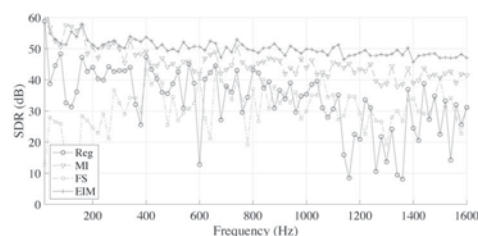


Fig.1: Reproduction accuracy was evaluated by signal-to-distortion ratio (SDR). SDRs of Reg, MI, FS, and EIM with respect to frequency are plotted.

2-2-1

2-2-1 Full-band LPCNet:
48 kHz リアルタイムニューラルボコーダ

Full-band LPCNet: A real-time neural vocoder for 48 kHz audio

☆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),
志賀芳則⁽²⁾, 河井恒⁽²⁾ (1 神戸大学, 2 情報通信研究機構, 3 名古屋大学)

- ◆近年の高品質音声合成はサンプリング周波数 24 kHz が多く検討されているが, 人間の可聴域は最大 20 kHz。
- ◆人間の可聴域をカバーする 48 kHz の音声合成は, 音質や合成速度に課題がある。
- ◆これらの課題を解決するフルバンド LPCNet を提案
- ◆48 kHz 音声合成でもリアルタイム性を維持, テキスト音声合成による主観評価では原音に匹敵する音質を達成した。

Table 1: Real-time factor for inference.

model	GPU	CPU	Multi CPU
BLSTM+Taco2dec	0.08	0.42	0.18(8)
Tacotron2	0.08	0.37	0.18(8)
WaveNet	672	-	6144(16)
Parallel WaveGAN	0.03	-	1.1(16)
LPCNet[384]	0.46	0.42	-
LPCNet[512]	0.62	0.61	-
LPCNet[640]	0.95	0.84	-

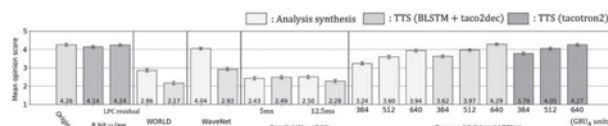


Fig.1: Result of MOS test with 25 listening subjects.

2-2-2

2-2-2 DNNに基づく音声ボコーダにおける
周期・非周期成分のモデル化の検討

An investigation of modeling periodic and aperiodic components
in speech vocoder based on deep neural networks

◎法野行哉, 高木信二, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一
(名工大)

本稿では、自己回帰構造を持たないニューラルボコーダにおいて、音声波形の周期成分及び非周期成分を適切にモデル化するため、並列型や直列型のモデル構造を検討する。並列型や直列型のモデル構造を用いることで、合成波形の自然性が向上した。サンプリング周波数が48kHzの音声波形を、実時間よりも高速に、自然音声と同等の品質で生成することを実現した。また、合成学習データの範囲外のピッチを持った音声波形の生成も可能であることを示した。

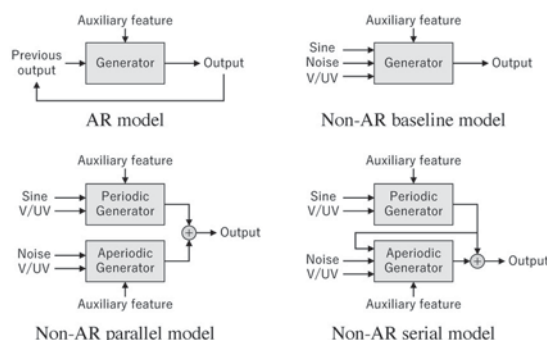


Fig. 1: Structures for speech waveform modeling

2-2-4

2-2-4 口唇動画をを用いた調音-音声変換の
大語彙連続発話への適用

Application of articulatory-to-speech conversion using lip movies to large
vocabulary continuous speech synthesis

☆田口史朗, 鍋木時彦

- ◆調音情報から音声への変換則を構築する調音-音声マッピングは、口頭摘出者などの音声障害者のための代用音声への応用が期待され研究が行われている。近年では、深層学習の登場によって、口唇動画を調音情報として用いた調音-音声マッピングの研究が行われるようになったが、多くは語彙が制限されたコーパスの利用にとどまっている。
- ◆そこで本研究では、畳み込み層を基底とした系列変換モデルを用いた口唇動画-音声変換法を提案し、TIMIT文章の読み上げ音声を収録したTCD-TIMITコーパスを用いて、最適な学習法の検討を行った。
- ◆本研究で検討した多くの手法で品質評価指標の改善が見られたものの、音源情報の推定精度の向上が達成されなかったため、新たな音源情報に関する制約を検討する必要があると考えられる。

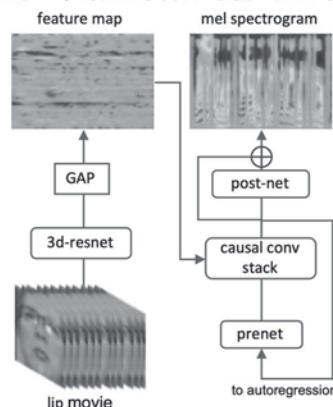


Fig. 1 Outline of the convolutional sequence-to-sequence lip-to-speech conversion model.

2-2-3

2-2-3 HMelGAN: 階層的構造を導入した敵対的
学習に基づく高速ニューラルボコーダ

HMelGAN: A fast neural vocoder based on generative adversarial networks
with hierarchical architecture

☆大西弘太郎, 中鹿亘, 松本光春(電通大)

- ◆リアルタイム推論可能な高速なニューラルボコーダとして、Generative adversarial network (GAN)を用いたニューラルボコーダが近年多数提案されている
- ◆Generator と Discriminator の両方が階層的構造を持つ hierarchical MelGAN (HMelGAN) を提案する
- ◆階層的構造により、低いサンプリング周波数から高いサンプリング周波数へと段階的に学習をでき、Discriminator が同時に複数のサンプリング周波数の信号を考慮できる
- ◆客観的評価と主観的評価の両方の実験で既存手法 (MelGAN) と同程度の計算量で同等以上の品質であることを確認

Table 1 PESQ and preference score with 95% confidence interval of
reconstruction speech.

Model	PESQ score	Preference score
MelGAN	1.83 ± 0.06	0.453 ± 0.046
HMelGAN (ours)	2.30 ± 0.05	0.546 ± 0.046
WaveNet	3.33 ± 0.02	N/A
Original	4.64	N/A

Table 2 PESQ and preference score with 95% confidence interval of speech from
generated Mel-spectrogram.

Model	PESQ score	Preference score
MelGAN	1.79 ± 0.06	0.506 ± 0.055
HMelGAN (ours)	2.12 ± 0.06	0.493 ± 0.055

2-2-5

2-2-5 音声プライバシーのための
ブラックボックス型音声加工法

Black-box voice modification method for voice privacy

☆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、音声データの匿名化を目的としたブラックボックス型のハイパーパラメータ最適化法を用いた手法を提案する。提案法のフロー図をFig. 1に示す。
- ◆EER と WER を用いて目的関数を定義し、設けた範囲の中で目的関数が最小となる音声加工法の最適なハイパーパラメータの探索を行なった。
- ◆様々な音声加工法を用い、それぞれの匿名化の達成度を得られた目的関数の最小値をもとに検証した。
- ◆音声のリサンプリングが最も匿名化の達成度が高かった。リサンプリングを施した際のEERとWERの推移をFig. 2に示す。

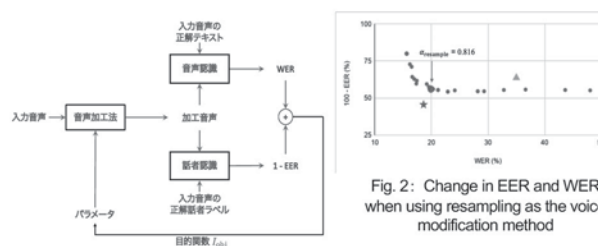


Fig. 1: Proposed method to optimize the hyperparameter of a voice modification method

Fig. 2: Change in EER and WER when using resampling as the voice modification method

2-2-6

2-2-6 実用的なささき声の生成法:
Phantom Silhouette 方式の改良

Practical whisper generation method: Improvement of Phantom Silhouette method

○内田照久(大学入試センター 研), 森勢将雅(明治大 総合数理)

目的 通常音声から、疑似的なさき声を生成するための、vocoderを利用したパラメトリックな変換法の改良と評価

方法 音声分析合成法 WORLD を使用, 合成時に操作を実現

(1) 低域抑圧と雑音駆動による無声化を拡張: ファントム化

(2) スペクトル軸の伸縮による F_1 , F_2 上方シフト量を制御

(3) 高域成分補填のための強調量を入力に対応して変更

評価 首都圏の国公立 11 大学 1 年生 123 名(男: 90 名・女: 20 名) 聴感的に「ささき声らしい」との高評価, 自然性には課題

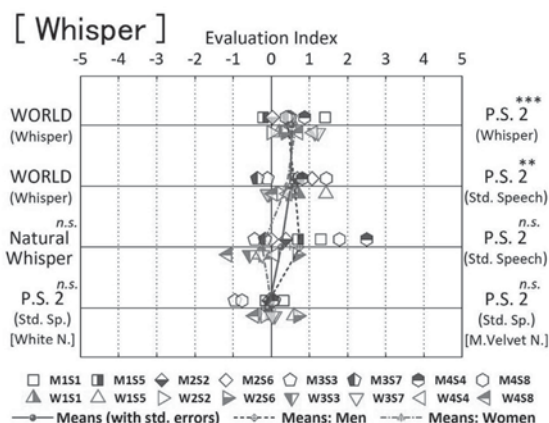


Fig.6 Evaluation index of whisper score for each comparison condition.

2-4-1

2-4-1 低速度の自動車接近音の知覚における
老人性難聴の影響: 模擬難聴システムによる検討

Effects of age-related hearing loss on perception of the low-speed approaching car sound: Investigation using hearing impairment simulator.

◎古屋孝基, 渡邊優也, 松井淑恵 (豊橋技科大院)

- ◆低速度の車走行音の聴取実験により、車両接近の認識や危険度の評価が老人性難聴によりどのように異なるか模擬難聴を用いて調査した。
- ◆車走行音の加工条件として、録音時の音圧を再現し (NH 条件)、70 代男性の聴力レベルを圧縮特性健全度 50%、100% で模擬した音 (HIS50% 条件、HIS100% 条件)、HIS50% 条件および HIS100% 条件と A 特性最大音圧レベルが同じになるよう音圧を低減させた音 (FA50% 条件、FA100% 条件) の 5 種類を設定した。
- ◆実験参加者は車走行音を聴取しながら接近車両を認識した瞬間にマウスのボタンを押下し、その危険度を 7 段階で評価した。
- ◆結果、老人性難聴により車両の接近を認識するまでの時間が健聴者よりも遅れ、危険度についても低く評価されることが示唆された。

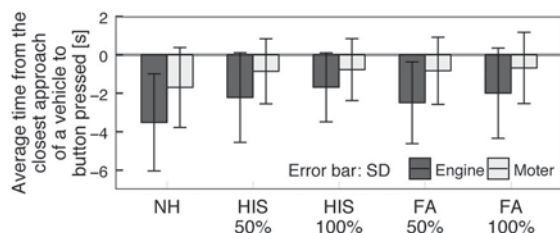


Fig. 1: Average time from the closest approach of a vehicle to button pressed. The ordinate 0 represents time of the closest approach of a vehicle. The positive values indicate that button pressed after the approach.

2-2-7

2-2-7 微分可能全域通過フィルタを用いた
ダイナミックレンジ圧縮

Differentiable all-pass filter for dynamic range compression

☆彦坂秀(名大), 関翔梧(NTT), 武田一哉, 戸田智基(名大)

- ◆観測信号の最大-最小音圧レベル差を低減するダイナミックレンジ圧縮 (Dynamic range compression: DRC) は、雑音環境下における音声の了解度向上や音圧増幅時における音響信号のクリッピング回避などに利用される。
- ◆高い音声了解度を保持することが可能である一方で、音声信号に対してゲインを操作するため、圧縮前後において音響特徴量の歪みが発生し音質が劣化するという問題がある。
- ◆本稿では、微分可能な全域通過フィルタ (All-pass filter) を用いたダイナミックレンジ圧縮法を提案する。
- ◆実験の評価より、1) 従来の DRC と同様に音声波形のダイナミックレンジを圧縮することが可能なこと、2) 従来の DRC と比較して圧縮率は大きいものの特徴量歪みを大幅に改善することが可能なこと、が明らかとなった。

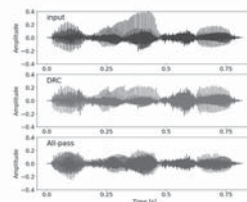


Fig. 1 Example of speech waveforms

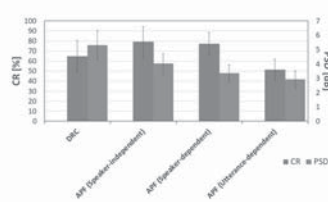


Fig. 2 Objective evaluation results

2-4-2

2-4-2 学童期の選択的聴取能力:
図と地課題および競合語課題を用いてSelective listening of school-age children
using figure-ground and competing words tasks

○加藤正晴, 嶋田容子(同志社大), 木谷俊介(北陸先端大)

子供達にとって教室内のような騒がしい状況で聴きたい音を聞く能力(選択的聴取能力)は極めて重要である。しかしいわゆる通常の聴力と比較して選択的聴取能力の発達はかなり遅く、しかもその原因は未だ十分にわかっていない。言語の音響的特徴にも影響を受けると考えられるが日本人の選択的聴取の発達についての研究もほとんどない。

本研究では雑音下での単語音声の聴取能力を測定する図と地 (Figure Ground: FG) 課題、左右の耳に呈示される異なる単語音声の聴取能力を測定する競合語 (Competing Words: CW) 課題を実施した。学童(小学2年生)および成人においてその成績を比較したところ、両課題とも学童の成績は成人の成績よりも低かった。日本人の選択的聴取の発達は英語圏の人々と同様ゆっくりと発達すると思われる。

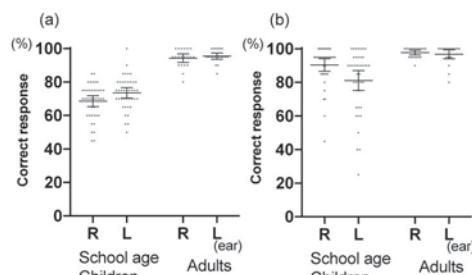


Fig.1: Percent correct response of FG (a) and CW (b) tasks as a function of ear (Dots: individual data, Bars and Error bars: the means and 95% CIs).

2-4-3

2-4-3 Discussion on cartilage conduction applied on mobile phones by measuring ear-canal sound pressure

© Xiuyuan Qin (Chiba Univ.), Sho Otsuka (Chiba Univ.), Junghun Noh (Samsung Display), and Seiji Nakagawa (Chiba Univ.)

- ◆ Cartilage-conduction (CC) has been proposed and applied to a hearing aid for atresia of the external auditory canal. Recently applications of CC on mobile phone have been developed.
- ◆ In this study, hearing thresholds and ear-canal sound pressure (ECSP) of CC were measured with different given placements by using a smartphone panel. A comparison between CC and air conduction (AC) was made, and the influence of placement on CC was discussed.
- ◆ The measurements were done when the panel contacted or did not contact the pinna as well as the panel covered the pinna fully and partly. The ECSPs on the contacted condition were higher than the results obtained on the uncontacted condition, showing the influence of CC by vibration of mobile phone.

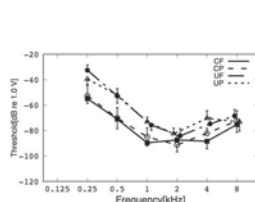


Fig. 1 The hearing thresholds on four different conditions.

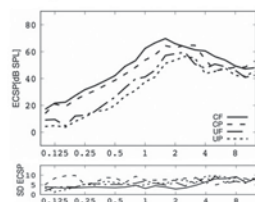


Fig. 2 The hearing thresholds represented by ECSP on four different conditions.

2-4-5

2-4-5 骨伝導イヤホンを使った雑音環境下での聞こえ

A study of listening capability of bone conductive earphone under noisy environment

☆堀内奨介, 高橋徹, △赤塚俊洋, △山田耕嗣, 中山雅人(阪産大)

- ◆ 環境雑音下での骨伝導イヤホンおよびオープンエア型ヘッドホンでの聞こえ方の違いを単語聞き取り試験により比較
 - ◆ 初期的検討として白色雑音下で比較
 - ◆ 骨伝導イヤホンの使用により MOS 値で 0.5~1.0 程度改善
 - ◆ 環境雑音レベル LAeq 67.5~75.0 dB 下で骨伝導イヤホン 4.0 以上
- 1: 全く聞こえない, 2: 聞こえない, 3: どちらでもない, 4: 聞こえる, 5: よく聞こえる

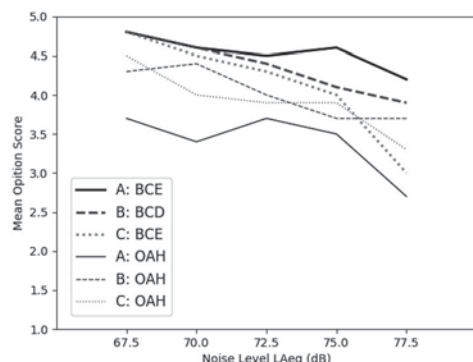


Fig. 1: Mean Opinion Score for 3 subjects (Thick lines: Bone Conductive Earphone, Thin lines: Open-Air Headphone)

2-4-4

骨導ヘッドホンを用いたダイコティック聴取及びモノラル聴取での非同時マスキング効果の計測*

Measurement of non-simultaneous masking effects in monaural and dichotic listenings by using bone conduction headphones

○平山龍朗, 岩城護(新潟大院・自然研)

- ◆ 骨導音聴取時に生じる順向性、逆向性マスキング効果をモノラル聴取とダイコティック聴取に対して調査した。
- ◆ 実験条件の中で最長の ISI における最小可聴値を基準として、そこからの上昇量をマスキング量として評価した。
- ◆ 骨導音聴取時にも非同時マスキング効果が確認された。気導音聴取時とは異なる特性だった。
- ◆ 順向性マスキングでは、モノラルの場合は気導音聴取の方が、またダイコティックの場合は骨導聴取の方が大きかった。

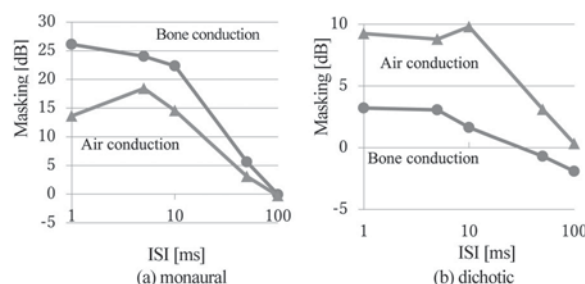


Fig. 1 Forward masking level of air conduction listening and bone conduction listening

2-4-6

2-4-6 両耳聴タスク遂行時の聴性脳幹反応と眼球反応との関連

Relationship between auditory brainstem frequency following response and eye responses during a dichotic attention task

○山岸慎平, 古川茂人 (NTT CS 研)

- ◆ 聴覚系の各段階の神経活動は選択的注意やタスクへの集中などによって動的に変動することが知られている。聴覚野やそれ以上のレベルでは注意の影響の証拠は多いが、脳幹レベルの活動については結果が一貫していない。
- ◆ 本研究では、眼球反応と脳波を同時に計測し、微小眼球運動 (Microsaccade: MS) が左右方向への注意を、瞳孔径の大きさが覚醒度を反映すると考え、両耳聴タスク中の脳幹反応との関連性を調べた。
- ◆ 両耳聴タスク中、MS は聴覚的注意の反対方向に偏る傾向があった。MS の方向が聴覚的注意方向の指標となることを示唆する。
- ◆ MS の方向と左右耳の音の方向が不一致であった場合、一致していた場合よりも脳幹反応が大きかった。これは MS の方向がタスクで課された注意方向と反対に偏る結果と一貫している。
- ◆ 中程度の瞳孔径の場合に脳幹反応の SN 比が最も高いことがわかった。これは中程度の覚醒度の際にパフォーマンスや神経活動が最適化されるという考えと一致しており、脳幹レベルでもそういった関係性があることを示唆する。
- ◆ 本研究の結果は、注意状態によって脳幹反応が時間的に変動することを示唆している。1000 試行以上の加算平均を行う従来の脳幹反応の計算方法では動的な変動を見逃していた可能性が高いと考えられる。

2-4-7

2-4-7 快適かつシャープネス値の大きい 自然環境音の間欠呈示による作業効率向上

Work efficiency improvement by intermittently presented comfortable sounds with high sharpness values

○星野博之, △白戸宏季, △永田知大(愛知工業大学)

- ◆目的 運転中などの人が活動する各種作業環境において、人の状態を適切かつ快適に保つための音刺激を実際の自然環境音から探索、創成することを目的としている。今回は、4種類の自然環境音(長さ5秒)を間欠的に呈示しながら作業課題(暗算課題、タイピング課題)を行う実験による作業効率向上(覚醒効果)の評価を行った。
- ◆方法 快適さ主観値が高く、シャープネス値が2acumより大きい「野鳥コーラス」「溪流」「砂浜」「フクロウと虫」の4音を用いて作業課題実験(各課題で条件ごとに5分間、被験者7or8名)を行った。
- ◆結果 防音室内において、無音および「野鳥コーラス」を10秒、20秒、30秒、40秒間隔で呈示した状況でタイピング課題を行った場合のタイプ数の結果をFig.1に、4音を40秒間隔で呈示した場合の暗算課題の解答数の結果をFig.2に示す(**: $p < 0.01$, *: $p < 0.05$, +: $p < 0.1$)。結果として、今回の4音の中では「野鳥コーラス」の20秒、30秒間隔での呈示により作業効率が向上する可能性が示された。

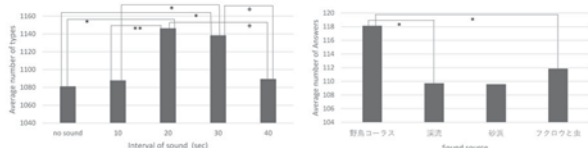


Fig.1 Average number of types when the sound "yachokohrasu" was presented

Fig.2 Average number of answers of mental arithmetic when each sound was presented

2-4-8

2-4-8 2純音の知覚的同時性の範囲の法則および複合音への適用

An application of the law of perceptual simultaneity range to complex tones

◎岡崎聡, 津崎実(京都市立芸大)

- ◆一連の研究から、我々は、あらゆる2純音の周波数組み合わせに対して同時性の範囲を予測可能な法則「岡崎・一川・津崎方程式」の導出に成功した。
- ◆本研究は、この法則が複合音の同時性の範囲にも適用可能か検討した。
 - 測定: 2純音調波複合音に対する同時性の範囲を測定した。
 - 予測: 岡崎・一川・津崎方程式を複合音中の各純音成分に対して適用し、平均をとった。
- ◆検討の結果、同時性の範囲の測定値と予測値は概ね重なった(Fig.1)。
- ◆以上から、周波数の次元において、純音の同時性の範囲を予測可能な岡崎・一川・津崎方程式は、複合音の同時性の範囲の予測にも適用可能であることが示唆された。

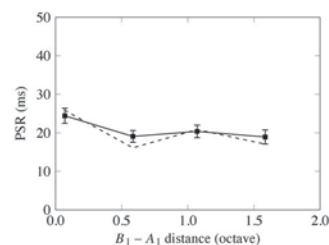


Fig.1 Perceptual simultaneity range of two-tone complex. Observed (with error bar) and predicted (dashed)

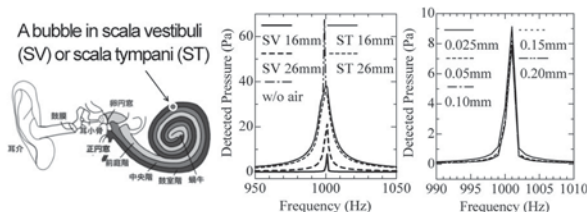
2-4-9

2-4-9 蝸牛内への気泡の侵入による聴覚特性への影響

Effects of Air Bubbles in Cochlea on Sound Detection of Auditory System

☆佐藤圭祐, 洪文甲, 玉木愛莉, 堀井康史(関西大)

潜水時の急浮上や外リンパ液の亀裂により蝸牛腔内に気泡が生じることがある。このとき、前庭階に気泡が入ると著しい難聴をきたすが、鼓室階に気泡が入っても聴力特性への影響はほとんど見られないという。また、極端な例として、鼓室階のリンパ液をすべて空気に置き換えると、蝸牛マイクロフォン電位が一時的に観測されなくなるものの、再び人工リンパ液を鼓室階に注入すると可逆的に聴力を取り戻すという。本稿では、そのメカニズムを著者が提唱する蝸牛定在波理論をもとに波動論的に説明する。気泡を含む蝸牛定在波モデルを構築し、COMSOL Multiphysicsを用いて解析を行ったところ、図1に示す通り、上記の報告に一致する結果を得ることに成功した。そして、これまで不明とされてきた気泡含有時の蝸牛内部の音響的振る舞いを明らかにするとともに、気泡位置と聴力特性との関係について説明を行った。



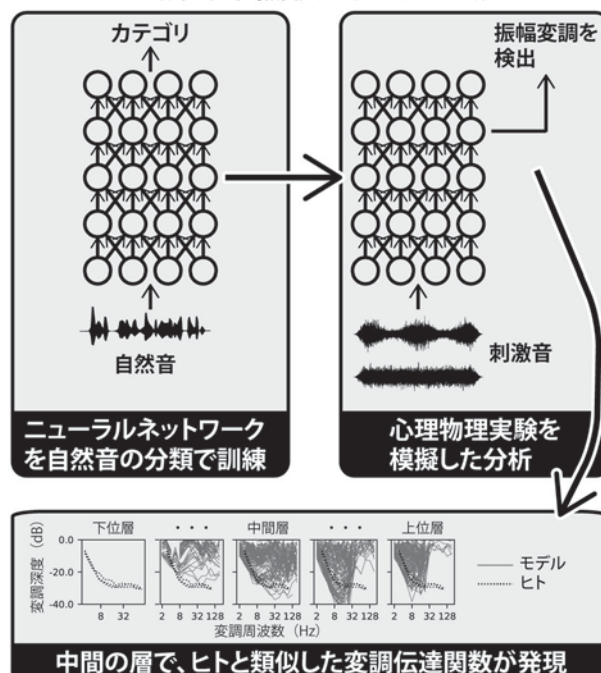
(a) A bubble in cochlea (b) Position of bubbles (b) Diameter of bubbles
Fig.1: Detected sound pressure by 1000 Hz outer hair cell

2-4-10

2-4-10 自然音分類で訓練した ニューラルネットワークにみられる 心理物理学の変調伝達関数

Psychophysical Modulation Transfer Function in a Neural Network Trained for Natural Sound Classification

○上村卓也, 寺島裕貴, 古川茂人(NTT CS 研)



2-5-1

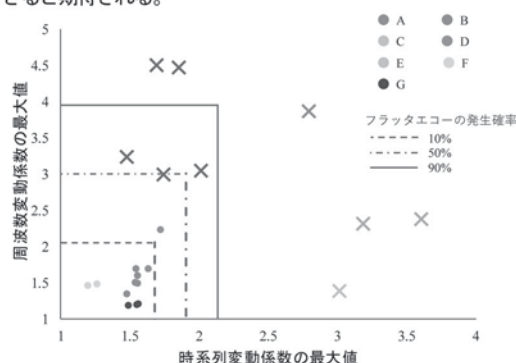
2-5-1 減衰除去インパルス応答の変動係数に基づくフラッタエコー発生確率の評価

Evaluation of occurrence probability of flutter echo based on variation coefficient of the decay-cancelled impulse response

○羽入敏樹, 星和磨, 鈴木諒一(日大・短大)

下山達志(ミサワホーム総研), 横田雅人(日大・院)

- ◆本研究では、減衰除去インパルス応答を用いた音場の拡散性の評価指標である「時系列変動係数」と「周波数変動係数」を用いたフラッタエコー発生評価手法を提案した。提案手法によって、フラッタエコー発生を判定する可能性を示した。
- ◆本手法により、音響シミュレーション等によるインパルス応答から、設計段階でフラッタエコーの発生を予測できる可能性がある。また、フラッタエコーが生じている実音場での低減対策効果を定量的に評価できると期待される。



2-5-3

2-5-3 深層学習を用いたパブリック空間における背景音内の聴取音声の音環境印象に及ぼす影響度評価についての基礎的検討

Evaluation Method of Acoustic Comfort in Public Space using Deep Learning Considering Audible Voice in Background Sound

○伊藤夏美(日本音響エンジニアリング), 橋本修(日大理工)

現代のパブリック空間は使われ方が多様化し、空間の印象には音環境要素の中でも周囲の会話音などの音声の聴き取り度合いが影響していると考えた。従来の音声聴取の評価法は単一音源で定常的な音環境を想定されており、多音源かつ非定常的であるパブリック空間の音環境では評価が難しいと考えられる。そこで本研究では、機械学習手法の一つである深層学習を用いて、パブリック空間の実録データから時間変化を伴う音声の聴き取り度合いを評価し、その背景音が音環境の印象評価に及ぼす影響度を考察した。この学習結果を用いて実録の背景音内の音声の聴き取り度合いの時間変化と音圧レベルの時間変化を比較した結果、聴き取り度合い変化とレベル変化は必ずしも一致しないことが示されたため、レベル変動だけでは音環境変化の全てを説明することはできないと言える。また、音声だけでなく他の妨害要素を考量したアルゴリズムを考え、それらの複合評価を行うことで、より詳細な音環境評価が可能になり、パブリック空間だけでなく音声聴取の影響を受ける様々な用途の空間においても応用できると考える。

2-5-2

2-5-2 異なる座席位置で得られるインパルス応答の主観的類似性に関する研究

A study on subjective similarity in impulse responses in audience area

○徳永泰伸(舞鶴高専), 寺島貴根(三重大・工)

- ◆音楽ホールで開催される演奏会では、客席部がゾーン分けされる場合がある。この場合、各ゾーンによって価格設定がなされ、座席には優劣が存在する。しかし、優劣のついた各ゾーンにおける音響状態の違いを知覚できなければ、音響的な側面では差異がないものといえる。
- ◆小規模ホールの客席部で採取したインパルス応答を対にした聴感比較試験を行い、被験者が回答した客席部の音響状態についての類似性を求め、その結果を多次元尺度法(MDS)により処理した。
- ◆多次元尺度法からは布置図が得られるが、その座標値と音響指標の関係性が把握できれば類似性の判断に関する手がかりが明らかになる。
- ◆実験の結果からは、対象とする客席部範囲の大きさによって、類似性の判断の手がかりとなる音響指標は変化する可能性が示された。

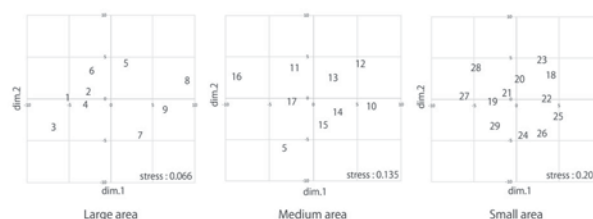


Fig.1: Results of MDS

2-5-4

2-5-4 調剤薬局におけるスピーチプライバシー保護のためのマスキング手法

Masking method for speech privacy protection in dispensing pharmacies

○石川あゆみ(岐阜高専)

- ◆調剤薬局におけるスピーチプライバシー問題対策として、スピーチプライバシーシステムの導入が効果的であることが明らかになっているが、その導入コストは高額である。
- ◆本報では、調剤薬局関係者へのスピーチプライバシーに関するアンケート調査、指向性の異なる3種のスピーカーと3種のマスキング音を用いた場合の会話のマスキング効果等に関する主観評価実験を行った結果を示す。

Kind of speaker		
No.	Feature	Pattern number
1	Speech privacy system	YAMAHA VSP-1
2	Small speaker for small stores and offices	BOSE 100PR
3	Super directional speaker	HSS audify mini

Kind of masking sound		
Name	Content	Feature
n	River murmuring	Synthetic sound of natural sound and voice
sn	Street noise	Voice, footsteps and vehicle traffic noise, etc.
pn	Pink noise	General test sound
m	String quartet	Dry source of classical music

2-5-5

2-5-5 交通騒音のうるささ評価に受聴環境が及ぼす影響に関するシミュレーション実験

Effect of the listening situation on noisiness evaluation of traffic noises using audio-visual simulation system

○池田拓海(東大院), 小野寺拓馬(元東大院)

米村美紀, 李孝珍, 坂本慎一(東大生研)

- ◆ 室外および室内環境で交通騒音のうるささ評価に視覚情報が及ぼす影響を調べる目的で、視覚情報の提示にドームスクリーンを用いて主観評価実験を行った。
- ◆ 室外と室内環境それぞれに応じた映像と試験音を同時に提示し、ME法によって聞こえてくる交通騒音のうるささを評価させた。
- ◆ 室外環境を想定した条件では、映像を提示する方が交通騒音のうるささが小さく評価される結果となった。一方で室内環境を想定した条件では、逆に映像を提示しない方が、うるささが小さく評価される結果になった。これは、交通騒音のうるささ評価における映像提示の影響が、受聴環境によって異なることが示された。

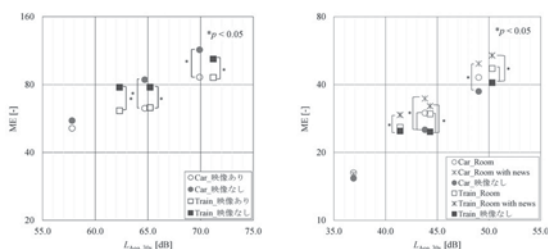


Fig.1 ME値と試験音の $L_{Aeq,30s}$ の関係(室外) Fig.2 ME値と試験音の $L_{Aeq,30s}$ の関係(室内)

2-5-7

2-5-7 FMBEMに基づく入射指向性解析に関する基礎的検討

Basic study on incidence directivity analysis based on FMBEM

○榎本貴之(東大・新領域/サイバネット), 井上尚久(前橋工大), 安田洋介(神奈川大・工), 佐久間哲哉(東大・工)

- ◆ 高周波数用FMBEMの後処理における入射指向性の効率的解析方法の構築を目指し、基礎的検討を行った
 - ◆ グリーン関数の平面波展開は次式で与えられる
- $$G(\mathbf{r}_p, \mathbf{r}_q) = \frac{1}{4\pi} \oint E_{pL}(\mathbf{k}) T_{LM}(\mathbf{k}, \mathbf{r}_{LM}) E_{Mq}(\mathbf{k}) d\mathbf{k}$$
- ◆ 次式の ζ が受音領域における入射指向性に相当する
- $$\zeta(\mathbf{k}) = T_{LM}(\mathbf{k}, \mathbf{r}_{LM}) E_{Mq}(\mathbf{k})$$
- ◆ Fig. 1に示す検証モデルで求めた $\zeta(\mathbf{k})$ は、Fig. 2左図の通り多重極展開点の位置に依存した
 - ◆ ζ を局所展開係数に変換した上で、低次項のみから指向性を再計算すると、Fig. 2右図の通り、多重極展開点の位置に依存せずに指向性が推定された

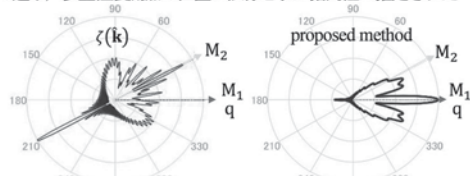
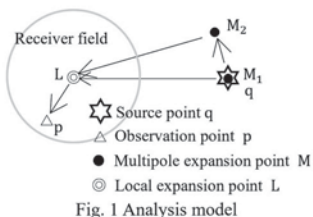


Fig. 2 Amplitude of incident directivity.

2-5-6

2-5-6 大学図書館のラーニング・コモンズにおける学習活動に及ぼす周囲音環境の影響

Effect of ambient acoustic environment on learning activities in the learning commons in academic library

☆丸山直也, 川井敬二(熊本大)

- ◆ 大学図書館のラーニング・コモンズにおいてアンケート調査・騒音レベル測定・利用状況観察を実施した。
- ◆ 学習環境やにぎやかさなどの音環境に関する選好度が利用エリアの選択に関係していることが示され、比較的高い周囲の騒音レベルによって「周囲の会話が気になる」傾向がみられた。
- ◆ 隣卓での大きめの会話音によって、学習の集中への妨害など心理的な影響が有意に示された (Fig.1)。

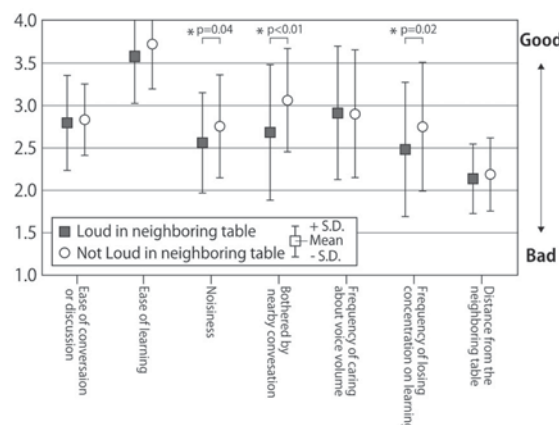


Fig.1: Comparison of subjective evaluation with/without loud conversation at neighboring table

2-5-8

2-5-8 斜入射解析のためのFDTD法を用いた音場における分散性及び安定性の検討

Dispersion and stability analyses of the FDTD method for a sound field under an oblique incidence

◎竹村美穂(日本建築総合試験所), 豊田政弘(関西大)

- ◆ 1次元的に周期配置した構造の斜入射透過損失は、周期構造の規約部分のみを解析領域としてFDTD法によって求めることが可能である。
- ◆ このとき、解析領域には連続境界条件を適用し、平面波の入射角に応じてFloque変換を施すことで斜入射が垂直入射と同等に扱える。変換によって支配式は新しい変数を用いたものに書き換えられる。
- ◆ 入射角 $\theta=0$ のとき、変換音場と従来のFDTD法の音場は一致する。しかし、 $\theta \neq 0$ の変換音場では波の位相速度が音速とは異なるため、変換音場は理論的に分散性をもった場であるといえる。さらに、FDTD法では空間や時間を離散化することから数値的な分散性も生じる。これらの分散性は波の挙動に影響を与えることがわかった。
- ◆ 安定性解析により、変換音場では従来のFDTD法の音場よりも細かい時間離散化幅を設定しなければ解が安定しないことが示された。

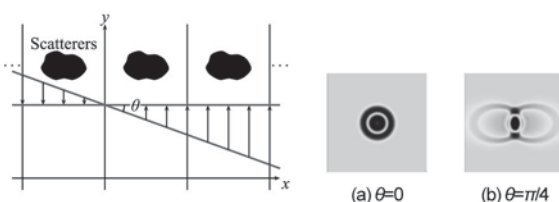


Fig. 1 Periodic structure in the x-direction Fig. 2 Sound pressure distributions at a certain time for the mapped fields.

2-5-9

2-5-9 周波数依存をもつ吸音境界を考慮した 時間領域 FEM による音響解析 —ADE 法を用いたインピーダンス境界—

Acoustic simulations using a time-domain FEM dealing with frequency-dependent absorption boundaries
—Impedance boundary using ADE method—

◎吉田卓彌(安藤ハザマ技研, 神戸大院), 奥園健, 阪上公博(神戸大院)

- ◆周波数依存のインピーダンス境界を有する室内音場解析のための陰的時間領域 FEM (TD-FEM) を提案した。
- ◆周波数依存の考慮に伴う畳み込みを効率的に扱うため、畳み込みを補助微分方程式の求解に置き換える Auxiliary Differential Equation 法を適用した。
- ◆伝達関数法に基づく音響管測定シミュレーションにより定式化の妥当性と補助微分方程式の求解に4次精度陽的ルンゲクッタ法 (RKM) の使用が有効であることを示した (Fig. 1)。また、追加の数値実験により、提案法の多孔質吸音体、ならびに、MPP 吸音体のモデル化への適用性を示した (Fig. 2)。

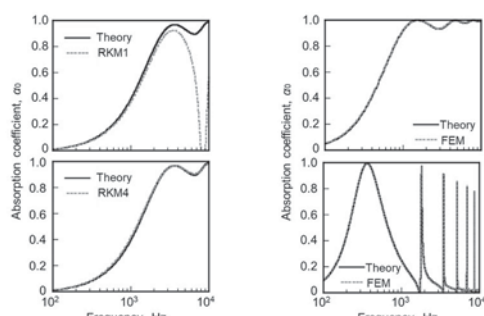


Fig.1 Comparisons of absorption coefficient of porous sound absorber calculated by theory and TD-FEM using 1st order RKM (RKM1) and 4th order RKM (RKM4).

Fig.2 Comparisons of absorption coefficient of porous absorber (Upper) and MPP sound absorber (Lower) calculated by theory and TD-FEM.

2-5-11

2-5-11 平面波エンリッチメントを用いた Partition of Unity FEM による室内音場解析 —微細穿孔板と通気性膜のモデル化—

Sound field analysis using plane-wave enriched FEM
—Modeling of microperforated panel and permeable membrane—

☆向江俊一, 奥園健, 田丸加奈子, 阪上公博 (神戸大院)

- ◆Partition of Unity FEM (PUFEM) に基づく室内音場予測手法は、Helmholtz 方程式の一般解を形状関数に導入することで場の近似を強化し高精度な解析を可能とする。
- ◆PUFEM 音場解析において、拡張作用性の吸音を示す吸音体として、微細穿孔板吸音体と通気性膜吸音体を扱う方法を提案する。
- ◆音響管モデルを対象に、理論値との比較から提案方法の妥当性を示す。次に微細穿孔板と通気性膜を含む2次元室内音場を対象に、従来FEMとの比較から PUFEM の有効性を明らかにする。
- ◆PUFEM は、著しく少ない自由度の単一の粗いメッシュと q -refinement を用いて、広帯域にわたる周波数応答をより高精度に計算できることを明らかにした。(Fig. 1)

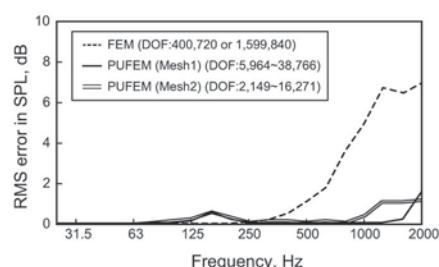


Fig.1: Comparisons of RMS error between PUFEM and FEM

2-5-10

2-5-10 周波数依存をもつ吸音境界を考慮した 時間領域 FEM による音響解析 —等価流体に基づく多孔質モデル—

Acoustic simulations using a time-domain FEM dealing with frequency-dependent absorption boundaries
—Modeling of porous materials based on equivalent fluids—

◎吉田卓彌(安藤ハザマ技研, 神戸大院), 奥園健, 阪上公博(神戸大院)

- ◆等価流体に基づく多孔質吸音体の時間領域有限要素モデルを提案する。
- ◆周波数依存の考慮に伴う畳み込みを効率的に扱うため、畳み込みを補助微分方程式の求解に置き換える Auxiliary Differential Equation 法を適用した。
- ◆2種類の多孔質吸音体 (PA-1, 2) を設置した音響管を対象とした数値シミュレーションにより定式化の妥当性を検証した。その際、修正積分法 (MIR) による高精度化 (Fig. 2) と陰的ルンゲクッタ法による安定性の改善を確認した。

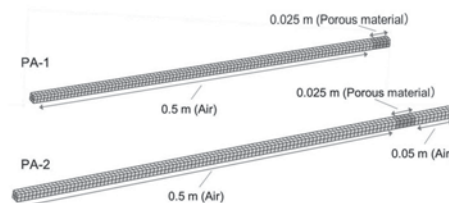


Fig.1 Finite element model of impedance tube including porous absorbers PA-1 and PA-2.

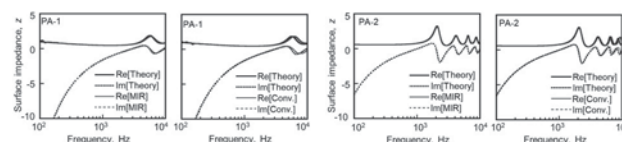


Fig.2 Comparisons of surface impedance of PA-1 (Left) and PA-2 (Right) calculated by theory, conventional TD-FEM and TD-FEM with MIR.

2-7-1

2-7-1 超音波音速CT装置を用いた建造物探査時の 再構成画像の精度向上

Improvement of accuracy of reconstructed images obtained by ultrasonic TOF-CT for structural inspection of buildings

◎林卓宏, 足立和成, 柳田裕隆 (山形大院理工学)

建造物の内部調査の際、非破壊的に内部状況を探査する方法の一つとして超音波音速CT装置の研究・開発を行っている。本研究で使用している装置は、対象物に取り付けた送信器から超音波パルスを打ち込み、受信器に到達するまでの透過伝搬時間 (TOF) を多数の経路について測定することで対象断面を相対的な音速分布の形で画像化するシステムである。その画像再構成においては、音波の伝搬経路を直線と仮定しているが、波動性等の影響により音波伝搬経路が直線から大きく逸脱して妥当な TOF の特定ができず、適切な画像が得られないことがある。そこで、波動伝搬シミュレーション並びに実験を行い、それらの結果に基づいて、超音波の TOF の同定方法を検討することで、より現実の断面に近い、妥当な再構成画像が得られるようにした。

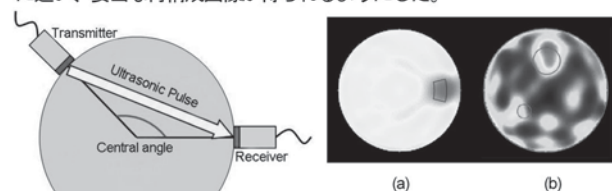


Fig.1 Schematic explanation of the ultrasonic TOF-CT system. TOF of ultrasound is measured for many propagation paths in the cross section.

Fig.2 Reconstructed TOF-CT images for (a) a simulated case and (b) an actual mortar pillar. The diameter of the both cross sections is 50cm.

2-7-2

2-7-2 強力超音波用ボルト締めランジュバン型振動子における部品界面の最適静的与圧に関する研究

Optimization of Static Prestress on the Interface Between Components of a Bolt-Clamped Langevin-Type Transducer for High-Power Ultrasonics

☆藤田捷希, 足立和成, 山吉康弘(山形大院・理工)

大振幅が必要な強力超音波用の振動源として、ボルト締めランジュバン型振動子(以下 BLT)が一般的に用いられている。圧電セラミックスと金属ブロックをボルト締めによって一体化させた構造の BLT において、それを高周波化すべく単にその全長を短くすると、それら部品間の界面外周部において静的圧縮と圧が不足し、大振幅駆動時に動的振動応力がそれを上回ること、部品間界面での接触が不安定になり、機械的損失の増大やそれに伴う特性劣化を引き起こす懸念がある。本研究では、そうした BLT の部品間界面における応力に着目し、有限要素解析によって、大振幅駆動時でもそこでの接触を保つ高周波用のその構造を検討した。BLT 製作時の締め付けにはデジタル角度レンチを使用し、製作時のその特性の再現性を高める方法を検討した。

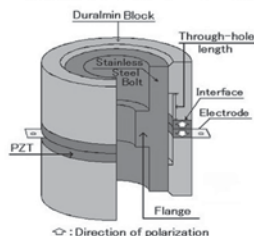


Fig. 1: Typical structure of a bolt-clamped Langevin-type transducer

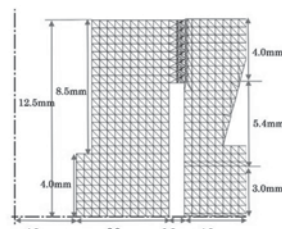


Fig. 2: Finite-element modeling of an optimized BLT structure

2-7-4

2-7-4 超音波接合のための切れ込みを入れた面対称構造を持つ縦一たわみ振動源の基礎検討

Basic Study of longitudinal-transverse vibration source with notched plane-symmetric structure for ultrasonic welding

○浅見拓哉, 三浦 光(日大・理工)

◆筆者らは、超音波複合振動を発生させるための振動源について検討を行っている。筆者らは、固定が容易で、かつ小型な単一振動子による縦一たわみ振動源を考えた。その振動源は、縦振動源の先端部に切り込みを用いることで、縦振動とたわみ振動の複合振動を発生させ、また切り込みに対して面対称構造とすることで発生するたわみ振動の位相を反転させ、先端部以外のたわみ振動を打ち消す構造とした。上述の構造の振動源を有限要素法解析ソフトウェアである COMSOL Multiphysics 5.5 を用いて基礎的な検討を行った。

◆Fig. 1 は、提案している縦一たわみ振動源の概略である。本稿では、図に示すような振動源の振動モード、及び振動分布を COMSOL Multiphysics 5.5 を用い求め、縦振動とたわみ振動の複合振動が得られるか、またたわみ振動の伝搬を打ち消すことが可能かなど検討を行った。

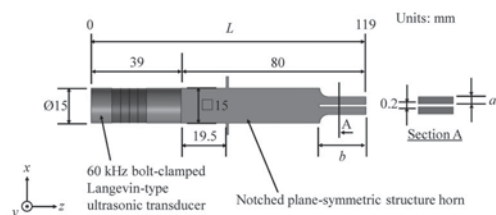


Fig. 1: Longitudinal-transverse vibration source.

2-7-3

2-7-3 超音波を用いた液体媒質の動粘度による温度上昇効果の検討

Examination of temperature rise effect by kinematic viscosity of liquid medium by ultrasound.

☆山田悠太郎(愛工大), 畑中信一(電通大),
△佐藤正典(本多電子), 小塚晃透(愛工大)

◆液体媒質中に超音波を照射させると熱エネルギーを生み出し、液体の温度は上昇する。

◆水槽内に比重の異なる液体媒質(水+シリコンオイル)を水槽内に入れた場合、超音波を照射すると液体媒質の温度が上昇する。このときシリコンオイルの動粘度別の結果を評価する。さらに、超音波を照射させた時、液体の境界面に現れる渦と温度上昇の関係について考察する。

◆水(界面活性剤入り)を媒質とし超音波を照射すると、液体の境界面に現れる渦の全長が大きくなり、シリコンオイルの温度上昇効果も低下した。

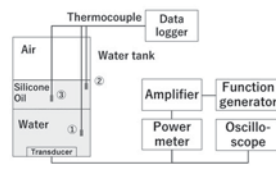


Fig. 1: Experimental apparatus.



Fig. 2: Boundary of liquid. (US ON)

2-7-5

2-7-5 角形状のステータを用いた多自由度超音波モータの試作

Design and measurement of multi-degree of freedom ultrasonic motor using polygonal stator

○小池義和, △佐藤康太(芝浦工大)

トルクを向上させるためにランジュバン構造を用いた超音波多自由度モータ(MDOF-USM)を提案している。実際に6枚のPZT素子を含んだステータを試作し、駆動を確認している。本報告では製作精度の改善のために八角形のステータを設計し、実現可能性について検討を行ったので報告する。有限要素法解析(FEM)を行い、多自由度駆動に必要な縦振動モード、縮退する2つのたわみモードの共振周波数を一致させることが可能なことを確認した。これまでと同様、2つのステータで球状のロータを挟み込む構造であり、駆動部と反対側に設けているバネ長により予圧を制御する。実際に円形状のPZTから超音波加工機により八角形のPZT素子を作成し、角形状の振動子となるステータを構成した。アドミタンス測定の結果、2つの縮退するたわみモードの共振周波数を確認している。今後、縦振動モードの改善し、駆動を試みる。

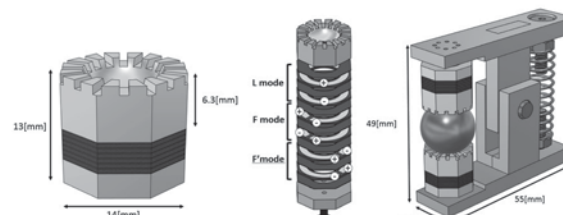


Fig. 1: Composition of an octagonal stator and whole structure of MDOF-USM.

2-7-6

2-7-6 マイクロバブルを含む水中における 圧力波の長周期非線形発展に 気泡抗力が及ぼす影響

An Effect of Drag Force Acting Microbubbles on Long-Range Propagation of Nonlinear Pressure Waves in Bubbly Flows

◎金川哲也(筑波大シス情), △谷田部貴大, △鮎貝崇広(筑波大院シス情)

多数の球形気泡を一樣に含む水中において、各気泡と水が初期に等速で流れており、ここを伝わる平面圧力進行波の伝播を理論的に調べる(図1)。気泡の運動として体積振動と並進運動を考え、気泡に働く力として付加質量力と抗力を考える(図1)。波長は気泡径と比較して十分に長く、波の周波数は単一気泡の固有周波数に比べ十分に低い場合を考える(低周波長波)。圧力波の無次元振幅は、有限ではあるが、十分小さいものとする(弱非線形波動)。気泡の生成・消滅・合体・分裂は考慮しない。熱伝導性、気体の粘性、相変化などを無視する。

特異摂動解析の結果、低周波長波が従う KdV-Burgers (KdVB) 方程式が導かれた。並進運動は波の非線形性に寄与し、非線形性を減少させることがわかった。抗力は波の散逸性に寄与し、抗力の考慮は散逸性を増加させることがわかった。抗力による散逸と音響放射による散逸は独立な物理であり、1000 周期に渡る数値解析の結果、散逸性の大きさは、初期気体積分率と初期気泡径に強く依存することが判明した。

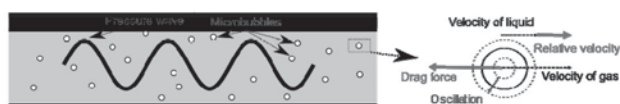


Fig.1: Pressure waves in flowing water containing many microbubbles. Oscillation and translation of bubble are considered.

2-7-8

2-7-8 平面円形に配置した超音波スピーカを用いた空中マニピュレーション

Aerial Manipulation Using Ultrasonic Speakers Arranged in a Circular Plane

☆谷智司(愛工大), 畑中信一(電通大),

△佐藤正典(本多電子), 小塚晃透(愛工大)

- ◆小型超音波スピーカ 36 個を、円形に配置した超音波ユニットを用いて音圧分布を測定した。
- ◆超音波ユニットの内側(1 周目・2 周目)と外側(3 周目)で内側の位相を基準として、外側との位相差を 90° , 180° , 270° で変化させたときの音場中の音圧をマイクロホンで測定した。
- ◆Fig. 1 に測定した超音波の中心軸を含む 2 次元面 ($80 \text{ mm} \times 160 \text{ mm}$) の音圧分布を示す。また、Fig. 2 に内側と外側で位相差を与えたときの音波の中心軸上の音圧を測定した結果を示す。

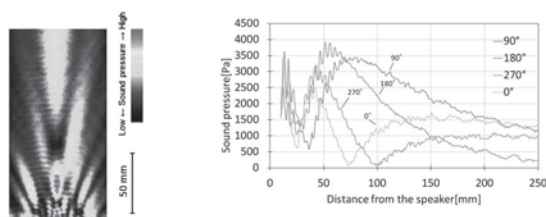


Fig.1 Sound pressure distribution on 2D plane including the sound beam axis measured by the microphone.

Fig.2 Sound pressure distribution on the sound beam axis for the phase difference of 0° , 90° , 180° , and 270° between outermost 18 speakers and inner 18 speakers.

2-7-7

2-7-7 蛍光マイクロバブルの 粘弾性的性質の音圧依存性

Viscoelastic properties of fluorescence microbubbles depending on sound pressure

○吉田憲司(千葉大・CFME), 金児千晶(千葉大・院融合), 章一汀(千葉大・医業), 豊田太郎(東大院・総合文化), 林秀樹, 山口匡(千葉大・CFME)

- ◆超音波・近赤外蛍光デュアルイメージング用の造影剤として開発する蛍光マイクロバブルの膜物質は複数種のリン脂質(卵黄レシチン)から構成される。
- ◆卵黄レシチンバブル懸濁液と DSPC バブル懸濁液(比較対照)の減衰特性とその音圧依存性を評価し、リン脂質膜の膜張粘弾性定数(E^S と κ^S)を推定した。
- ◆両者のバブルにおいて減衰特性には音圧依存性があり(Fig.1 参照)、粘弾性的性質の音圧に対する依存性が示唆された。
- ◆低音圧条件の減衰特性より推定した卵黄レシチンバブルの E^S および κ^S は DSPC バブルと比較して 1/5 および 1/10 程度の値であった。
- ◆卵黄レシチン膜は液状もしくはゲル状であると想定され、バブルの安定性(寿命)の面において特性改善が必要であることが示唆された。

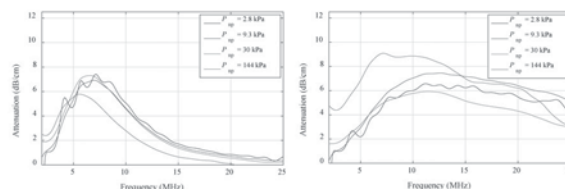


Fig.1: Frequency-dependent attenuation in case of (a) Egg-Yolk Lecithin bubble and (b) DSPC bubble at four conditions for sound pressures.

2-7-9

2-7-9 強力空中超音波励起による 平板内 Lamb 波の伝搬シミュレーション

Numerical simulation for Lamb wave propagation by aerial ultrasonic irradiation

○山田健太, 大隅歩, 伊藤洋一(日大・理工)

- ◆強力空中超音波による弾性波源走査法を利用した材料内部の欠陥を検出する方法について基礎検討を行っている。
- ◆本報告では、提案手法の Lamb 波伝搬について、有限要素法による数値シミュレーションと実測の比較検証を行った。
- ◆その結果、解析定数の設定等の誤差により、高周波領域において減肉部付近の Lamb 波伝搬特性に若干の違いは見られるものの、シミュレーションと実測による Lamb 波伝搬画像の傾向は概ね一致した。

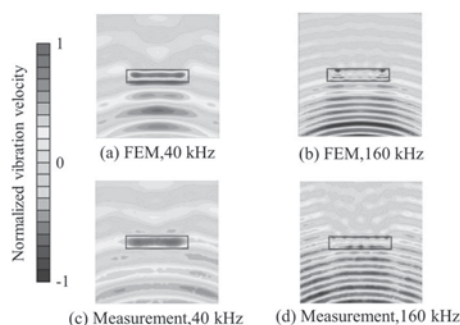


Fig.1: Propagation of Lamb wave at same time

2-7-10

2-7-10 汎用空中超音波センサ素子による
近距離センシングの試み

Short range sensing with commercial airborne ultrasonic transducer

○中村健太郎(東工大・未来研)

- ◆40 kHz 前後で動作する空中超音波トランスデューサが古くから市販され、ロボット、ドローンなどをはじめ、0.1~10 m の距離センシングに広く使われてきた。
- ◆開放型ではQ 値が30~40 であり、単一素子のパルスエコー法による場合、その過渡応答から、0.1 m 以内の近距離測定が困難である。
- ◆本報告では、0~10 mm の極めて近距離のセンシング利用のために、トランスデューサの電気特性に着目した手法を検討した。
- ◆空中超音波トランスデューサは空気との結合がよく、近傍の物体の影響が電気端子から見た共振周波数、Q 値などに現れる。
- ◆Fig. 1 のように、反射面からの距離と、電気特性の関係を調べた。
- ◆例えば、共振周波数 f_r と反共振周波数 f_a は半波長4.2 mm ごとに変化する (Fig. 2)。この変化は、距離が1 波長程度までは顕著である。

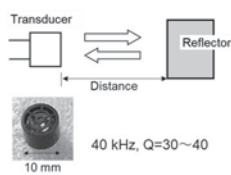


Fig. 1 Airborne US transducer.

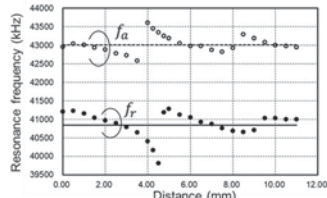


Fig. 2 Resonance frequencies vs. distance.

2-7-12

2-7-12 高分子系音響整合材の付加による
圧電高分子振動子の挙動について

Behavior on the P(VDF/TrFE) piezoelectric element with using an acoustic matching layer of polymer

○高橋真幸(山形大学)

- ◆本研究では、観測対象物の高分解能画像化を目的としている。
- ◆試作したトランスデューサの圧電材料は、圧電高分子材料 P(VDF/TrFE)(75/25mol%)を使用した。トランスデューサの動作周波数は、2MHz である。P(VDF/TrFE)トランスデューサの空気中での高効率動作を実現するため、P(VDF/TrFE)圧電振動子の表面に、同じく高分子系音響整合層(ポリエステル)を付加し、その挙動を実測値と計算値から評価した。
- ◆Figure.1 に、厚み約 200 μ m の P(VDF/TrFE)圧電振動子と、この P(VDF/TrFE)圧電振動子上に、ポリエステル・フィルム(接着層を含む厚み 95 μ m)を張り付けた構造(概略)図を示した。
- ◆結果：この P(VDF/TrFE)圧電振動子の挿入損失は、2MHz で約 60dB であることが分かった。

Figure 1.

Schematic configuration of the P(VDF/TrFE) piezoelectric element with an acoustic matching layer (polyester film).

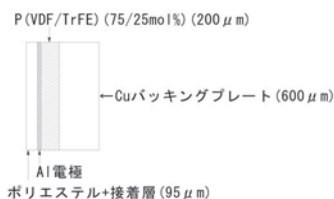


Figure 1

2-7-11

2-7-11 堅牢ハイドロホンの前面板形状と
受波感度の指向性の関係

Relationship between Front Plate Shape of Tough Hydrophone and Directivity of Receiving Sensitivity

○貝瀬不二丸, 小林篤史(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, 竹内真一(桐蔭横浜大)

- ◆当研究室では、高強度超音波に 70 時間曝露しても受波感度が低下しない堅牢性を有した 2 種類のハイドロホンを開発した。
- ◆今回の研究では、2 種類の堅牢ハイドロホンの受波感度の指向性の測定を行い、堅牢ハイドロホンの前面板形状と指向性の関係について検討した。
- ◆Fig. 1 に堅牢ハイドロホンの受波感度の指向性を示す。円柱型堅牢ハイドロホンの受波感度の指向性は最大電圧の方向から 1 度までが -6 dB の範囲であった。円錐型堅牢ハイドロホンの受波感度の指向性は最大電圧の方向から 6 度までが -6 dB の範囲であった。円錐型前面板は先端が小さく後方に向かって徐々に大きくなっており、側面からの信号を受け取りやすいため、受波感度の指向性が低いと考えられる。

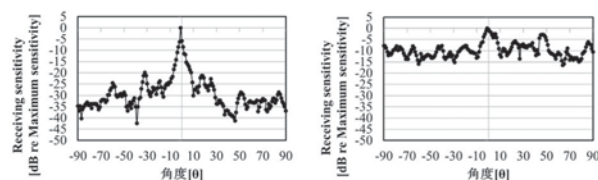


Fig. 1 堅牢ハイドロホンの受波感度の指向性

(左：円柱型堅牢ハイドロホン 右：円錐型堅牢ハイドロホン)

2-7-13

2-7-13 MUT 型音響メタマテリアルによる
音響インピーダンス整合の実験的検証

Experimental Study of Acoustic Impedance Matching by MUT-type Acoustic Metamaterial

○田中宏樹, △町田俊太郎, △南利光彦 (日立製作所)

筆者が提案する MUT (Micromachined Ultrasound Transducer) を利用した音響メタマテリアルによる新たな音響インピーダンス整合器の有効性を実験的に検証した。Fig. 1 に設計構造と試作デバイスを示す。Fig. 2 に水中音響評価により得られた試作デバイスを透過する音波の割合 (透過率、1=100%透過) を示す (Test device Exp.)。MUT 型音響メタマテリアルを含まないシリコンバルク (Si bulk Exp) に対し、有意に高い透過率を示し、提案構造の有効性が示された。

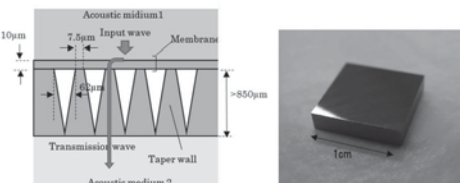


Fig. 1 Proposed design and test piece of MUT-type Acoustic Metamaterial

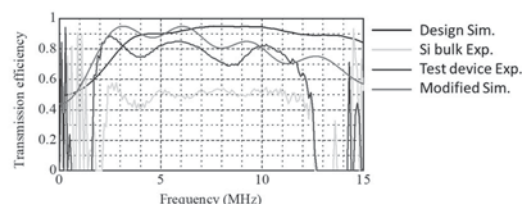


Fig. 2 Transmission efficiency

2-8-1

2-8-1 BLSTM を用いた音素-調音変換

Phoneme-to-Articulatory Mapping Using BLSTM

☆飯山智晴, 大村英史, 桂田浩一(東京理科大)

音声が生体内で共鳴する際には、舌や歯といった調音器官が動き、声道形状の調整が行われる。この調音器官の動きを調音運動という。調音運動のデータ(調音データ)を用いることで、生成される母音の制御などが可能となる。先行研究では深層ニューラルネットワーク(以下DNN)を用いることで音素-調音変換を実現していたが、全結合層のみからDNNを構成していたため、調音データの時系列上での依存関係はモデル化されていなかった。そこで本研究では、long short-term memory(LSTM)とbidirectional LSTM(BLSTM)から構成される再帰型ネットワークを用いることを検討する。これらのネットワークは時系列の依存関係を学習できるため、より高精度な調音データの生成が実現できると考えられる。実験の結果、表1に示すように提案手法が先行手法と比べて高精度な調音データの生成を実現した。このことから、音素-調音変換におけるBLSTMの有効性が確認できた。

Table 1 Experimental results

	RMSE
先行手法(全結合DNN)	0.675
提案手法(BLSTM)	0.651

2-8-3

2-8-3 鼻音生成時の伝達関数に生じる零点の生成要因の検討

Examination of causal factors which generate spectral dips on transfer functions of nasals

☆杉浦唯, 竹本浩典(千葉工大), 北村達也(甲南大), △鴻信義(慈恵医大)

- ◆ 声門から鼻孔に向かう経路に対して口腔、喉頭蓋谷、梨状窩、副鼻腔は鼻音のスペクトルにディップを生成する分岐管となる。
- ◆ 本研究では、鼻音発声時の声道形状をCTで撮像したデータと計算機シミュレーションを用いて、口腔(OC)、左右喉頭蓋谷(EVL, EVR)、左右梨状窩(PFL, PFR)がそれぞれ単独で生成するディップ周波数を求めた(Fig. 1 ▲)。
- ◆ ディップ生成時の音圧分布を検討した結果、鼻音では左右前口蓋弓裏面の中咽頭腔が分岐管として作用することが新たに判明した(Fig. 2○)。
- ◆ これら7つの分岐管は連成振動し、口腔を除いて各分岐管が単独で生成するディップとは大きく異なる周波数にディップが生成されることが明らかになった。
- ◆ 鼻音のスペクトルから口腔以外の6つの分岐管に由来するディップを特定することは困難であると推測される。

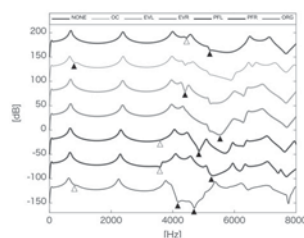


Fig. 1: Transfer functions of seven models



Fig. 2: Front left view of the vocal tract model from the glottis to the lower nasopharynx

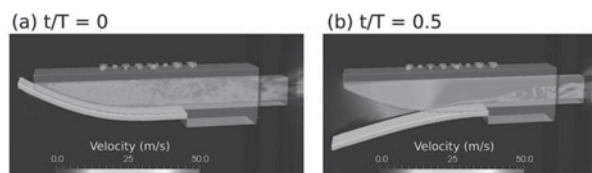
2-8-2

2-8-2 リード式人工声帯の空力音響シミュレーション

Aeroacoustic simulation on reed-type artificial vocal folds

○吉永司(豊橋技科大), △Rafia Inaam(豊橋技科大), 荒井隆行(上智大), △横山博史(豊橋技科大), △飯田明由(豊橋技科大)

- ◆ 声帯は肺からの気流による自励振動により音が発生することが知られ、その音の特性について古くから研究の対象となってきた。また、声帯の音発生を模擬するリード式人工声帯を用いて、声道の特性が調べられてきた。
- ◆ 本研究ではリード式人工声帯に対して、リードの振動、気流と音発生の連成シミュレーションを行うことにより、自励振動による音の発生メカニズムを調査する。
- ◆ リードの振動は1次元梁の運動方程式を計算し、埋め込み境界法により気流内でのリードの振動を表現した。
- ◆ シミュレーションにより予測した発生音は実験と良好に一致し、十分な精度でリードの振動と気流の乱れ、音の伝播を解析できていることが確認できた。

Fig.1: Velocity magnitude in the artificial vocal fold for (a) $t/T = 0$, and (b) $t/T = 0.5$.

2-8-4

2-8-4 鼻腔・副鼻腔模型の造形精度の調査

Verification of modeling accuracy of 3D-printed nasal and paranasal model

○北村達也(甲南大) 杉浦唯 竹本浩典(千葉工大) 鴻信義(慈恵医大)

- ◆ 背景: 3Dプリンタの普及によって、鼻腔・副鼻腔を含む声道の模型を作成しその音響特性を計測する研究が増えている。しかし、模型の造形精度はほとんど検証されていない。
- ◆ 目的: 産業用CT装置を用いて模型の造形精度を検証する。
- ◆ 方法: 成人男性1名の/m/発声時の頭部領域をCTにて撮像し、声道領域を抽出した(杉浦ら, 2020.3)。それに基づき民間の3Dプリントサービスにて模型を一体成型した。模型の素材はナイロンである。その模型を産業用CT装置にて撮像し、原データと比較した。
- ◆ 結果:
 - 模型の造形に使用した3Dプリンタの精度(± 0.30 mm)の範囲の誤差が模型全体に観察された。
 - 一部では上記の精度を超える誤差が見られた。

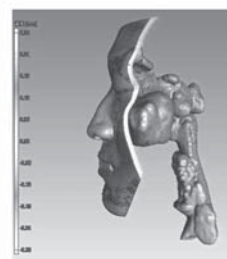


Fig.1: Color map for deviation between the original and measured data.

2-8-5

2-8-5 rtMRI 動画から抽出した発話器官の輪郭データに基づく平均声道の生成と分析

Analysis of mean vocal tract shape based on edge lines extracted from rtMRI movies

☆後藤翼, 天野沢海, 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)

- ◆日本語調音声学の精緻化のため, 18名の被験者の発話運動を rtMRI 動画で記録してデータベースを構築した。
- ◆本研究では, 全被験者の安静呼吸時の声道形状の輪郭を抽出して前鼻棘と口蓋平面が一致するように座標変換した後, 平均声道を計算して主成分分析とクラスター分析を行った。
- ◆平均声道に対して男性平均声道は前下方へ拡大し, 女性平均声道は後上方に縮小していた (Fig. 1)。この変形は第一主成分 (Fig. 2a) と一致した。
- ◆クラスター分析では男性型と女性型に大きく分類された (Fig. 2b)。この分類は, 第一主成分を反映していると思われる。

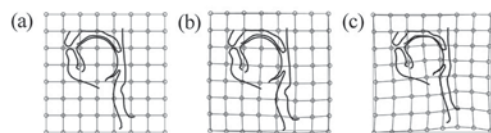


Fig.1: 平均声道 (a), 男性平均声道 (b), 女性平均声道 (c)



Fig.2: 第一主成分 (寄与率0.40) による変形 (a) (太黒線: 平均声道, 細黒線: 平均声道+1標準偏差, 太灰線: 平均声道-1標準偏差), 声道形状のクラスター分析 (b)

2-8-7

2-8-7 若年の補聴器・人工内耳装用者における語音検査と自身の発話明瞭性の関係

The relationship between intelligibility of speech production and perception in young hearing aid/cochlear implant users

○安啓一, △種子田尚人, △石井悠貴 (筑波技術大)

- ◆補聴器および人工内耳装用者の聴取明瞭性と発話明瞭性の関係を検討した。平均聴力の測定と言葉の聴き取り検査, そして健聴者による録音した聴覚障害者の音声の評価を行った。
- ◆補聴器装用者の聴き取りは聴力損失が大きくても聴き取りが良好な場合が多く, 人工内耳装用者の聴き取りは装用時期が早く, 装用年数も長いほど聴き取りの明瞭性は高かった。

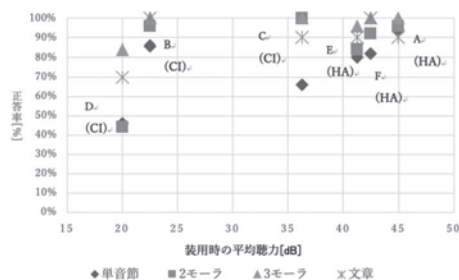


Fig.1: 装用時の平均聴力レベルと正答率。縦軸は正答率, 横軸は装用時の平均聴力を示す。図中のCIは人工内耳装用者, HAは補聴器装用者を示す

2-8-6

Sounds of Englishes:
English pronunciation database

ONoriko Nakanishi (Kobe Gakuin University)

This is a preliminary report of an ongoing project that seeks to build a database for elucidating the trends of accents of World Englishes.

Discussed in the presentation will be:

- ◆Speakers' background
 - List of countries, distribution of gender and generation
- ◆Selection of words and sentences
 - LPD3 inventory, lexical sets, and sentences
- ◆How the data is gathered
 - Online survey form, and calls for participants
- ◆How the data is analyzed
 - Online analysis form, and labeling criteria
- ◆How the data can be used
 - For pedagogical and research purposes



Fig.1: Database "Sounds of Englishes"

2-8-8

2-8-8 時間分解精度の操作がモザイク音声の明瞭度に与える影響: 日本語と中国語の比較

The effect of degraded temporal resolution on intelligibility of mosaic speech: A comparison between Japanese and Mandarin Chinese

☆江口 輝, 上田 和夫, レメイン ジェラード B., 中島 祥好 (九州大)

- ◆音声をスペクトログラム上で一定の時間幅, 周波数幅に区切り, 各区画のパワー平均値を用いて白色雑音で駆動した音声をモザイク音声と呼ぶ。
- ◆本研究では, 日本語および中国語モザイク音声の明瞭度を比較した。
- ◆両言語によるモザイク音声の明瞭度は区間長により決定づけられ, 言語による影響は小さかった。
- ◆両言語において, モザイク音声の方が, 局部時間反転音声よりも区間長の増加に伴う明瞭度の減少が緩やかであった。

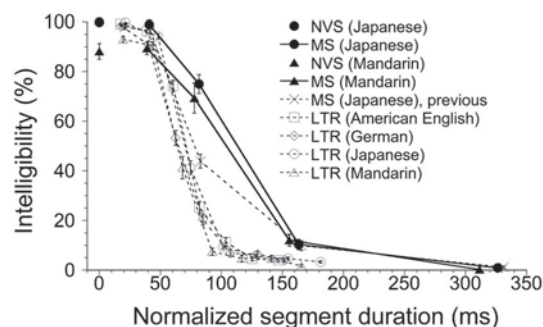


Fig.1: Percentage of mora or syllable intelligibility of noise-vocoded speech (NVS) and mosaic speech (MS) in Japanese (n = 15) and Mandarin Chinese (n = 9) as a function of normalized segment duration. The previous results for MS obtained by Nakajima et al. (2018) and locally time-reversed speech (LTR) by Ueda et al. (2017) are shown together for comparison. Error bars show standard errors.

2-8-9

2-8-9

萌え声の知覚印象

Impressions of Moe voices

☆松富晃之, 平出勇太, 山田真司(金沢工大院)

近年、「クールジャパン」のコンテンツでは「萌え」が重要視されている。この「萌え」関連の市場では、「キャラクター」がコンテンツの大幅を占めている。アニメーションやゲームコンテンツにおいて、キャラクターに声を付与することが主流になり、キャラクターにとって「声」は重要な要素の一つになっている。本研究では、女性声優の「おにいちちゃん」という発話音の印象をSD法によって評定した。因子分析の結果、発話音の印象は「力量性」、「活動性」および「評価性」の3次元の印象空間で表せることが明らかになった。さらに、重回帰分析から得られた標準化係数を基に生成したベクトルを印象空間上に布置したものをFig.1に示す。Fig.1より「評価性」が高い声が「萌え」を感じ好まれることが明らかになった。

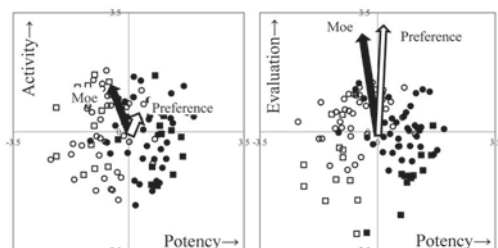


Fig. 1 Impressions of spoken voices. The vectors indicate the degrees of Moe and preference

講演取消

2-8-11

音韻象徴性における調音ダイナミズムの研究-異なる調音点軌跡を示す無意味語による検証-

Articulatory Dynamism of Sound Symbolism
-An Examination on Nonsense Words Having Different Articulatory Trajectories-

○横山安紀子(日本大学)

音韻象徴性(Sound Symbolism、音象徴)はことばにおける音と意味の非恣意的関連性を示すと考えられる言語現象である。このような言語現象はどのようにして生じるのだろうか。その要因や過程についてはいまだ十分に明らかになっていないといえる。発表者は音韻象徴性を生む要因の1つに、ことばを発する際の話者の身体的な調音活動が関与していると考え、発話の際の調音点の移動軌跡の特徴が意味概念の特徴と関係するという仮説を立て、話者の調音活動を「調音ダイナミズム」と呼び、1つのモデルとして提案したいと考える。この「調音ダイナミズム」を検証するため、調音点の移動軌跡を異にする無意味語と異なる空間距離を表す2次元および3次元の視覚刺激を作成し、聴覚刺激をブライム、視覚刺激をターゲットとした知覚実験を行った。その結果、話者がことばを発する際の調音運動過程が、空間を示す意味概念の認知と関係する傾向が見られ、調音ダイナミズムが音韻象徴性を生じさせる要因となる可能性が示唆された。このことはまた、「距離」という概念を人間がことばで表す際に、異なる感覚間の相関関係があることを提起させるものとなった。発表では実験結果とその考察を報告する予定である。

2-9-1

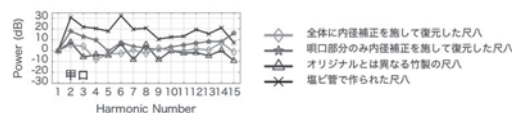
付加製造法を用いて復元した尺八の倍音構造の評価

Harmonic structure of reproduced shakuhachi by additive manufacturing.

☆倉本 有紗(都立産技高専)、高橋 義典、△水野 明哲(工学院大)

△高崎 和之、△山本 昇志(都立産技高専)

- ◆和楽器の一種である尺八は、竹を使用して作られていることから、一度破損・消失してしまうと全く同じ形状の復元することは困難である。
- ◆尺八はある程度硬化な材料であれば材質による音色の変化は見られないことから、これまで尺八をCTスキャンしたデータを元に3Dプリンタを用いて複製を作ること、半永久的に尺八の音色・形状を残すことを目的とした。
- ◆尺八の内側に塗ってあるトノコや石膏の影響で輝度値が高くなることが原因で、複製の内径が狭くなることが確認されてきた。
- ◆3Dモデルを作成する前に、CT画像に内径調節を行うプロセスを導入することで、内径が実物に近い3Dモデルを作成することができた。
- ◆これまでは3Dプリンタの材料の関係で内径調節を全体の1/4のみしか適用できなかったが、本報告では尺八全体に適用し出力を行った。
- ◆復元した尺八の倍音構造について評価を行った。



2-9-2

2-9-2

音響学における三線の研究

Research of Sanshin based on the acoustical evaluation

©西宮康治朗(青山学院大・工)

- ◆沖縄の弦楽器である三線は棹が音色を決定する重要な役割をすると考えられてきた。しかし音響学的には弦楽器の音色は弦・駒・響板が主要因であるとされており、棹の影響は科学的に評価されていない。そこで棹の影響を含めた三線の発音機構を解明する事を目的とする。
- ◆三線には形状の異なる多数の棹の型が存在する。棹の形状差による音色への影響を評価し、曲げ剛性が重要である事が分かった。
- ◆弦と棹の接点である歌口が振動の腹になる棹の固有振動モードの周波数に調弦が一致した際に音色に影響を与える事が分かった。

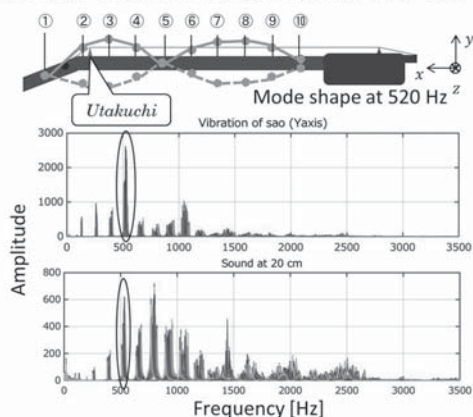


Fig.1: Mode shape of sao at 520 Hz and the frequency spectra of vibration of sao and the sanshin sound.

2-9-4

2-9-4 東南アジア民族音楽の自動採譜に関する研究(邦楽楽器三味線からヴェトナム楽器ダンバウまで)

Research of an automated music scoring system for East Asia folk songs (Japanese traditional instruments Shamisen to Vietnam musical instruments Dan Bau)

©小坂谷壽一(八戸工大)

- ◆古来より邦楽楽器の伝承法は、口伝、又はネウマと言われる記号方式が主であった為、正確な伝承や後継者の育成が難しかった。これらの問題を解決する為、三味線を弾けば自動的に譜面化する自動採譜装置(Fig.1)を開発し、高精度の西洋譜(Fig.2)と三味線譜(Fig.3)を提供する。
- ◆採譜アルゴリズムには、AI手法の1つである“マルチ・エージェント”方式を採用し、実施例として演奏に最も難曲と言われている「津軽じょんがら節」の自動採譜の様を、デモ形式で披露する。
- ◆今後の応用展開として、日本の伝統音楽の譜面化~東南アジアの譜面の無いフレットレス民族楽器の譜面化に取り組む予定である。

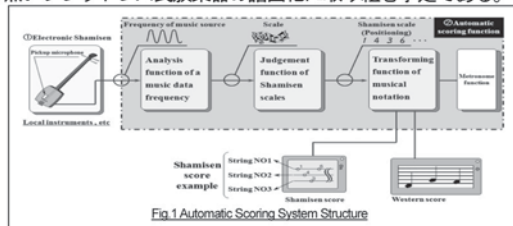


Fig.1 Automatic Scoring System Structure

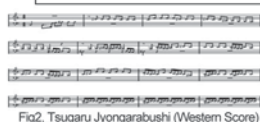


Fig.2 Tsuruga Jyongarabushi (Western Score)

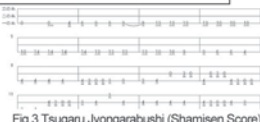


Fig.3 Tsuruga Jyongarabushi (Shamisen Score)

2-9-3

2-9-3

音響学における箏・三味線の研究

Study on Acoustics of Koto and Shamisen

©安藤珠希(箏曲)

- ◆「芸は盗め」と言われてきた伝統音楽にも、以前から科学的な研究も存在する。本稿では箏と三味線に焦点を当て、音響学的にどのような研究が行われてきたのか、主な文献を紹介しながら内容別に概観した。
- ◆その結果、取り上げた1931年~2019年までの36件の文献(演奏動作解析・学習支援・歌・自動採譜・旋律の分析等は除く)の中では、一時的な空白期間はあるものの、箏・三味線ともに定期的に研究が行われていることが分かった。
- ◆内容は、楽器の材料・振動計測、音の周波数分析、音の放射(箏)、箏糸糸、箏柱・三味線皮の代替材料、音律など幅広い内容が扱われている。
- ◆箏は、特に近年、それぞれの研究内容が単独に行われている傾向が多いが、三味線は、音と振動解析が一連の研究として行われていることが多い。
- ◆特に三味線のサワリは、西洋楽器と比べて独特な機構・音色のためか、合成シミュレーションも含めて数多く研究対象とされている。
- ◆楽器・付属品・演奏者・演奏曲・録音状態などの条件が違えば結果が違ったり、演奏者は無意識のうちに求める音を出そうとすることもあつたため、個人差・個体差を考慮しながら共通点を抽出していくことが課題である。
- ◆伝統の維持のためにも研究者・演奏者・専門の技術者の三者が協力し、一連の楽器研究を進めていくことが大切であるとする。

2-9-5

2-9-5 ヘッドの不均一性を特徴とする膜鳴楽器とその物理モデル化・数値解析手法

Physical modeling and numerical analysis of a membranophone characterized by nonuniformity of the head

©鮫島俊哉(九大・芸工)

- ◆Fig.1に示すインドの伝統打楽器である「タブラ」は、ヘッドの材料特性をあえて不均一にすることで、ピッチ感のある独特な音色を表現している。
- ◆Fig.2に示す能楽の「小鼓」では、キャリア・ヘッドの裏側に鹿皮(調子革)が貼られている。また、ヘッドの張力を与える紐(調緒)を手になぐって演奏するが、そのことでヘッドの張力に不均一性が生じている可能性がある。
- ◆膜鳴・民族楽器においては、「不均一性」がその音色を特徴づけるキーポイントになっている。そのような、不均一性を有する膜鳴楽器の物理モデル化と数値解析手法を構築し、音響振動特性を明らかにしたい。

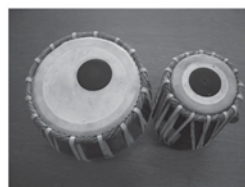


Fig.1: Popular Indian musical drum tabla.



Fig.2: Japanese traditional musical drum kotsuzumi.

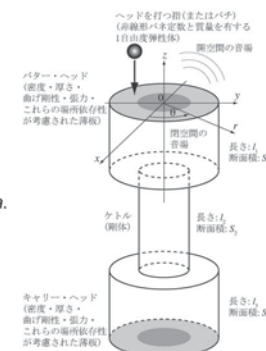


Fig.3: Physical model of a musical drum.

2-9-6

2-9-6 フルートのジェット変動と発生音に関する空力音直接計算

Aeroacoustic Simulation on Jet Behavior and Acoustic Radiation of the Flute

☆小野木君枝, △横山博史, △飯田明由 (豊橋技科大)

人の実演奏状態を模擬した計算モデルを用いて、3次元空力音直接計算を行い、得られた流れ場と音を実験と比較して妥当性を検証した。

- ◆計算で、実験と同等の基本周波数での音の放射を再現できた。
- ◆流れ場を可視化した結果(図1)、共鳴管内が低圧時($t/T=0$)、ジェットはエッジの外側を向き、共鳴管内が高圧時($t/T=2/4$)、ジェットはエッジの内側を向いていた。このジェットと圧力の位相関係は、基本周波数での音の発振に良好な関係であり、基音放射のために妥当な流れ場が再現されていることが確認できた。
- ◆計算結果の流れ場の実効的偏心値と、計算結果の2次、3次倍音の優位性には整合性があり、流れ場と音の整合性を確認できた。

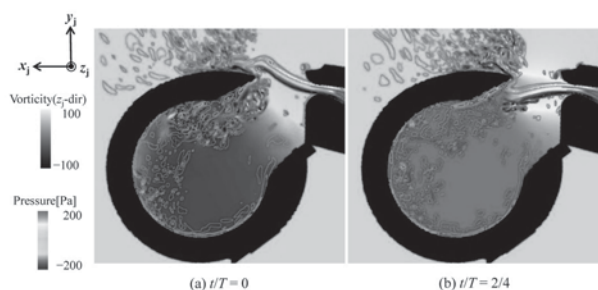


Fig.1 Temporal variation of vorticity and pressure field.

2-9-8

2-9-8 異なる断面積プロファイルを持つ金管楽器における吹鳴時の抵抗感および反射関数の比較

Comparison of Blowing Resistances and Reflection Functions among Brass Instruments with Different Cross-Sectional Profiles

○佐藤 周平, 若槻 尚斗, 水谷 孝一, 前田 祐佳 (筑波大)

- ◆本研究では、金管楽器の反射関数と抵抗感の関係性に着目し、金管楽器と同様の形状を有する気柱モデルを作成し、有限要素法によるシミュレーションによる周波数応答解析を行い、反射関数を算出した。
- ◆シミュレーションでは気柱モデルを3つ作成し(Fig. 1)、それぞれにおける音響インピーダンスを計算し、反射関数を算出した(Fig. 2)。
- ◆その結果、管体のプロファイルに応じて反射関数におけるパルスの形状と振幅が変化する様子が確認された。
- ◆今後はこの結果を用いて反射関数と唇の物理モデルを用いて自動振動のシミュレーションを行い、抵抗感との関係を考察する予定である。



Fig.1: Shapes of air column models

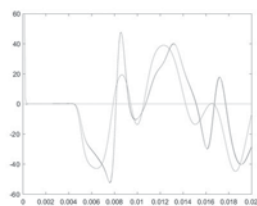


Fig.2: Reflection functions

2-9-7

2-9-7 組み立て式ホイッスルの実用化に伴う音量の向上検討

Study for improving the volume of the assembled whistle

○黒沢良夫 (帝京大), △小野真 (レイトロロジー)

- ◆災害時に救助を呼ぶため、携帯に便利な組み立て式のホイッスルを開発した。名刺サイズの平面素材(通常の名刺に用いるケント紙やコート紙)を組み立てて作成する。本内容は既に特許化した。十分な音量(音圧レベル)が出なかった。救助者に聞こえやすいよう、内部構造や吹き方を工夫して音圧レベルやピーク周波数を改善した。
- ◆その結果、40cmの距離で98.4dBの音圧レベル(O.A.)を実現した。さらに、孔から出口までの距離とピーク周波数の実験関係式を導出できた。また、息の通り道の隙間は息の量(吹きやすさ)に関連し、音量とのバランスの最適値を見つけた。吹き方は特にコツはなく初心者でも十分大きな音が出せるが、側面の穴の塞ぎ方により、音圧レベルを向上できることを発見した。



Fig.1: Assembled whistle

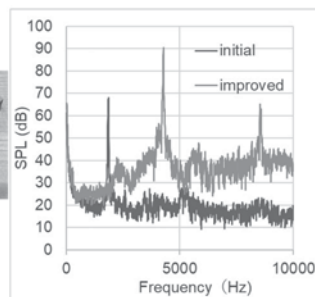


Fig.2: Comparison of the sound pressure level

2-9-9

2-9-9 非楽音の打音推定システムの開発

Development of Identification System on Non-musical Percussive Sound

☆對馬朗, △伊藤和寿 (芝浦工業大学)

- ・本研究では、非楽音である打楽器(ドラムセット)の自動探譜を行うシステムの開発を目的とする。
- ・開発するシステムの最終的な目的は、打音ファイルを入力とし、打楽器を叩いたタイミングを推定し、七種類の打楽器の推定と、三種類の皮物楽器は演奏時の奏法の推定を行う。これらの結果を踏まえて、以下のFig. 1に示すようなドラムセットの譜面を出力することである。



Fig.1: Example of drum set musical score.

- ・提案するプログラムは、事前に取得する参照データとFFT解析後のパワースペクトル密度を比較することで楽器を推定する手法である。この手法は連続打音や和音の解析に対して容易に拡張が可能である。
- ・全ての楽器・奏法において10データ用意し、解析・推定を行った。
- ・単音における楽器名・奏法の推定精度の平均値は98%であった。

2-9-10

2-9-10 明珍火箸の打撃音の計測と解析

Measurement and analysis of striking sound of Myochin fire chopsticks

☆島崎湧, 小長谷悠, 及川靖広, 菅野由弘(早大理工)

- ◆ 明珍火箸とは、平安時代から続く甲冑師の家系である明珍家が、明治時代に入って火箸作りに着手し、昭和40年代に現当主52代明珍宗理によって楽器として創り出された火箸である。
- ◆ その特徴である澄んだ音色と長い余韻は現代の作曲家に注目されている一方で、工学的な分析はまだ十分でない。
- ◆ 打撃音のスペクトログラムのそれぞれのピークを拡大してみると、近接する位置にピークが2本ずつ存在することが確認できた (Fig.1)。
- ◆ 火箸の振動のシミュレーションとして、有限要素法による固有値解析を行った。さらに、得られた固有振動数を用いて火箸の打撃音を生成した。実測結果とシミュレーション結果の比較から、近接した2つの固有振動数によって生じるうなりが火箸の音色の形成に寄与していることが確認できた。

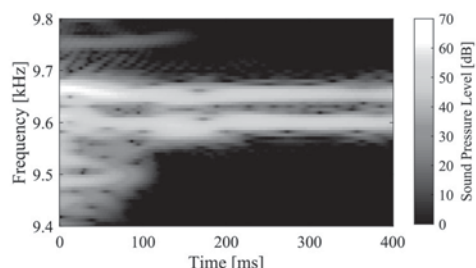


Fig.1: Spectrogram of striking sound of Myochin fire chopsticks

2-10-1

2-10-1 保育施設周辺の音環境に関する意識調査

Questionnaire survey on sonic environment around a childcare facility

◎片岡寛子, △樋下真路, 高田正幸(九州大・芸工)

- ◆ 本研究では、住民の居住環境に対する満足度や聞こえてくる音などから特徴づけられる地域性や音に対する個人の感受性を把握したうえで、保育施設から生じている音の捉え方とどのように関係するかを検討するため、アンケート調査を実施した。
- ◆ 調査の結果 回答者の過半数が住環境の静けさに満足しており調査対象施設が立地する地域は比較的静穏な環境と言える。
- ◆ その中で聞こえてくる音について、悩まされている音として最も多く挙げられたのは飛行機の音であった。飛行場が比較的近くにあり、着陸時の飛行経路になっているという地域の特徴が表れている。
- ◆ 保育施設新設の賛否に関する質問では、ネガティブな回答が全体の数%程度であり、先行研究で保育施設の近隣住民に対して行った調査と比較して少ないことが示された。
- ◆ 個人の特性として、回答者の騒音感受性を評価し、保育施設新設に関する意見との関連を検討したところ、保育施設新設に対して否定的な意見を持つ回答者は騒音感受性が高い傾向にあることが明らかになった。

2-9-11

2-9-11 非負値行列因子分解を利用した複数楽器音で構成される楽曲の分離再合成の高精度化

Improvement of accuracy of separation and resynthesis of music composed of multiple instruments by using non-negative matrix factor decomposition

☆前田梨江, 星野博之(愛知工業大学)

- ◆ **目的** 一度混ざってしまった多重音を単一音として扱う上で重要な音源分離の問題に注目し非負値行列因子分解を使用して3曲中に用いられている楽器音が音源分離によりどのように音質劣化するかを調べた。
- ◆ **方法と結果** 3曲の音楽データの約11秒間ずつを切り取り、非負値行列因子分解によって分割し再合成する。再合成音について楽器音ごとの原音に対する明瞭度劣化と発生するノイズを主観評価する実験を行った。Fig.1に使用曲①、フレーム長512*6の場合の結果を示す。また劣化の度合いとスペクトル差分値との相関係数がどの帯域に限定すると大きくなるか調べ、分離再合成による音の減衰帯域とノイズの発生帯域を明らかにした。

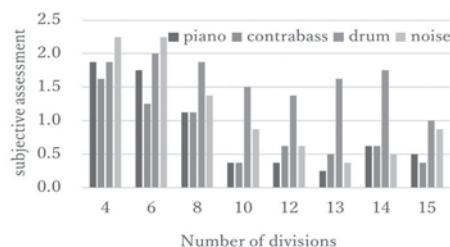


Fig.1 Results of evaluation of intelligibility degradation by number of divisions (Music 1, Frame Length 512*6)

2-10-2

2-10-2 音環境からみた大学図書館における学生の居場所に関する検討

図書館を題材とした「自分事」化プロジェクトの実践

Study on student's "lbasho" in university library from soundscape perspective: A case report on project "struggle as his/her own issues" in KAIT library.

○小橋侑生, 寺嶋将隆(神奈川工科大),

三浦貴大(産総研), 須藤康裕, 松本一教, 上田麻理(神奈川工科大)

神奈川工科大学では、図書館を題材として2019年度より自分事化プロジェクトを実践している。その一環として、五感を活用した様々な視点・専門分野からプロジェクトを進めている。聴覚の視点からも本プロジェクトに寄与すべく調査を開始した。本研究では、神奈川工科大学附属図書館のサウンドスケープ調査を学生自身が行い、自分事化プロジェクトの第一歩とした。



Figure: Project "struggle as his/her own issues"

2-10-3

2-10-3 残響時間・音源の種類・提示レベルが知識創造活動時の発言しやすさに与える影響

Influence of sounds, presentation level and reverberation time on ease of speaking of knowledge creative activity

☆仁平真由(茨城大院), 矢入幹記, △沖田隆慶(鹿島建設),
辻村壮平(茨城大院)

- ◆天井の高い空間において複数人で知識創造活動を行う状況を想定させ、残響時間、音源の種類、提示レベルを変化させた空間で主観評価実験を実施し、これらの音環境の要因が知識創造活動時の発言しやすさに与える影響を検討した。
- ◆主観評価実験の結果から、大空間における複数人での知識創造活動において残響時間約0.6s~2sの範囲で残響時間の影響は小さいことが示唆された。また、本実験の条件では、暗騒音のみの音環境より音源を付加した音環境が「発言しやすい」印象になることが分かった。知識創造活動時の「発言しやすさ」には「静かさ」と「活気」が関係していることが示され、さらに「リラックス」も関係していることが示唆された。

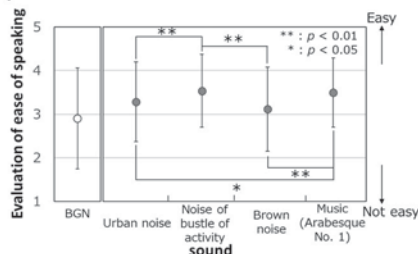


Fig.1: "Evaluation of ease of speaking" for each sound environmental condition

2-10-5

2-10-5 視覚情報の有無が音環境に対する評価構造に与える影響

The influence of visual stimuli on the evaluation structure of sound environment

☆山崎尚斗(茨城大院), 辻村壮平(茨城大院), 朝倉巧(東京理科大)

- ◆音環境の快適及び不快の印象に着目し、視覚刺激が音環境評価に及ぼす影響を捉えるため、聴覚刺激の提示手法が異なる三手法(実環境の提示、音刺激のみ提示、VR及び音刺激の提示)による主観評価実験を実施した(Fig.1, 2)。
- ◆対象環境が快適及び不快な場合を比較すると、快適な環境を対象とした場合の方が音環境評価に及ぼす視覚刺激の影響は大きい。
- ◆「イライラする一落着く」、「激しい一穏やかな」、「うるさい一静かな」の評価は視覚刺激の影響を受けやすい。
- ◆聴覚刺激の提示手法を比較すると、実環境における提示が音環境評価のための潜在的な軸が多い。
- ◆不快な音環境を対象とした場合の方が快適な場合よりも音環境評価のための潜在的な軸が多い。

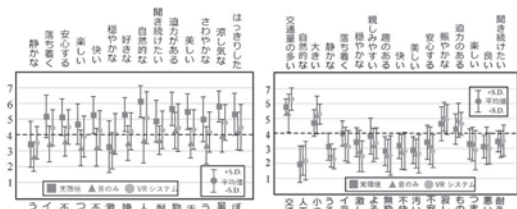


Fig. 1 Results in the condition of a comfortable sound environment

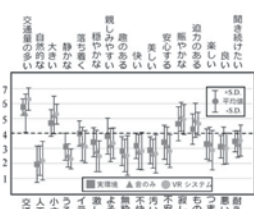


Fig. 2 Results in the condition of an uncomfortable sound environment

2-10-4

2-10-4 リフレッシュしやすさに及ぼす視覚と聴覚の調和感の影響に関する研究

Study on the effect of vision and hearing harmony on the ease of refresh

☆北條寛人(茨城大院), 佐藤浩浩, 三浦太郎, 小林真人,
△科部元浩(飛鳥建設), 辻村壮平(茨城大院)

- ◆業務時間中にリフレッシュする状況を想定させ、提示映像と付加音の種類を変化させた各条件下で主観評価実験を実施し、リフレッシュしやすさの印象に与える視覚と聴覚の複合影響を調和感に着目して検討した。
- ◆主観評価実験の結果から、複合刺激により快適と評価される背景として、単体の刺激が快適な印象であることだけでなく、複合させた刺激の調和感が影響を及ぼすことが示された。さらに快適さの評価には、映像による快適感よりも音による不快感の影響の方が大きいことが示唆された。また、音及び映像の調和感を捉える場合、音と映像のイメージの調和感のみならず、各提示刺激間において調和している刺激群を把握することが重要である可能性がみられた。

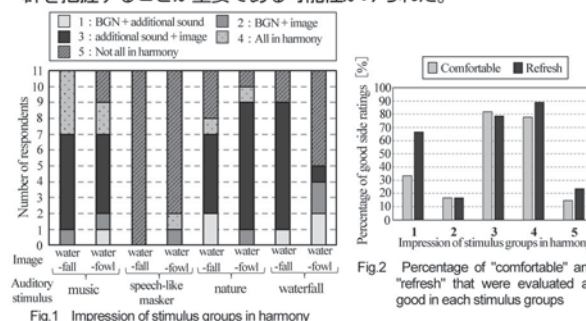


Fig.1 Impression of stimulus groups in harmony

Fig.2 Percentage of "comfortable" and "refresh" that were evaluated as good in each stimulus groups

2-10-6

2-10-6 報知音のデザイン
報知強度と統一感のデザインに向けて

UI sound design: Towards a design method of message sounds that easily notify each strength of message and express one unified impression

和氣早苗, 和田美咲(同志社女子大)

- ◆【目的】通常、一つの機器においても異なる報知強度の報知音が複数利用される。複数の報知音を併用する際、各報知音の「報知強度」の違いを明確に理解でき、一方で全体の「統一感」も感じることが出来る報知音セットのデザインについて検討する。
- ◆【実験概要】報知強度 弱・中・強の3音で1セットを5セット準備し、一対比較評価と絶対評価にて、「報知強度のわかりやすさ」と「統一感」を評価した。音色を統一した実験と異ならせた実験の2回の実験を行った。
- ◆【結果】報知音の強度差をわかりやすくしながら、報知音セットとしての統一性を保つには、繰り返しの発音パターンで統一感を保ちつつ、強度に伴って音価を短くテンポを上げるのがよいことがわかった。なお、強度差をより強めたい場合には音色の違いを用い、統一性を優先したい場合は同一音色で作成するのがよいこともわかった。

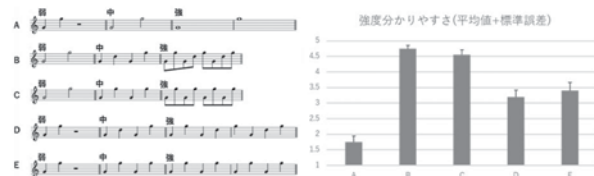


図1: 実験1(音色統一)に用いた報知音セット

図2: 実験1(音色統一)での強度のわかりやすさ

2-10-7

2-10-7 加齢性難聴に伴うポピュラーソングの
印象変化の検討～模擬難聴を用いて～Changes in the impressions of Japanese popular songs
due to age-related hearing loss○寺澤洋子、水野真由美、山本雄也、大中悠生、石川嘉秀(筑波大)
松井淑恵(豊橋技科大)、安啓一(筑波技大)

- ◆高齢難聴を有する音楽リスナーは、どのようにポピュラーソングを聴いているだろうか？
- ◆1970—2000年代の日本のポップスを選曲し、それらの抜粋に模擬難聴処理を施し、オリジナル音源と模擬難聴音源を作成した。それらを若年健聴者25名に聴試してもらい、健聴条件と模擬難聴条件における楽曲の印象変化を評価した。評価には、SD法による尺度評価と、自由記述による印象変化の描写を用いた。
- ◆SD法による分析の結果、模擬難聴処理によって、地味な印象へと遷移する傾向があること、不快度が上がる傾向があること、曲のスピード感が中庸な印象へと変化して行く傾向があることがわかった。尺度評価と同時に、印象に残った楽器を自由記述で回答してもらったところ、低音の楽器が相対的に聞こえやすく、印象に残ることが示された。
- ◆二つの条件の間でどのように印象が変化するかを尋ねたところ、全ての楽曲において模擬難聴によって不快度が増すわけではないことも示された。楽器数がもともと少ない曲、バラード、年代の古い曲などにおいては、模擬難聴条件の方が好まれる場合もあった。

2-10-8

2-10-8 加齢による聴覚特性変化と
食関連音の印象との関係** The effects of age-dependent degradation of auditory characteristics on
sound impression.

☆於勢真紀子、脇田由実(大阪工大)

- ◆若年者と高齢者との食関連音に対する印象強度には差があり、同じ食関連音を聞いても、高齢者は若者に比べおいしいと感じない傾向がある。そこで両者の印象強度の違いが音の時間変化量の違いであると仮定し、様々な食関連音に対して両者の聞こえの音響特徴(パワー、MFCC、スペクトル重心)の時間変化量を比較した。
- ◆その結果、パワーが大きい箇所のMFCC1次の若年者の時間変化度合いは、試聴時の音の大きさにかわらず、若年者の聞く音の方が大きい傾向にあった。音の印象の違いがMFCC1次の変化の大きさの違いに起因する可能性がある。今後は、さらに試聴実験を行い、試聴時の印象と音響特徴量との関係を明らかにしたい。

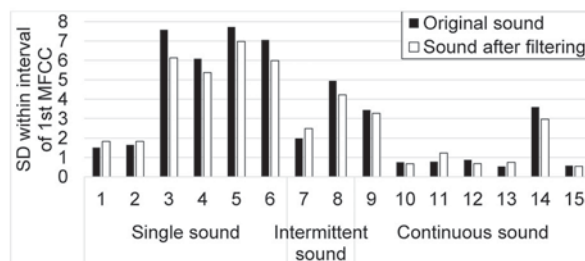


Fig. 1 1st MFCC standard deviation within interval of the highest section of sound pressure power in each sound source

2-10-9

2-10-9 評価グリッド法の理論と実際

Theory and practice of the Evaluation Grid Method,

○讃井純一郎(関東学院大学)

- ◆適切なデザイン目標を設定するためには、利用者の要求や価値観を把握することが不可欠であるが、漫然とヒアリングを行ってもうまくはいかない。本稿では、人々のデザイン評価のメカニズム(≒ニーズ)を構造的に明らかにすることを目的に開発した面接調査手法「評価グリッド法」について、その概要を解説するとともに「音のデザイン」に応用する際に留意すべき課題について論じる。
- ◆評価グリッド法は、認知心理学的理論を背景に、臨床心理学分野において治療を目的に開発された面接調査手法(レポートリーグリッド法)をデザイン評価に適した形に改良発展させたものである。複数の評価対象を比較評価させ、なぜそのように評価したのか理由を尋ねることで、デザイン評価の観点(評価項目)を回答者自身の言葉で効率よく抽出する。さらに得られた個々の評価項目に対し、そのことによって得られるベネフィット、あるいは逆にそのことを実現するために必要と考える具体的な特性をラダーリングと呼ばれる補完質問によって明らかにすることで、評価判断の全体像(評価構造)を階層的ネットワーク図として表現することを可能とした。
- ◆評価グリッド法が「音のデザイン」においても活用されることを期待するが、その際、対象音に対する評価構造を有している人々を回答者とするのが不可欠であること、また時間の経過とともに消えてしまう「音」という評価対象を、いかにして比較評価が可能な形で提示するかといった留意すべき課題についても述べる。

2-10-10

2-10-10 (招待講演) 被験者自身の言葉で評価語
とした音事象の評価構造の抽出Evaluation structure of environmental sounds: Experiment of subjective
evaluation using subjects' own terms

○川井敬二(熊本大・院先端科学)

- ◆サウンドスケープの記述に関する ISO/TS 12913-2(2018)では、pleasantness, eventfulness といった主観評価語群が例示されている。
- ◆サウンドスケープという語が示唆するように環境音から感じる印象評価は多様である。それを把握するために、環境心理評価で多用されてきたSD法は有用と思われるが、どの形容詞対を用いるのが妥当かを議論する必要がある。たとえば辞書などの語彙集や先行研究から網羅的に選んだ、といった研究も多いが、「かたいーやわらかい」「澄んだーにごった」など音色研究で使用された形容詞対はサウンドスケープ評価に適しているのか？
- ◆実際に、人々が音事象からうける印象評価がどのようなものであるか、個人毎の印象表現語を抽出し、個別評定尺度による主観評価実験を考案し、実施した。実験は、被験者による音事象のグループ化、音事象グループへの印象表現語の付与、その印象表現語を用いた主観評価の3ステップから成る。評定結果を主成分分析により分析した。
- ◆得られた主成分は「選好性(いいーわるい、など)」「活動性(にぎやか、など)」「日常性・生活感」の順で、寄与率は32, 13, 7%であった。
- ◆選好性・活動性については、概ね全被験者が相関の高い印象表現語を挙げた。第3主成分以降は個人差が大きくなり、みられた。
- ◆pleasantness, eventfulness という ISO 規格の主要軸と類似した主成分の構成であり、その妥当性を支持する結果といえる。

2-10-11

2-10-11 ホールにおけるフルートの鳴りの評価
—研究対象に応じた評価方法の工夫—

Evaluation for 'projection' of flutes in a concert hall

○小林まおり(JAIST), 上野佳奈子(明治大・理工)

- ◆どのような研究対象でも安定して反応が得られる万能な主観的評価法は存在せず、研究対象に応じて工夫する必要がある。
- ◆その一例として、楽器の遠鳴り・そば鳴りの観点から行ったホールにおけるフルートの鳴りの評価に関する研究を取り上げる。
- ◆この研究では、遠鳴り・そば鳴りという言葉に対応する聴感印象が評価者間で一致することを前提とせず、定義や聴感印象の多様性について把握を試みた。
- ◆具体的には、まず実験参加者の遠鳴り・そば鳴りの定義の多様性について調べた。次に、実際に複数の演奏音を参加者に提示し、遠鳴り・そば鳴りの程度を評価させ、どの音が遠鳴り・そば鳴りと評価されるのか調べた。その際、評価に用いる言葉を分析することで、どのような聴感印象と対応づけられているのか検証した。

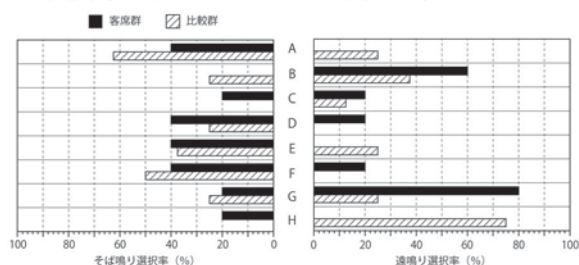


Fig.1: 遠鳴り・そば鳴り上位2位までに選ばれた選択率。左図がそば鳴り、右図が遠鳴りに選ばれた選択率(評価者 □: 比較群, ■: 客席群)

2-10-13

2-10-13 音のデザインへの評価グリッド法の
応用的展開に関する検討An investigation on applied development of
the evaluation grid method to sound design

○辻村壮平(茨城大院)

本稿では、まず評価グリッド法の実施において重要なポイントを確認した(Table1)。さらに、音を扱った研究テーマにおける評価構造の抽出に評価グリッド法を適用するための課題や注意点を整理し、音響工学分野への評価グリッド法の応用可能性について検討した。

- ◆音をエレメントとする場合、回答者が差異を判断できるような聴感的な違いが明確なものをを用いる。
- ◆音に対する評価構造抽出において、概念が場と直接結びつく場合、一対比較法のように、音刺激を二つ組にして順に提示し、総合評価を行う基準となる観点でどちらが良いかを回答者に判断させ、その理由を訊ねる方法が考えられる。
- ◆概念が一定の場に限定されないものである場合、イメージグリッド法のように、音刺激を一つずつ全て提示し、音に関して感じた印象とその理由を思いつくり回答してもらう方法が考えられる。

Table1: Failure factors of the evaluation grid method.

認知構造的側面	テーマが抽象的すぎる
	対象者の意識に評価構造が存在しない
技術的側面	エレメントに偏りがある (種類が相応しくない)
	インタビューアの問題
	記録・まとめ方の問題

2-10-12

2-10-12 評価グリッド法による製品への
潜在価値・要求の導出Derivation of potential value and requirements for products
by evaluation grid method

○山岸和子(ソニー), 関研一(千葉工大)

新製品の開発において、新たな感性価値を導出し、感性に関わる性能を付加する場合、製品の主機能と両立する構造案を見出すことは難しい。本論文では、評価グリッド法とDSMを用いて顧客の潜在要求を導出し、機能・構造との関係を可視化する方法を検討する。提案する方法をキーボードに適用した結果、音に対する要求、機能、構造の関係がDSMクラスターで可視化され、操作音、音質、機能音としてどう性能を付与すればよいかを関係者で検討しやすくなった。また、要求に基づく音に対する新たな機能や、キーボードに対する新たな価値創出とその実現手段が関係付けて導出され、音のデザイン面でも有効性を示すことができた。

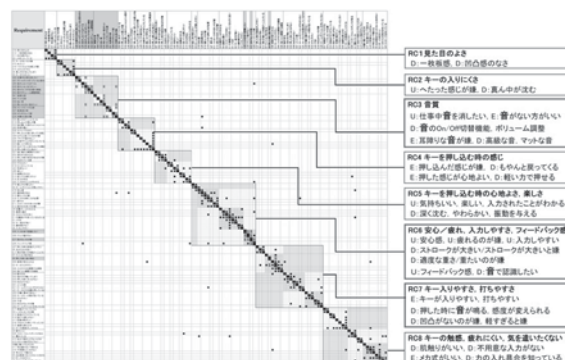


Fig.1: Requirement DSM of keyboard

2-P1-1

2-P1-1 End-to-End 複数言語音声認識モデルに
おける様々なマルチタスク学習の検討Investigation of Various Multi-task Learning for End-to-End
Multi-Language Speech Recognition Model

☆早川友瑛, 西崎博光(山梨大・工), 山本一公(中部大・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)

◆研究目的:

- 複数言語音声認識におけるマルチタスク学習の評価
- ✧ CTCベースのEnd-to-End音声認識システムで効果を確認

◆マルチタスク学習

- CTCモデルと言語識別器
- CTCモデルと話者識別器
- CTCモデルとスタイル^{*}識別器(*読み上げ音声 or 自然発話)

◆実験結果

- DNN-HMM 音声認識システム(ベースライン)と性能を比較
- 複数言語音声認識においてマルチタスク学習の有効性を確認
- ベースラインのDNN-HMM音声認識システムの精度に匹敵

Table 1 PERs [%] in each E2E model with multi-task learning.

言語	DNN-HMM (Baseline)	事前学習 (CTC)	CTC and 言語識別器	CTC and 話者識別器	CTC and スタイル識別器
		(CTC)	GRL ×	GRL ×	GRL ×
チェコ語	11.6	10.0	8.7	8.5	8.7
英語	26.5	34.9	33.7	32.6	31.2
フランス語	13.1	9.2	8.2	8.2	7.9
ドイツ語	19.0	17.7	15.3	15.2	14.9
日本語	14.7	20.3	16.3	16.8	19.7
スペイン語	10.8	12.1	9.7	9.5	9.2
全言語	15.7	18.1	15.7	15.6	16.3

2-P1-2

2-P1-2 音声認識エンジン Julius における DNN 計算のための CUDA アーキテクチャの検討

Examination of CUDA architecture
for DNN-HMM Computation
in Speech Recognition Engine Julius

○鈴木駿, 李晃伸

- ◆音声認識エンジン Julius におけるリアルタイム DNN-HMM 計算の高速化を行った。
- ◆CPU でのマルチスレッド化および SIMD 拡張命令では、平均処理速度(実時間比)が3.19 xRT から0.83 xRT へ改善されたが、メモリバンド幅の上限からそれ以上の改善は困難である。
- ◆CUDA C を用いた GPU での並列計算を実装した結果、CUDA のアーキテクチャを適切に設定することで最終的に約 0.20 xRT と CPU のみの場合と比較して約 4 倍の処理速度の向上が得られた。またブロックの構成によって大きく性能が変化することが示された。

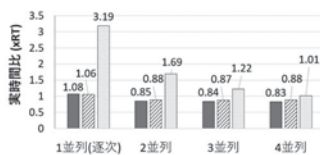


Fig.1: マルチスレッドとSIMD拡張命令の併用による処理速度の比較

Table.1: CUDA C を用いた認識処理の実時間比 (xRT)

X	2	4	8	16	32
2	0.719	0.434	0.378	0.326	0.332
4	0.462	0.301	0.279	0.263	0.266
8	0.339	0.264	0.282	0.280	0.281
16	0.235	0.197	0.196	0.205	0.240
32	0.240	0.239	0.244	0.258	0.276

2-P1-3

2-P1-3 Kaldi ベースの低遅延リアルタイム 音声認識システムの開発と評価

Development and evaluation of Kaldi-based low-latency
real-time speech recognition system

☆レオ チーシャン, 早川友瑛, 西崎博光 (山梨大・工), 北岡教英 (豊橋技科大)

◆研究目的:

- 低遅延リアルタイム音声認識システムの開発と評価
- Google API との認識精度・速度を比較

◆音声認識システム

- Kaldi で学習したモデル+デコーダで構成
- スタンドアロンで動作
- ファイル入力, マイク入力, 仮想オーディオデバイスに対応
- Google と比較して低 RTF と低遅延を実現

◆認識実験

- 音響モデルは CSJ・JNAS・S-JNAS から, 言語モデルは BCCWJ から学習 (辞書サイズは 164k)
- CSJ Eval3 テストセットに対して認識精度と速度を評価
- 4 種類のマイク (認識環境を表現) に対する認識精度を比較

Table1 WERs of the CSJ eval3

	Google	Our ASR
WER [%]	28.16	6.45
RTF [s]	0.324	0.167
Latency [s]		0.436

Table.2 WERs for four types of mic.

	Google	Our ASR
Tabletop mic.	17.65	10.78
MacbookPro Built-in	11.76	9.8
AirPods Pro	14.71	19.61
ReSpeaker Mic	12.75	15.69

2-P1-4

2-P1-4 顔表情と音響情報を用いた対話破綻検出

Dialogue breakdown detection
using facial expressions and acoustic features

☆鈴木海斗, 入部百合絵 (愛泉大), 北岡教英 (豊技大)

- ◆7名の被験者を対象にシステムのふりをした人間がユーザと対話する Wizard-of-Oz 法を用いて収録された雑談対話動画からユーザの顔情報と音響情報を抽出し、それらの特徴を用いて対話破綻を検出
- ◆収集した動画のシステム発話に対して 2 名がアノテーションを実施し、タグが一致した発話のうち、破綻、非破綻それぞれ 54 発話ずつの計 108 発話を対象として破綻/非破綻を識別
- ◆予備実験より明らかとなった、表情変化の現れやすいシステム発話終了後 3 秒間を分析対象とし、顔の特徴点から距離や面積を算出し計 112 次元の顔特徴量を抽出
- ◆音響情報の分析区間はシステム発話終了からユーザ発話終了とし、破綻後に変化の現れやすい 384 次元の音響的特徴量を抽出
- ◆抽出した特徴量に対してラッパー法を用いた特徴量選択により次元削減し、破綻/非破綻の識別を行った結果、顔の特徴量と音響的特徴量を組み合わせた正解率が最も高く 85%を得た (Table.1)

Table.1: Results of Classification

Feature	Accuracy
Face	0.79
Acoustic	0.75
Face + Acoustic	0.85

2-P1-5

2-P1-5 書き起こしのための遠方発話 音声認識技術の検討

Study of distant speech recognition for transcription system

◎佐藤裕明, 萩原愛子, 伊藤均, 三島剛, 河合吉彦, 小森智康 (NHK), 佐藤庄衛 (NHK エンジニアリングシステム), 小川哲司 (早稲田大学)

- ◆NHK は音声認識技術を用いた制作現場の書き起こし支援技術の研究を進めている。今回、記者会見における記者質問など、遠方発話音声の認識精度改善を以下の 4 手法により検討した。

1. ルームシミュレータによる室内インパルス応答(RIR)付加
2. 周波数の高域減衰
3. 白色雑音付加
4. SpecAugment

- ◆各種手法を適用した場合における遠方発話の WER を Table 1 に示す。
- ◆手法の単独適用では RIR 付加に最も改善効果がみられ、各種手法の組み合わせによる適用で更なる改善効果がみられた。

Table 1 WER for distant speech recognition using each method

Add RIR	Decay High Frequency	Add Noise	Spec Augment	WER [%]
-	-	-	-	98.6
✓	-	-	-	88.8
-	✓	-	-	98.1
-	-	✓	-	98.3
-	-	-	✓	98.4
✓	✓	-	-	86.4
✓	-	✓	-	88.5
✓	-	-	✓	88.1
✓	✓	-	✓	85.6
✓	✓	✓	-	84.6
✓	✓	✓	✓	84.0

2-P1-6

2-P1-6 Transformer 音声認識システムの
進化的最適化

Evolutionary Development of Transformer Speech Recognition System

☆日野健人, 篠崎隆宏(東工大)

- ◆進化最適化法により Transformer 音声認識システムを最適化
 - Transformer 音声認識システムとして ESPnet の ami レシピを使用
- ◆CMA-ES による最適化と DI-ES による最適化を認識エラー率とモデルサイズの点から比較
 - DI-ES における文化伝承法として知識蒸留を使用
- ◆どちらの進化戦略においても ami レシピによる初期個体よりも認識エラー率の小さいモデルを実現
- ◆DI-ES により ami レシピによる初期個体よりもモデルサイズと認識エラー率の同時削減を実現

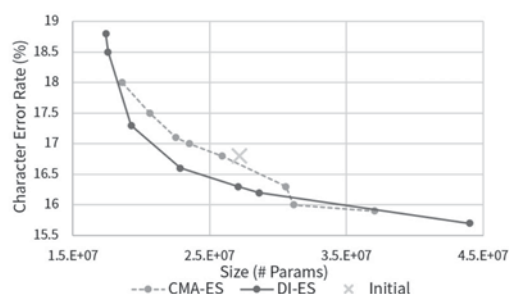


Fig.1: Comparison of CMA-ES vs. DI-ES

2-P2-8

2-P2-8 CTC を用いた複数の音声認識結果から
の高精度な音素列推定High-Accurate Phoneme Sequence Estimation from
Multiple Speech Recognition Results Using CTC Loss

☆李曉陽, 西崎博光(山梨大・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)

- ◆研究目的:
 - 複数の音声認識 (ASR) システムからの音声認識結果を統合し, 正しい音素列を推定するモデルを検討
 - ✧ CTC 損失により音素推定精度が大きく改善できることを確認
- ◆正解音素推定器と損失関数の組み合わせを比較 (4 通り)
 - DNN 正解音素推定器 と CRNN 正解音素推定器
 - 交差エントロピー と CTC 損失関数
- ◆実験結果
 - ベースライン (ベストな ASR システム) と提案手法の性能を比較
 - DNN モデルより CRNN モデルの性能が優れることを確認
 - 交差エントロピーより CTC 損失関数で正解音素推定器の性能が大きく改善することを確認

Table 1 PER [%] in each data set.

Data set	Baseline	DNN		CRNN	
		CE	CTC	CE	CTC
Eval1	9.08	9.07	8.33	8.18	6.77
Eval2	8.23	8.21	7.36	7.14	5.56
Eval3	8.12	7.98	7.33	7.1	5.76
SDPWS	10.56	10.49	9.97	9.92	8.05

2-P1-7

2-P1-7 超多言語事前学習による
低資源音声認識の検討

A Study on Low-resource Speech Recognition by Multilingual Pre-training

☆侯汶昕, 董越, 庄佰融, 楊龍飛(東工大), 史嘉彤(JHU), 篠崎隆宏(東工大)

- ◆42 言語を含む計約5000時間の学習データを用いて音声認識モデルを事前学習する, 超多言語事前学習を行う
- ◆14 の低資源言語への転移学習を行い, 言語による違い等について比較検討する
 - 単一言語転移: 概ねどの言語においても文字誤り率 (CER) が大きく低下し言語間転移学習が有効であることが確認できた
 - 多言語転移: 適応用データが特に少ない場合においても安定した言語適応化ができた

Table: Results of monolingual (left) and multilingual (right) transfer

Language	w/o pre-train	w/ pre-train	Language	w/o pre-train	w/ pre-train
Arabic	42.0	14.2 (66.2%↓)	Arabic	19.7	15.7 (20.3%↓)
Breton	39.5	16.3 (58.7%↓)	Breton	21.1	15.8 (25.1%↓)
Hakha Chin	34.2	11.0 (67.8%↓)	Hakha Chin	14.7	10.0 (32.0%↓)
Chuvash	51.3	68.4 (33.3%↑)	Chuvash	19.2	14.4 (25.0%↓)
Dhivehi	31.1	10.3 (66.9%↓)	Dhivehi	13.7	10.8 (21.2%↓)
Esperanto	5.2	2.3 (55.8%↓)	Esperanto	3.8	2.7 (28.9%↓)
Estonian	30.2	9.9 (67.2%↓)	Estonian	14.5	10.1 (30.3%↓)
Indonesian	33.2	11.3 (66.0%↓)	Indonesian	14.5	10.1 (30.3%↓)
Interlingua	35.1	18.2 (48.1%↓)	Interlingua	14.4	10.5 (27.1%↓)
Kyrgyz	21.8	7.7 (64.7%↓)	Kyrgyz	11.1	8.0 (28.2%↓)
Latvian	30.6	9.4 (69.3%↓)	Latvian	13.7	10.2 (25.5%↓)
Kinyarwanda	84.8	169.5 (99.9%↑)	Kinyarwanda	45.3	31.7 (30.0%↓)
Sakha	36.4	13.3 (63.5%↓)	Sakha	15.3	11.9 (22.2%↓)
Slovenian	26.1	23.2 (11.1%↓)	Slovenian	10.7	8.5 (20.6%↓)
Weighted Avg.	23.0	9.6 (58.3%↓)	Weighted Avg.	11.1	8.2 (26.1%↓)

2-P2-9

2-P2-9 ボトルネック特徴量を用いた
ゼロ資源言語の音声検索の検討Spoken Term Detection Using Bottleneck Features
for Zero-Resource Language

☆水落 智, 伊藤 彰則, 能勢 隆(東北大)

- ◆本研究では, ゼロ資源言語のデータベースに対する単語検索として, 学習に用いる言語と対象言語の言語的な差異による影響を低減するために, 単一言語で学習した音素識別器から得たボトルネック特徴量を用いる手法を検討した. 検索対象言語はカクチケル語とする.
- ◆全ての対象言語において, 単一言語のボトルネック特徴量を用いても検出性能は向上しなかった.
- ◆全ての対象言語において, 複数の言語のボトルネック特徴量を連結した特徴表現を用いても, Posteriorgram を用いた場合とは異なり, 検出性能は向上しなかった.
- ◆今後は複数の言語で学習したモデルから得られる特徴量の使用を検討する予定である.

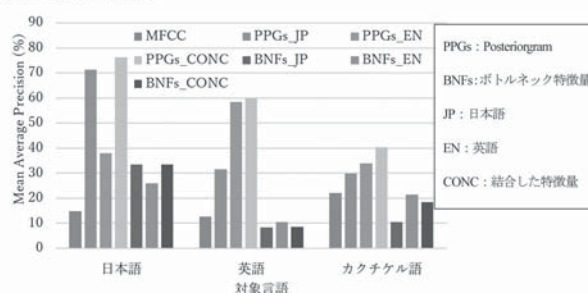


Fig.1: Result of STD using Posteriorgram

2-P3-14

2-P3-14 WaveNet を用いた 言語情報なし感情音声合成における 感情の強さ制御の検討

Study of controlling the Strength of Emotions
in Speech-like Emotional Sound Generated by WaveNet

©松本剣斗, 原直, 阿部匡伸(岡山大学・HS 統合科学研)

- ◆音声は言語情報と非言語情報を通じて感情情報を伝え、感情音声合成においては言語情報と非言語情報を独立して扱えることが望ましい。
- ◆我々はこれまでに WaveNet を用いた言語情報なし感情音声合成方式について検討した [K. Matsumoto et al., APSIPA 2019]。
- ◆HCI への応用のためには感情の強さの制御も必要である。
- ◆本稿では、提案方式を拡張して感情ラベル操作によって感情の強さを制御する方式を検討する。
- ◆感情の強さを評価するために主観評価実験をおこない、実験結果より、感情 Happy において感情の強さを制御可能であることがわかった。
- ◆合成音は <https://kmatu.github.io/SES-emotional-strength/> で試聴できる。

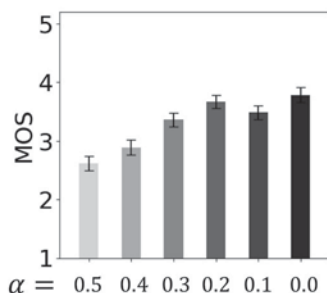


Fig. 1: MOS score for emotional strength of Happy

2-P3-16

2-P3-16 アカペラ歌唱における歌唱速度の 変化を考慮した歌声合成に関する検討

Segment tempo prediction for a cappella singing voice synthesis.

☆片平健太(神戸大), 足立優司(メック株式会社),

田井清登(メック株式会社), 高島遼一(神戸大), 滝口哲也(神戸大)

- ◆歌声合成の研究における発声タイミングの推定に関して、楽譜に記されているテンポを基準として、歌唱におけるノリやタメなどの発声タイミングのずれを再現する手法が提案されている。
- ◆アカペラ歌唱での歌唱速度が頻繁に変化する特徴を捉えるため、音符単位のテンポを推定し発声タイミングを求める手法を提案したが、不自然なテンポ変化が現れることがあった。
- ◆本稿ではテンポの変化を楽譜上の各セグメントの経過時間の変化と捉え、より自然なテンポ系列を推定する手法を検討する。



Fig. 1: Musical score.

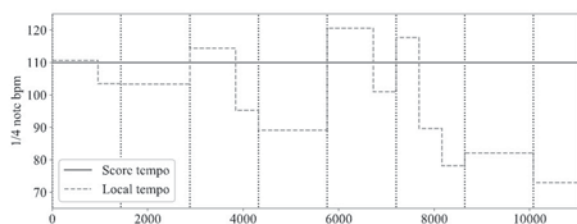


Fig. 2: Comparison of Tempo changes between global and local.

2-P3-15

2-P3-15 LJSing: 単一歌唱者による大規模日本語 歌声コーパス

LJSing; Large-Scale Singing Voice Corpus of Single Japanese Singer

☆藤村拓雄, 能勢隆, 伊藤彰則(東北大)

- ◆ここ数年、東北きりたん, PJS, JVS-MuSiC など研究用歌声コーパスの開発が盛んである。
- ◆現在利用可能な公開歌声コーパスは長くても 60 分程度であり、深層学習を利用した歌声合成手法において十分であるかは不明である。
- ◆本研究では統計的歌声合成の検討のための女性プロ歌唱者一名による計 5 時間の日本語の歌声コーパス LJSing の収録・整備を行った。
- ◆LJSing は歌唱者が選んだ J-POP を中心とした 51 曲からなる SongSet (126 分) とフレーズ単位で音韻・韻律/バランスを考慮して選んだ PhraseSet (174 分) からなる。
- ◆歌声データについては自動でプレス挿入・音素アラインメントを行った後、手作業によりプレスおよび時間情報の修正を行った。
- ◆既存の歌声コーパスと比較した結果、特に音高と音符長のバランスが取れていることをエントロピーにより示した。

Table 2: 各コーパスのコンテキストのエントロピーと音素数、有音時間

コーパス	HTS-M	HTS-F	Kiritan	SongSet	LJSing PhraseSet	All
音素 [bit]	4.08	4.09	4.13	3.96	4.02	4.00
楽音継続長 [bit]	3.29	2.32	2.57	2.71	3.30	3.06
音高 [bit]	3.36	3.35	3.93	4.11	4.43	4.33
音高差 [bit]	3.08	3.00	3.24	3.14	3.42	3.30
音素数 [個]	6,223	6,326	20,556	33,486	37,278	70,764
モーラ数 [個]	3,501	3,555	11,289	18,770	21,176	39,956
有音時間 [分]	30	29	64	126	174	300

2-P3-17

2-P3-17 スペクトログラムを用いた CycleGAN に 基づく高品質ノンパラレル声質変換

CycleGAN-based high-quality and parallel-data-free
voice conversion using spectrogram

☆金垣葵(東北大), 田中雅也(東京農工大), 能勢隆(東北大),

清水遼平(東大), 伊藤彰則(東北大)

- ◆パラレルデータを用いない声質変換では CycleGAN に基づく手法の有効性が示されているが、CycleGAN-VC2 に代表される従来法ではパラメトリックボコーダによる品質劣化により高品質化が困難。
- ◆提案法 Scyclone では音響特徴量にスペクトログラムを使用し WaveRNN ベースのボコーダにより高品質な変換音声を作成。
- ◆CycleGAN による変換時のエイリアシングを回避するため非エンコーダ・デコーダモデルを採用し、スペクトル正規化を用いることで生成器・識別器の安定した学習を実現。
- ◆自然性と話者性に関する主観評価により CycleGAN-VC2 より大幅に主観品質が改善されることを確認。

Tab.1: MOS on speech naturalness.

Target speaker	Ayanami	F009
Target (natural)	4.59 ± 0.12	4.76 ± 0.10
Target (vocoder)	4.63 ± 0.10	4.68 ± 0.11
CycleGAN-VC2	1.80 ± 0.14	1.87 ± 0.17
Scyclone	3.92 ± 0.17	3.36 ± 0.20

Tab.2 :MOS on similarity to the target speaker.

Target speaker	Ayanami	F009
Source (natural)	1.00 ± 0.00	1.06 ± 0.07
CycleGAN-VC2	2.81 ± 0.16	3.20 ± 0.17
Scyclone	4.39 ± 0.14	4.49 ± 0.14

2-P3-18

2-P3-18 Cross-lingual voice conversion with Multi-codebook Hierarchical Vector-Quantized Variational Autoencoder

○ Tuan Vu HO, Masato Akagi (Japan Advanced Institute of Science and Technology)

In this paper, we proposed a cross-lingual voice conversion model based on Multi-codebook Hierarchical Vector-Quantized Variational Autoencoder (VQ-VAE), which can avoid the posterior collapse problem of previous Variational Autoencoder model (VAE) by using a discrete latent representation with constant Kullback-Leibler Divergence loss. In order to capture the linguistic information at different timescale and avoid the phonetic mismatch between different languages, a hierarchical structure for VQ-VAE is proposed. Evaluation results on VCTK dataset and JVS dataset show that the multi-codebook structure yields better quality of converted speech in term of Mel-cepstral Distortion and Modulation Spectrum.

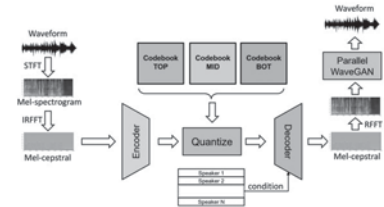


Fig. 4: Overview of proposed multi-codebook hierarchical structure VQ-VAE VC

2-Q1-1

2-Q1-1 バースト駆動による液滴移動
—ガラス基板への強力弾性表面波の励振(III)—
Removal of water droplets by burst waves

- Excitation of high intensity ultrasonic waves on glass plate(III)-

☆淵脇空輝, 中村健太郎(東工大・未来研)

- ◆これまで、ガラス板上に強力な弾性表面波を励振することによって、ガラス板上の異物を除去する方法を検討してきた。
- ◆圧電素子への電圧印加方法を、液滴の固有振動数を繰り返し周波数とするバースト駆動にする方式を提案する。(Fig. 1)
- ◆入力電力一定(8.85 ± 0.03 W)でバースト波の duty 比を変化させても、圧電素子の温度上昇量に有意な差はなかった。(Fig. 2)
- ◆Duty 比を 50 % に固定して、繰り返し周波数を変化させたところ、繰り返し周波数が 20 Hz のときに、水滴を動かすのに必要な入力電力を連続波と比較して約 34 % 軽減できた。(Fig. 3)

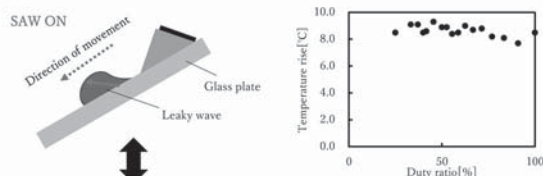


Fig.2 Temperature rise vs. duty ratio.

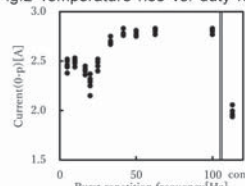


Fig. 3 Current vs. repetition characteristics.

Fig. 1 Excitation of water droplets by burst waves.

2-P3-19

2-P3-19 自己聴取音声の生成で生じる特徴量の変化

Changes of frequency features caused by self-voice generation

☆野々山 大樹, 鈴木 千文, 西野 隆典(名城大)

- ◆自らの声を録音し音声の聴取を行うと違和感を覚える。その原因は、音の伝達経路の違いによるものであると考えられる。本稿では、録音音声をもとに自らが発話時に聴く自己聴取音声の生成と、生成後の音声の評価および分析について報告を行う。
- ◆周波数領域における自己聴取音声 $Y(f)$ の生成モデルとして次式を仮定した。

$$Y(f) = \{H(f) + G(f)\}X(f) = A(f)X(f)$$

$X(f)$: 録音音声信号, $H(f)$: 口元から耳までの伝達特性

$G(f)$: 頭内伝播の伝達特性

伝達特性 $A(f)$ を声質変換手法で代替し、自己聴取音声候補音を生成した。また、主観評価により自己聴取音声を決定した。

- ◆生成した自己聴取音声の特徴の分析として、録音音声と自己聴取音声に含まれる母音のフォルマントを求め、各被験者での共通点や差異がみられるか調査した。分析結果から、第1フォルマント周波数の上昇がみられるといった傾向が得られた。

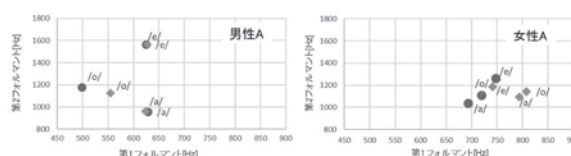


Fig.1: Formant frequencies of recorded and generated voice

2-Q1-2

2-Q1-2 近距離場浮揚から跳躍現象が生じる際の板状浮揚物体の回転

Overture of levitated tabular object by shifting from near-field acoustic levitation to jump phenomenon

☆青野浩平, 青柳学(室蘭工大・院)

- ◆中央から偏心した位置に溝を有する振動子を用いることで、PPC フィルム(0.215g, 40mm×40mm×0.05mm)が跳躍現象により非接触で回転し、さらに跳躍方向が定まることを確認した。
- ◆Fig.1に試作した振動子を示す。幅 15mm および深さ 3mm の溝が中央から 10mm ずれた位置に設けられており、40,590Hz で振動面(40mm×40mm)が縦振動する。
- ◆Fig.2に Fig.1 の振動子を使用して跳躍現象により PPC フィルムが回転するときの様子を示す。PPC フィルムの端を振動面の溝部上にあるように配置することで、駆動時に PPC フィルムが跳躍現象と共に反時計回りに回転し、左側の着地面に裏返って着地した。また、回転中の PPC フィルムは振動面に接触せず空中で回転した。

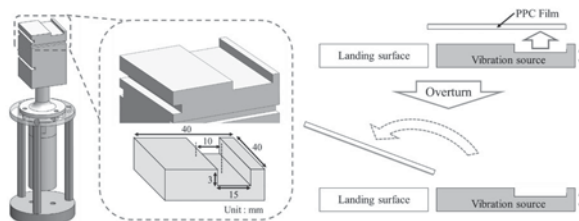


Fig.1 3D view and shape of the vibration source.

Fig.2 Overture of PPC film.

2-Q1-3

2-Q1-3 屋内移動体環境における遅延広がりと
ドップラー広がり計測Measurement of delay and Doppler spreads of sound waves in indoor
mobile environment

☆和田康平(筑波大・エシス), 海老原格, 水谷孝一, 若槻尚斗(筑波大・シス情系)

- ◆屋内移動体環境において、可聴音を用いた音響通信を行う際に、通信路で発生する遅延広がり、および、ドップラー広がりを計測した。
- ◆室内でスピーカーからチャープ信号、および、10 kHzの連続正弦波を出し、マイクロフォンを周回させて計測を行った[Fig. 1(a)]。その結果、最大で66 msの遅延広がり[Fig. 1(b)]、および、 ± 7 Hz程度のドップラー広がりが生じることが判明した[Fig. 1(c)]。また、水中通信路と比較して、遅延広がり、および、ドップラー広がりが大きく、かつ、ドップラー広がり形状も異なることが明らかになった[Fig. 1(d)]。

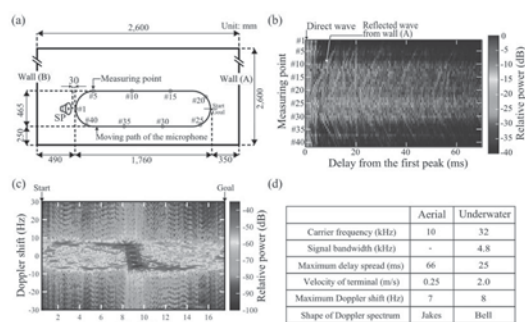


Fig. 1: (a) Experimental site (top-view), (b) measured impulse response, (c) spectrogram of received signal and (d) comparison of delay and Doppler spreads of aerial and underwater acoustic communication channel

2-Q2-5

2-Q2-5 音源方向推定技術を用いて対向させる
パラボラ反射鏡を用いる
水中音響通信シミュレーションUnderwater acoustic communication simulation using parabolic reflectors
facing each other using sound source direction estimation technology

☆茅根涼太郎, 海老原格, 水谷孝一, 若槻尚斗(筑波大)

- ◆指向性トランスデューサとしてパラボラ反射鏡を用いた水中での指向性通信を実現するために (Fig. 1), パラボラ反射鏡を使用した指向性トランスデューサを設計し (Fig. 2), 到来方向を推定することで指向性トランスデューサを向かい合わせにするシングルユーザー (1対1) の通信手順を設計した。さらに、指向性トランスデューサを使用してリンクを確立するシミュレーションを実行し、通信品質を評価した。
- ◆シミュレーションの結果、モデム同士のなす角が 45° 以下のときは誤差 5° 以内で到来方向が推定可能であり、その条件においては、通信品質の指標のひとつである Output signal-to-noise ratio (OSNR) が 10 dB 以上向上できることを明らかにした (Fig. 3)。

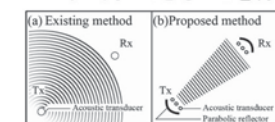


Figure 1: Underwater acoustic communication using parabolic reflector.

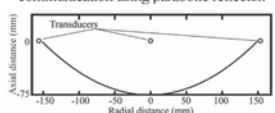


Figure 2: Designed parabolic reflector

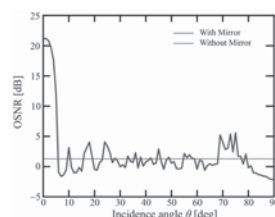


Figure 3: Simulation results; relationship between incident angle θ and OSNR

2-Q1-4

2-Q1-4 散乱群を含む音場における
音源可視化像の評価

Evaluation of sound source visualization in sound field containing scatterers

☆今川俊平, 若槻尚斗, 水谷孝一, 海老原格(筑波大)

- ◆葉が生い茂っている場所や、水中の泡などのような散乱体により遮られた場所に存在する物体を検知する際には、散乱体の間を回折して進んでいく特徴のある音を使うことが有効であると考えられる。
- ◆本研究では、これまでに、散乱体群が存在する二次元のシミュレーション空間を想定し、マイクロホン位置で記録した音圧情報の時間情報を反転させた信号を逆伝搬させることにより音源を可視化することに取り組んできた。
- ◆本稿では、その音源可視化シミュレーションで得られた音源可視化像に散乱体と与える関係の定量的な評価として、散乱体群の密度と可視化像の広がりとの関係および、散乱体群の密度と可視化像の理想音源位置からのずれの関係を見出した。

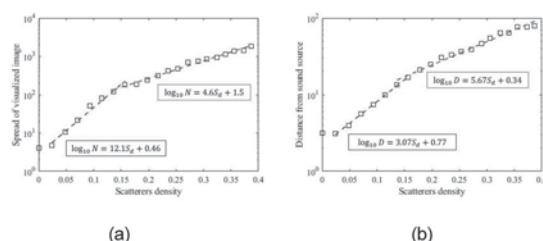


Fig. 1: Evaluation of Influence of Scatterers on Sound Source Visualization
(a) Evaluation of Spread of Visualized Image
(b) Evaluation of Distance from Sound Source

2-Q2-6

2-Q2-6 音響シミュレーションを使用した
エコー源の可視化による
コウモリのセンシング戦略の検討Study of a strategy for bat sensing
by visualizing echo sources using acoustic simulation☆手嶋優風(同志社大・生命医科学院), 山田恭史(広島大),
土屋隆生(同志社大・理工), 飛龍志津子(同志社大・生命医科)

- ◆コウモリのセンシング戦略は、ユニークかつ優れていることが知られており、工学応用を目指しそのモデル化が試みられているが、コウモリの飛行中のエコーを計測できないため、先行研究では、飛行軌跡とパルス放射方向を因子としたモデルしか存在しない。
- ◆本研究では、コウモリの行動計測に加え、FDTD法を用いた音響シミュレーションを使用することで、コウモリの飛行中のエコーの復元を可能にした。
- ◆その結果、コウモリが飛行中に認識している空間の可視化 (Fig. 1A) に成功し、また旋回速度とエコー到来角度に相関があることが分かった (Fig. 1B)。
- ◆この結果より、旋回速度がエコー到来方向によって記述可能となり、エコーを基に飛行しているコウモリ本来の戦略をモデル化することができた (Fig. 1D)。

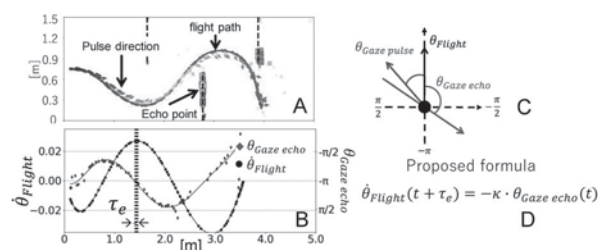


Fig. 1: Flight trajectory, pulse direction and source of echoes during flight (A), Temporal changes in angles of flight turn rate and echo direction (B), Angle definition (C), Proposed formula (D).

2-Q2-7

2-Q2-7 ベトナム産コウモリの耳介運動と
エコーロケーションの関係について

On the relationship between ear movement and echolocation of Tonum bats

☆清水元貴(同志社大・生命), 長谷一磨(名古屋大・環境), 吉野寿紀(同志社大・生命), △小森大輔(香港市大・獣医), △福井大(東京大・農), △Tu Vuong Tan (Vietnam Academy of Sciences and Technology), 藤岡慧明(同志社大・研究開発), 飛龍志津子(同志社大・生命)

- ◆本研究では、ベトナムに生息するキクガシラコウモリ科(Rhinolophidae)及びカグラコウモリ科(Hipposideridae)の計5種のコウモリに対して、エコーロケーション時の耳介運動の計測を行った。
- ◆Fig. 1に *R. pearsonii* の標的(振り子)を接近させた際の耳介先端の運動と、パルス放射・エコー取得タイミングの関係を示す。CF音をパルスに有する種は左右の耳介の開きが最大になるタイミングとエコーの取得タイミングがほぼ一致していた。一方、FM音を用いる種では、規則的な耳介先端の前後運動は行わず、他種と比較すると頭部を頻りに前後に動かしていることがわかった。
- ◆FM音に比べて情報量が少ないCF音を使う種では、耳介運動でエコーに周波数や音圧の変調など付加的な情報を積極的に生み出し利用していると考えられる。

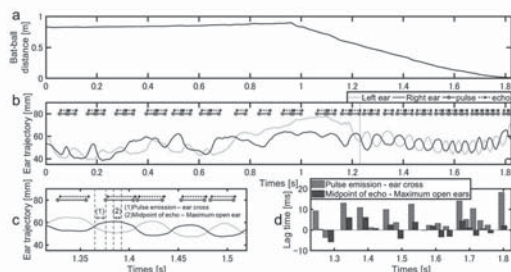


Fig. 1: Temporal changes in position of right and left pinnae of *Rhinolophus pearsonii* during echolocation.

2-Q2-8

2-Q2-8 静脈血流内の高輝度エコー信号の検出

Detection of high-intensity echo signals in the venous bloodstream

○上坂拓実, 大村真朗, 長岡 亮(富山大院・理工),

八木邦公(富山大院・医業), 長谷川英之(富山大院・理工)

- ◆高速超音波イメージングにより抽出した頸静脈血流エコー画像において、高輝度エコーが観察されることがある。これは微小血栓である可能性があり、血液性状などの異常を早期に検出できる可能性がある。本研究では、そのような高輝度エコーの検出方法について検討した。
- ◆血管の任意の断面の深さ方向の位置 z と時間 t における輝度値 $L(z, t)$ の図を作成した。次に、ヒストグラムの解析により高輝度エコーの閾値を決定した。輝度分布図において、閾値以上の箇所を赤色に表示したものをFig. 1(a)に示す。
- ◆塊を目視でカウントした結果と自動でカウントした結果を比較することで結果の妥当性を評価した。結果をFig. 1(b)に示す。検出数は目視、自動検出でそれぞれ15、16となり、両者がおおよそ対応することから、妥当な結果が得られたと言える。

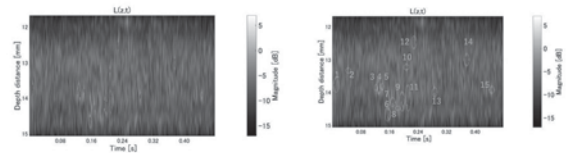


Fig. 1(a)

Fig. 1(b)

Fig. 1: (a) Spatiotemporal (depth-time) distribution of brightness. The detected high brightness regions are colored in red. (b) Comparison with manual detection.

2-R1-1

2-R1-1 適応アルゴリズムを必要としない能動正弦波騒音低減法

An active sinusoidal noise reduction method not requiring adaptive algorithms

○藤井健作(コダウエイ研), △棟安実治(関西大・工), 荻木禎史(千葉工大)

本検討では、1次正弦波騒音の採取、その周波数を検出するセンサー、2次系フィルタ、誤差マイクロホンの出力を最小化する適応アルゴリズム、以上の全てを必要としない低減法を提案する。但し、参照正弦波と参照余弦波を発生させる、例えば1Hzの正弦波の1周期分を保存するメモリ、事前推定した2次系の周波数特性を保存するメモリを用意する必要がある。

本検討は、到来正弦波の周波数を誤差マイクロホンの出力信号から推定する手順。スピーカから送出する正弦波の振幅と位相を調整する手順からなる。また、後者では事前推定した2次系を利用する方法、2次系の事前推定を必要としない方法の2つに分けて説明している。但し、2次系を利用する方法でも2次系フィルタは使用しない。これによって、強い共振特性をもつ応用例において懸念される2次系フィルタの長大化、参照信号がもつ強い自己相関関数によるFiltered-x法の不安定化などの問題への対応は不要となる。

本提案法で実行される主な処理は、低域通過フィルタと直流ノッチフィルタだけである。また、Filtered-x法を適用する方法では安定動作条件を満たすようにステップサイズを事前に選定する必要があり、動作が不安定となる可能性がある。これに対して本提案法では誤差マイクロホンの出力をゼロとする最適係数が毎回算定される。徐々に最適係数に近付ける処理ではない。これによって従来の手法にあるシステムの動作が不安定になるという問題は生じない。

2-R1-2

2-R1-2 能動騒音制御システムにおけるバイアス誤差の低減法に関する検討

A study on bias error reduction method for active noise control systems

○藤井健作(コダウエイ研), △棟安実治(関西大), 荻木禎史(千葉工大)

通常の音響空間にフィードフォワード型能動騒音制御システムを設置する応用例では、騒音検出マイクロホンに周囲騒音が混入する。この混入はバイアス誤差を発生させ、騒音低減効果を低下させる。本検討では、このバイアス誤差を低減し、その混入がない場合と同等の低減効果をFiltered-x法が適用されるシステムで実現する方法を提案する。

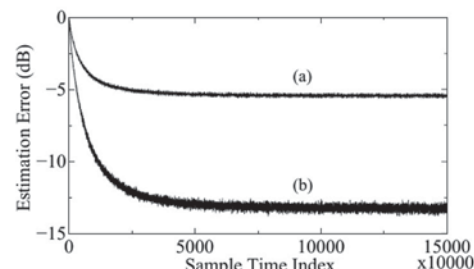


Fig. 1: Transition of noise reduction effect. (a): non-revision, (b) revision

Figure 1は音響系インパルス応答の標本値を指数減衰正規乱数で、1次騒音対周囲騒音のパワー比を0 dBとして計算した騒音低減効果の推移(a)を示している。但し、(a)は補正しない場合の結果、(b)は本提案法による補正を施した場合の結果を示している。また、(b)として示される最大の低減効果は、バイアス誤差を発生させる周囲騒音が無視できる場合の結果に等しい。この結果は本提案法の有効性を示している。

2-R1-3

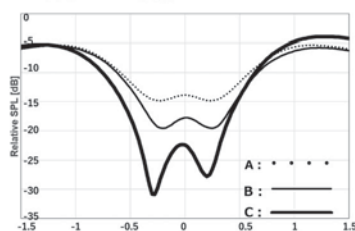
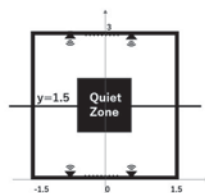
2-R1-3 遺伝的アルゴリズムを用いた Quiet Zone 生成に関する検討

A Consideration on Generating Quiet Zone Using Genetic Algorithm

☆周桐(龍谷大院・理工学研), 東絵莉香(キヤノン(株)),

安枝和哉(東京医大・医療保健), 片岡章俊(龍谷大・先端理工)

- ◆多点制御法を用いた Quiet Zone の生成において、遺伝的アルゴリズムによって抑圧制御点の配置を決定する手法を提案する。
- ◆より音圧を抑えるため、各抑圧制御点に重みを与える。与えた重み係数は遺伝的アルゴリズムを用いて決定する。
- ◆計算機シミュレーションによる評価実験から、遺伝的アルゴリズムを用いた重み付き多点制御法によって、より静かな Quiet Zone が生成できることが確認した。

Fig.1: Relative sound pressure level when $y=1.5\text{m}$ of each method for generating Quiet Zone in a $3\times 3\text{ m}^2$ room.Fig.2: The image of Quiet Zone in a $3\times 3\text{ m}^2$ room.

A: Conventional method which suppression points are allocated at boundary of Quiet Zone
 B: Suppression points are allocated by Genetic Algorithm (Weighting coefficients = 1)
 C: Suppression points' position and their weighting coefficients be generated by Genetic Algorithm

2-R1-5

2-R1-5 局所時間周波数構造に基づく 深層パーミュテーション解決法の実験的評価

Experimental evaluation of deep permutation solver based on local time-frequency structure

☆山地修平, 北村大地 (香川高専)

- ◆周波数領域独立成分分析 (FDICA) では、パーミュテーション問題の解決が必要である。
- ◆先行研究[1]では、FDICA が周波数毎に理想的な分離を行った場合、局所的な時間周波数構造から分離信号の正しいパーミュテーションを予測する深層学習 (DNN) モデルを提案した。
- ◆本稿では、上記の DNN モデルに基づくパーミュテーション解決法を拡張し、実際の FDICA の分離結果に対して、より頑健なパーミュテーション解決を達成するアルゴリズムを新たに提案する。
- ◆2話者の混合音声を対象に分離実験を行い、異なる音源分離手法と比較することで、提案手法のパーミュテーション問題解決性能を示す。

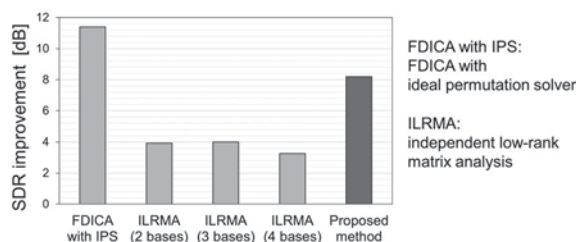


Fig.1: Average SDR improvements.

[1] 山地修平, 北村大地, “局所時間周波数構造に基づく深層パーミュテーション解決法,” 日本音響学会 2020 年春季研究発表会講演論文集, pp. 317-320, 2020.

2-R1-4

2-R1-4 会話音声低減のための サブバンド ANC システムの検討

A Study on Subband ANC System for Reducing Speech Signal

☆牧山雄祐, 梶川嘉延(関西大学)

- ◆開けた場所で多くの人が同時に話す状況では、周囲の騒音や他の会話などのために会話が困難であったり、情報漏洩といった問題が生じる。この問題を解決するためにパーティションにアクティブノイズコントロールを導入する研究を行ってきた。
- ◆ANC パーティションでは、快適な会話を実現するために音声信号を低減する必要がある。しかし、音声信号は非定常な信号であるため、全帯域に焦点を当てた通常の ANC では十分に低減できない問題があり、有色信号のノイズ低減性能を向上させるサブバンド ANC システムを用いることで音声信号に対して有効であると考えられる。
- ◆本稿では通常の ANC とサブバンド ANC の低減効果の比較検討を行い、通常の ANC よりサブバンド ANC の方が低減効果が高いことがわかった。これより、サブバンド ANC は音声信号に対して有効であることがわかった。

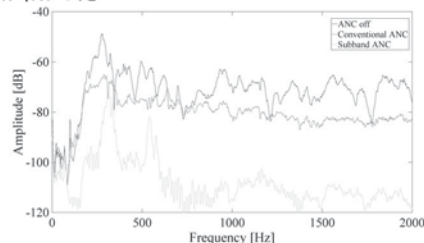


Fig.1: Comparison of noise reduction effect

2-R1-6

2-R1-6 ユーザーからの補助情報を用いる 独立低ランク行列分析

Independent low-rank matrix analysis informed by user annotation

☆大島風雅(香川高専), 中野将生(筑波大), 北村大地(香川高専)

独立低ランク行列分析 (ILRMA) は音声信号に対してまとまった周波数帯域がブロックとしてパーミュテーション不整合が起こることがある。この問題を回避するためにユーザーからのアノテーション情報を受け取り、ILRMA における最適化をより良い解へと導くためのシステムが文献[1]で提案されている。本稿では従来のアノテーション情報の活用方法を変更し、その効果を実験的に調査した。

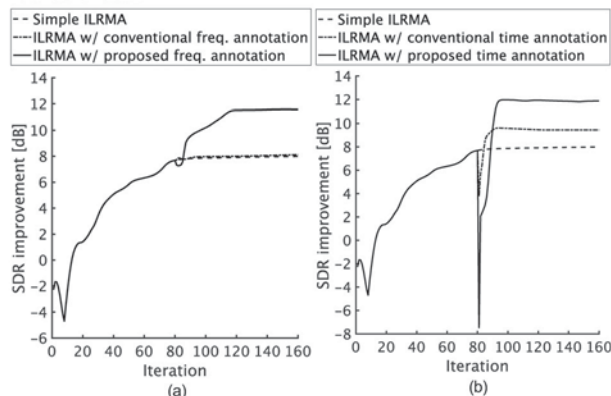


Fig.1: SDR improvements by frequency annotation (a) and time annotation (b).

[1] 中野将生, 北村大地, “ユーザーからの補助情報を用いるインタラクティブ音源分離システム,” 日本音響学会 2020 年春季研究発表会講演論文集, pp. 421-424, 2020.

2-R1-7

2-R1-7 船舶通知システムのための
LSTM-RNN を用いた船舶航行検知Investigation of boat noise detection by recurrent neural network
using long short-term memory for boat notification system

☆山口晴己, 武藤憲司(芝浦工大), 小林洋介(室蘭工大)

- ◆本研究の目的は音と視覚情報の主観評価を利用した運河沿い住民への船舶騒音印象軽減のための船舶接近通知システム開発である。
- ◆先行研究において CNN モデルによって音のスペクトログラムを特徴量とした航行音検知を行ったが、更なる精度向上が必要とされた。
- ◆本研究では入力特徴量間の時系列情報を考慮できる LSTM-RNN モデルを用いて航行音検知を行った結果、CNN モデルによる航行音検出よりも高い精度で航行音を検出することが可能となった。
- ◆入力特徴量間の時系列を学習したことで、CNN モデルではヘリコプターやオートバイの音を航行音と誤認識したものを、LSTM-RNN モデルでは正しく航行音のみを認識することができた。
- ◆各モデルで航行音の有無を 30 分間の評価データを用いて分類した結果を Fig.1 に示す。LSTM-RNN モデルではヘリコプターやオートバイの音を航行音ではない音と認識していることがわかる。

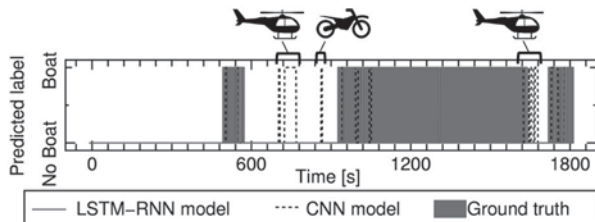


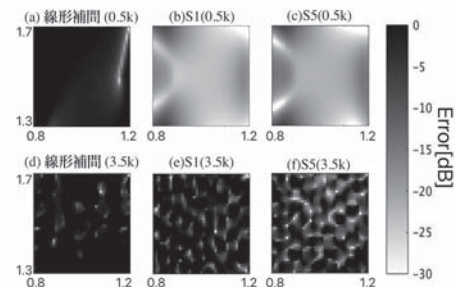
Fig.1: Result of each model classified boat noise using test data.

2-R1-8

2-R1-8 等価音源法を用いた室内インパルス応答
外挿法の数値シミュレーションNumerical simulation of Room Impulse Response Extrapolation
using the Equivalent Source Method

☆松橋 遼, 津國 和泉, 池田 雄介, 小坂 直敏(東京電機大)

- ◆問題点
空間インパルス応答の計測は、大規模なマイクロホンアレイや装置が必要で高コスト
- ◆目的
少数のマイクロホンのインパルス応答から局所音場の高次の反射音を含む空間インパルス応答の外挿
- ◆提案手法
等価音源法に基づき、仮想音源の線形和によって伝達関数を推定
スパース最適化によって仮想音源の重みを決定
仮想音源はn次反射音まで虚像の位置に配置、残響成分は虚像全てを取り囲む閉曲面に配置する
- ◆実験
FDTD シミュレーションを用い推定精度を比較
線形補間に比べ提案手法は0.5~2kHzでは約20dB程改善した

Fig.1: Error Map for each virtual source condition
(upper row: 0.5kHz, lower row: 3.5kHz)

2-R1-9

2-R1-9 衝突音のパワースペクトルを用いた
シジミの良否判別の試みA trial to discriminate between good and bad clams
using power spectrum of collision sound.

☆堀口遼太郎, 篠原寿広, 上保徹志, 中迫昇(近畿大)

- ◆漁師たちは出荷可能なシジミ(良シジミ)と出荷不可能なシジミ(不良シジミ)を手作業で判別している。しかしこの作業は重労働であるため自動判別技術が望まれており、本研究ではシジミの良否判別を行うアルゴリズムの開発を目指す。
- ◆本稿ではシジミをコンクリートに打ち付けた際の衝突音のパワースペクトルを利用しマハラノビス距離を用いた判別分析の結果を報告する。
- ◆衝突音を 689Hz ごとの周波数帯域のパワースペクトル (dB) で分析した。このパワースペクトルを各帯域の特徴量として使用した。
- ◆判別を行った結果、中心周波数 1723 Hz で最も強い特徴がみられた。

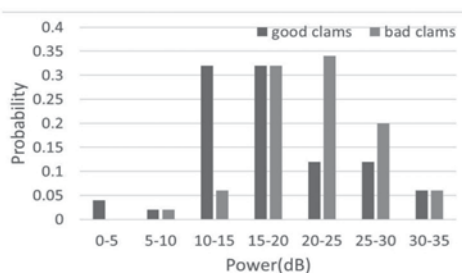


Fig. 1: Distribution of power at the center frequency 1723Hz.

2-R2-10

2-R2-10 キーワード推定を内包した
オーディオキャプション法

Audio Captioning through Keyword Estimation

○小泉 悠馬, 増村 亮, 西田 京介, 安田 昌弘, 齊藤 翔一郎 (NTT)

実現したいこと: 環境音の説明文生成 (audio captioning)

課題: 一つの音響イベント/シーンとそれを説明する単語が一対多対応するため、入力音を説明する文の候補が組み合わせ爆発を起こし、学習中に正解文を出力することが困難となる

解決法: 音説明文生成と音響イベント/シーンに対応するキーワードを同時推定する TRACKE を提案

実験結果: (i) 推定したキーワードを利用して説明文を生成し (ii) キーワード推定を行わないものと比べて性能が向上することを確認

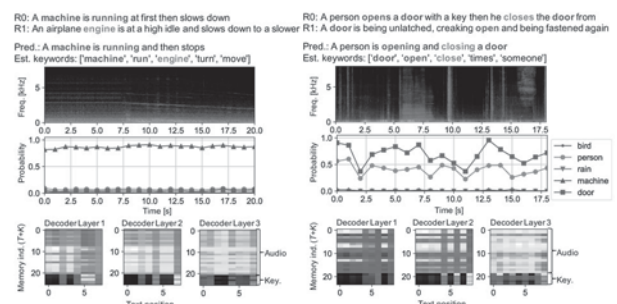


Fig.1: Examples of TRACKE outputs. (i) Ground truth (R0 and R1) and estimated caption (Pred.) and keywords (Est. keywords), (ii) input spectrogram, (iii) keyword posterior of each time index, and (iv) attention matrices of decoder.

2-R2-11

2-R2-11 調波打撃音分離の排他的マスクングに基づくブラインド音源分離

Blind source separation based on exclusive mask obtained by harmonic/percussive source separation

☆大藪宗一郎, 北村大地 (香川高専), 矢田部浩平 (早稲田大)

時間周波数マスクングに基づく優決定BSS (TFMBSS) は, 時間周波数マスクに基づいて線形の (歪みの少ない) 多チャネル音源分離が可能である. この利点を活かして, 調波打撃音分離 (HPSS) に基づく時間周波数マスクをTFMBSSの音源モデルとして用いた手法を提案してきたが, 従来のBSSに比べ優れた性能とは言えなかった.

本稿では, この従来手法の時間周波数マスクの生成方法を改良し, より排他的な (他音源を抑圧するような) 時間周波数マスクが得られる処理をTFMBSSの内部に導入する.

そして, 提案手法と従来手法を比較し, より性能が向上したことを示す. また, 両手法において, HPSSの最適化の反復回数やマスクのスムージングパラメータについて実験的に調査し考察する.

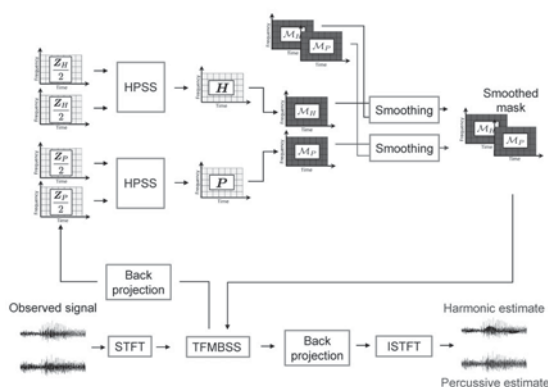


Fig. 1: Block diagram of proposed HPSS-based TFMBSS.

2-R2-13

2-R2-13 スペクトログラム無矛盾性を用いた独立低ランク行列分析

Independent low-rank matrix analysis based on spectrogram consistency

☆豊島直, 北村大地 (香川高専), 矢田部浩平 (早稲田大)

- ブラインド音源分離法の独立低ランク行列分析 (ILRMA) の改良
- 反復更新式にスペクトログラム無矛盾性を担保する処理を挿入
- アルゴリズムの軽微な変更で性能が改善することを実験で確認

Algorithm 1 Consistent ILRMA

Input: $\{\mathbf{x}_{ij}\}_{i=1, j=1}^{I, J}, \text{maxIter}$

Output: $\{\mathbf{y}_{ij}\}_{i=1, j=1}^{I, J}$

- 1: Initialize $\{\mathbf{T}_n\}_{n=1}^N, \{\mathbf{V}_n\}_{n=1}^N, \{\mathbf{W}_i\}_{i=1}^I$
- 2: for iter = 1, 2, ..., maxIter do
- 3: Ensure consistency by (21) $\forall n$
- 4: Update $\mathbf{W}_i$ by (15)–(18) $\forall i, j, n$
- 5: Update $\mathbf{T}_n$ and $\mathbf{V}_n$ by (19), (20) $\forall i, j, k, n$
- 6: Apply back projection by (22) $\forall i, j, n$
- 7: Update parameters by (23)–(25) $\forall i, j, k, n$
- 8: end for

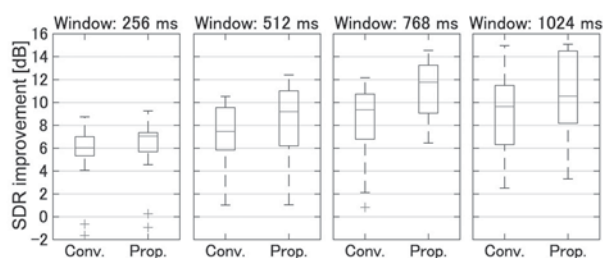


Fig. 1: Example of performance (music, two-source mixture, $T_{60} = 470$ ms)

2-R2-12

2-R2-12 22.2 マルチチャネル音響を用いた球面調和関数展開に基づく3次元音像制御

3-D sound image control for 22.2 multichannel audio based on spherical harmonics expansion

☆鈴木大夢 (立命館大院), 岩居健太, 西浦敬信 (立命館大)

22.2 マルチチャネル音響は, 立体的なスピーカ配置によって音像を3次元に制御可能な音響システムである. この音響システムによる音像提示では, 正確な3次元音像の提示手法が必要とされる. 従来の音像制御手法である vector base amplitude panning (VBAP) では, 所望の音像位置と3台のスピーカ位置から再生信号の分配比を求めることで, 3台のスピーカに囲まれた領域において音像方向を制御することができる. しかし, この手法では音像距離の制御ができないという問題点がある. これを解決するために, 本研究では, 音場に伝搬する音波に対する指向性制御に基づく新たな3次元音像構築手法を提案する. 本手法は, 各方向における基本的な指向性の強さを球面調和関数によって展開し, 目標の指向性と実空間の指向性が合致するように球面調和領域上でスピーカ制御係数を決定する. VBAP, 提案手法を用いて構築した3次元音像位置の精度に関する客観評価実験により, 提案手法が平均 0.28 m 誤差が減少することを確認した (Fig. 1).

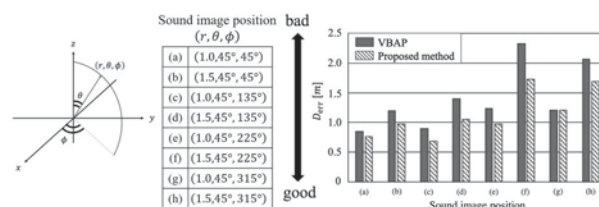


Fig. 1: Distance errors between target sound image position and generated sound image position

2-R2-14

2-R2-14 オンライン会議ツールのDLR法を用いた短時間IR計測による音響的一考察

An investigation of the acoustic propagation in an on-line meeting tool using short term impulse response estimation by DLR method

○福島学, 山下涼介, 大里一矢 (NBU), 大霜康晃, 加藤弘士朗, 石川直生, 上原正志 (千葉工大), 宮崎翔也 (株)ノア, 河納隼一 (株)アーネット, 近藤善隆 (株)ジェイテック, △宇治川大吾 (東京鋼鐵工業(株)), 松本光雄 (風間道子 (log a), △窪田泰也 (A-Lab), 大川茂樹 (千葉工大), 柳川博文 (arsl)

- ◆遠隔講義で使用したツールの受聴者環境を調査した
- ◆短時間IR計測を, IR推定にDLR法, 音源に白色雑音, で実施した.
- ◆比較項目

時間領域: 過渡特性, 精度: 残差エネルギー, 周波数領域: 振幅特性

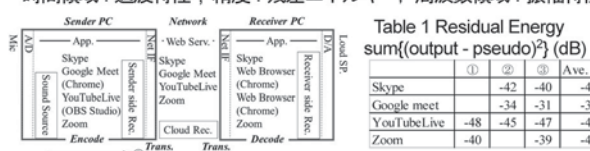


Fig.1 Overview of tele-conference system

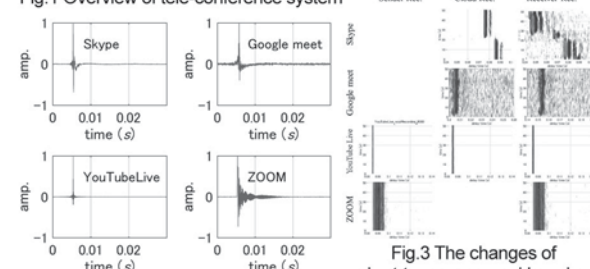


Fig.2 Measured Impulse Response (IR) samples in case of ③

Table 1 Residual energy sum $\{(\text{output} - \text{pseudo})^2\}$ (dB)

	①	②	③	Ave.
Skype	-42	-40	-41	
Google meet	-34	-31	-33	
YouTubeLive	-48	-45	-47	
ZOOM	-40	-39	-40	

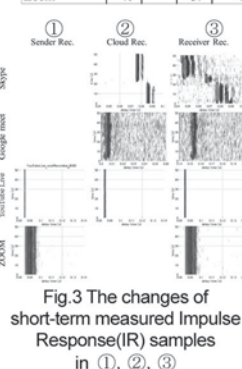


Fig.3 The changes of short-term measured Impulse Response (IR) samples in ①, ②, ③

2-R2-15

2-R2-15 距離スペクトルによる
ソフトな境界領域計測の実験的検証

An experimental study of soft boundary surface measurement
by distance spectrum

☆山下涼介, 大里一矢, 田中歩, 温水啓介(NBU), 大霜康晃, 石川直生,
上原正志(CIT), 河納隼一((株)アーネット), 近藤善隆((株)ジェイテック),
福島学(NBU), 風間道子(log a), 柳川博文(arsl)

- ◆ソフトな境界面を含む領域を音響測距法により計測を行う。
- ◆実験内容: 0.1 (m) 間隔でインパルス応答を計測
各観測点を point 番号で記す
距離による減衰を補正した上で距離スペクトルを求める
- ◆実験結果: point 6 で検出反射物の出現パターンが変わった
point 18 で検出反射物の出現パターンが変わった
距離スペクトルに変化をもたらすことが実験的に確認できた。

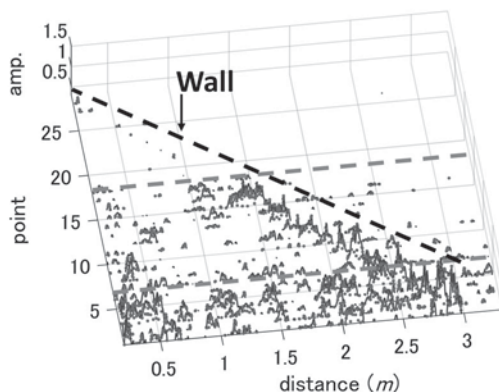


Fig.1 Measurement result (blue: hard, red: soft) (1 point = 0.1 (m))

2-R2-16

2-R2-16 MMG 計測による微小筋活動計測手法の
エコー画像から抽出した筋繊維座標・時間
波形による検証の一検討

An investigation on verification of muscle activity measurement method
by MMG measurement using muscle fiber coordinates time waveform
extracted from echo image

☆大里一矢, 山下涼介(NBU), △樋口幸(大分県立看護科学大),
△市田秀樹(信州大), 福島学(NBU), △森竹隆広(シエルエレクトロニクス
(株)), 柳川博文(arsl)

- ◆エコー画像から筋繊維座標・時間波形を抽出

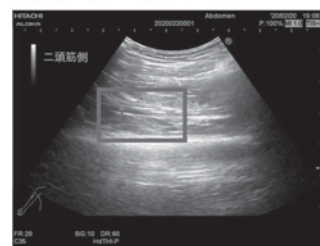


Fig.1 Echo Image

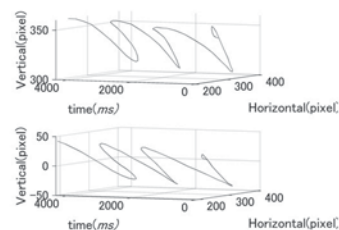


Fig.2 Time Wave