1-1-1

1-1-1　フィードフォワード型能動騒音制御システムにおける２次系の常時推定法

A method of constantly estimating secondary path in feedforward active noise control systems

○藤井健作（コダウェイ研），△棟安実治（関西大），菖木禎史（千葉工大）

騒音制御の安定継続のためには，２次系と帰還系の常時推定が必要になる．このうち２次系の推定は通常，スピーカから白色雑音を送出し，誤差マイクロホンの出力を望みの応答として行われる．これに対して・推定される２次系に帰還系推定誤差と騒音制御フィルタが含まれることが問題点として指摘される．本稿では，この問題の解決法を示す．

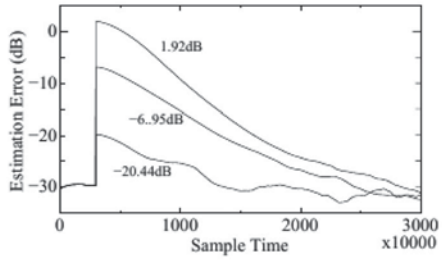


Fig. 1: Transition of secondary path estimation error

Fig.1 は本推定法によって得られた２次系推定誤差の推移を示している．図に示される 1.92 dB，−6.95 dB 等は変化後の２次系インパルス応答との差を表している．この結果において推定誤差は−30.31 dB を初期として制御が開始され，300×10,000 標本化周期の時点で図に示す変化が加えられている．但し，帰還系推定誤差は−30.29 dB に固定されている．この結果から，２次系が変化しても発散することなく，安定して推定誤差が初期設定時の−30 dB まで減少していることが確認される．

1-1-2

1-1-2　ANC システムのバーチャルセンシング技術におけるオンラインモデリングの改善に関する検討

Improvement of online modeling for virtual sensing method in ANC systems.

☆豊岡祥太，梶川嘉延（関西大）

◆ANC システムは騒音に対して同振幅，逆位相の擬似騒音を生成し，騒音と重ね合わせることで騒音低減を行うシステムである．

◆本稿では，所望地点でマイクロホンを設置することなく騒音の低減を可能とするバーチャルセンシング技術（AFVS）にオンラインモデリングを導入し，二次経路変動に対してロバストな AFVS の構成を提案する．

◆提案法では，二次経路だけでなく補助フィルタの再推定も可能である．また，先行研究と比較して騒音低減効果の向上のため，パワースケジューリングの構成を導入している．

◆Fig. 1 に所望地点のスペクトルを示す．提案法では，誤差マイクロホン地点の二次経路が変動しても変動前と同様の低減効果を保つ．

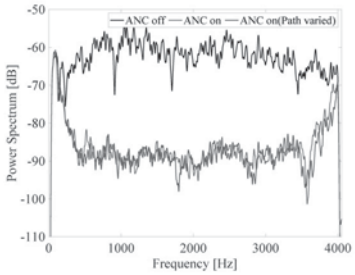


Fig.1 : Frequency spectra at desired position.

1-1-3

1-1-3　多マイクロホンを活用した非定常騒音の時空間予測モデルの基礎検討

Study of a spatio-temporal unsteady noise signal prediction using multiple microphones

◎小塚詩穂里，伊藤弘章，鎌土記良（NTT）

◆広帯域の騒音を能動的に抑圧する手法として Feedforward 型のアクティブノイズコントロール（ANC）が広く用いられている．

◆Feedforward 型 ANC では対象騒音が非定常である場合，その構成要素である騒音を観測する参照マイクや消音対象となる空間に設置した誤差マイクの相対配置，そのアルゴリズム遅延や二次系の遅延などの要因で消音性能が大きく劣化することが知られている．

◆多チャンネルのマイクで観測した過去の信号のうち対象のマイクの未来の信号を予測するのに有用なマイクや時間差を探して重ね合わせる手法を提案し，シミュレーション実験により性能評価した．

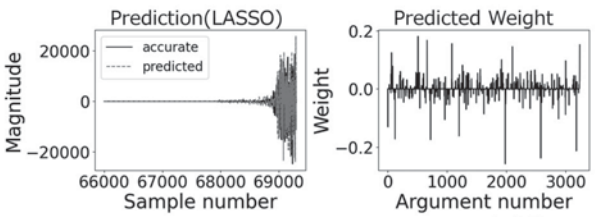


Fig.1 : Signal prediction of proposed method and estimated weights.
(Magnitude ---: accurate, - -: predicted)

1-1-4

1-1-4　日本語母音発声におけるマスクの種類および着用形態のフォルマントへの影響

Effects of Mask Type and Wearing Style on Formants in Japanese Vowel Vocalization

☆木戸孝，　△中村雅昂，　秋田昌憲，　緑川洋一（大分大）

◆マスクを着用する世の中は今後もしばらく続くと考え、マスクの問題点である会話の不明瞭性に着目する。今回は女性話者におけるマスク着用時の発話阻害度、高性能マスクと呼ばれるマスクに種類を変えた場合の比較をし、口唇の動きから発音方法など改善策を示す。

◆日本語母音「あ・い・う・え・お」を発音してもらい、録音した音声から Praat を用いて第１・第２ フォルマントの推移を観察する。また、得られたフォルマント五角形から面積を算出し、増減を調べる。

◆マスク着用時には面積が減少し、口唇の動きが制限される傾向にあるが、自分の顔より少し大きいサイズのマスクを着用すると面積の増加を示した。また、高性能マスクにおいても半数が面積増加を示し口唇の動きの制限が改善される傾向が見られた。

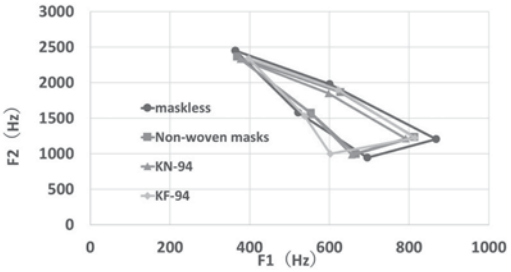


Fig1:Diagram of formants for each type of mask.

1-1-5

1-1-5　肉伝導マイクロフォンを用いた不快感情の検出方法の検討

Investigation of a method for detecting unpleasant emotions using a NAM microphone

○野中彰人，秋田昌憲，緑川洋一(大分大)

本研究室では、肉伝導マイクロフォンにより、胸部から得られた音響信号を用いた入眠予兆検出及び感情変化検出の研究に着手してきた。

本稿では、雑音聴取によって脈の変動や周波数特性にどのような変化が表れるのか調査していく。測定した体内音を 25 次ケプストラム解析し、測定信号の周波数特性の大局的変化を求めるため特徴パラメータを定義する。*v* フレームの *i* 次のケプストラムを *cv* とする時、m~n 次のケプストラムの和(特徴パラメータ)を以下の式で表せる。ここで、パラメータは m 次から n 次までのケプストラムを記述する時の零周波数の値である。雑音聴取区間を赤い枠線で示すとパラメータの値が減少している傾向が見られ、不快感情検出に利用できる可能性が示された。

$$P_v = \sum_{i=m}^{n} c_v(i)$$

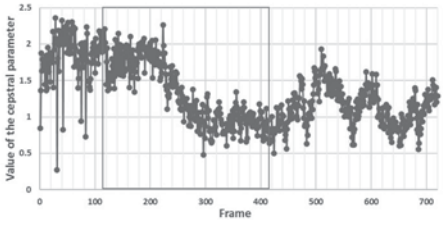


図1 ケプストラムの和の時間変化図（グループＢの被験者Ａ、m=1,n=3）

1-1-6

1-1-6　ディジタル音響信号のサンプル置換による情報秘匿のブラインド検出

Blind detection of information hiding by permutating samples in digital audio signals

○西村 明（東京情報大）

ステガノグラフィは、ディジタルコンテンツになんらかの情報を秘匿する情報ハイディング技術とその利用法であり、秘匿通信がそのひとつの目的である。この技術は、インターネット上の様々なコンテンツへ機密情報を秘匿できるため、テロリストや犯罪組織内部の秘匿通信に使われている、という疑いが、9.11 テロ以前から報道されていた。このような秘匿通信を検出するため、元のコンテンツ無しに、ステガノグラフィによる情報秘匿の有無を検出する手法である、ステガナリシスの必要性が高まっている。なかでも Steghide は、サンプルデータの最下位ビット値が秘匿情報と一致するようにデータを入れ替えて情報を秘匿するソフトウェアであり、統計的性質によるステガナリシスが困難である。

本稿は、これまでステガナリシスが成功したという報告のない、IEEE Dataport で公開されている Audio Steganalysis Dataset 10,000 ファイルを対象とする。最初に、検査対象信号に対して情報秘匿を行い、秘匿前後のコヒーレンスから、ステガナリシス可能な信号を選別する。その後、秘匿前後のパワースペクトル変化の特徴量を用いた SVM 学習を経て、情報秘匿の有無を検出する。

その結果、Dataset のうち 27% のファイルに対しては、ステガナリシスが可能と判定され、それらに対する情報秘匿の有無の予測は、正確性 95.0% を示し、有効なステガナリシスが可能であった。一方 73% のファイルについては、情報秘匿有無判定不能とされ、今後の課題となった。

1-1-7

1-1-7　格子状配置 4 マイクロホンによる音圧比に基づくスピーカー放射方向推定

Estimating direction of loudspeaker based on matching between observed levels by four microphones in a lattice and a directional radiation template

☆津田龍星(阪産大院)，前川大樹，栗谷朋留(阪産大)，
中山雅人，高橋徹(阪産大院)

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

手段 スピーカーの平均放射特性とマイクロホンで観測される音圧レベル比から方向を推定
(1)1/3 オクターブバンドの音圧レベル比を用いる手法
(2)手段(1)の相対音圧レベルの偏差の比を用いる手法
→レベル変動に頑健

評価 マイクロホンが 4 つ(M=4)の場合と 8 つ(M=8)の場合で放射方向推定の正解精度 (=8 方向中から正解方向を推定する頻度) を比較

結論 手段(2)は M=4 でも M=8 と同等の精度で方向推定が可能
2000Hz 以上の帯域の情報が有効

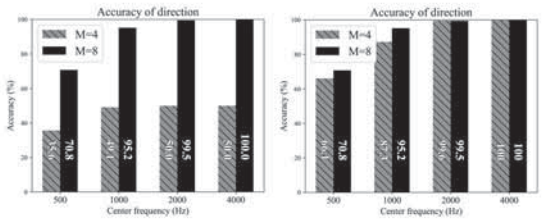


Fig : 1/3 オクターブバンドレベルの中心周波数ごとの方向推定の正解精度
(左図が手段(1)，右図が手段(2))

1-1-8

1-1-8　LSTM を用いた多チャンネル音響信号の空間性識別

Spatial Discrimination Using LSTM with Multichannel Sound Signal

☆柴田咲葵，伊勢史郎(東京電機大)

◆本研究では BoSC マイクで収録された複数の多チャンネル音響信号を用いて機械学習により空間性を識別できるかどうかを検討する．

◆空間性識別を行うための機械学習のために音響信号を時系列データとして扱い，時系列データの学習に適している LSTM を用い 2 クラス分類の学習を行う．一方のクラスは収録した信号をそのまま使用した正データのクラスとする．もう一方のクラスは収録した多チャンネル音響信号に含まれる空間的性質をなくすため，チャンネル番号を学習データごとにランダムにシャッフルした偽データのクラスとする．

◆9 種類の多チャンネル音源を入力として学習を行った．チャンネル番号をランダムにシャッフルした偽データを識別できることがわかった．ただし，学習精度にはバラツキがあり，4 種類の音源については識別できなかった．今後の課題としては機械学習の諸パラメータを変えることで識別できるかどうか，人間が識別できるかどうかなどの検討が必要である．

Table 1 Learning accuracy and similarity

Title	accuracy	similarity
Sea	100%	305
Train	100%	294
Guitar	50%	57
Nou	62%	31
Kizai	50%	5
Orchestra	96%	66
Pipeorgan	90%	247
Semi	48%	277
Song	100%	148

1-1-9

1-1-9　正中面の PNP モデルによる全天空個人化頭部伝達関数の生成

Generation of the individualized head-related transfer functions in the upper hemisphere using parametric notch-peak models in the median plane

☆中村風香，飯田一博（千葉工大・先進工）

◆３次元音像定位を実現するため，正中面７方向の PNP モデルを構築し，それを用いて聴取者が天頂方向の N2 周波数を個人化することにより全天空の個人化 HRTF を生成する方法を提案し，この方法を実現するツールキットを開発した(Fig. 1).

◆生成した個人化 HRTF の音像定位実験を行った結果，水平面，正中面，横断面のいずれにおいても良好な音像定位が可能であることが示された(Fig. 2).

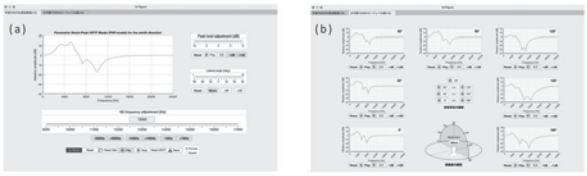


Fig.1 GUI of the PNP model for individualization of HRTF for the zenith direction (a) and for the seven directions in the median plane (b).

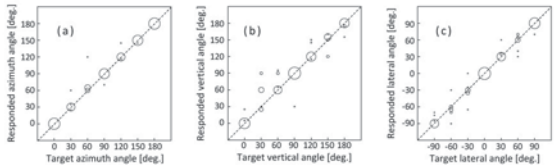


Fig.2 Responses to the individualized HRTFs in the horizontal plane (a), in the median plane (b), and in the transverse plane (c).

1-1-10

1-1-10　時変空間フィルタリングに基づく再生信号適応型マルチスポット再生

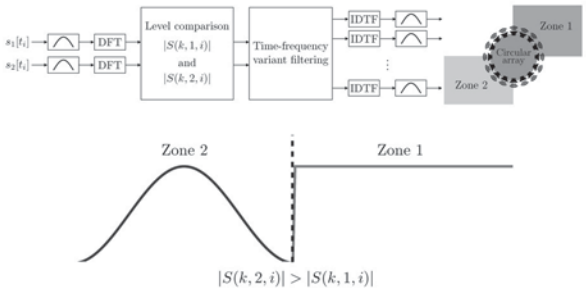
Signal-adaptive multiple sound spot synthesis based on time-variant spatial filtering

○岡本拓磨 (情報通信研究機構)

Q：「再生する音の種類によって適応的に空間フィルタリングを切り替えるマルチスポット再生ってできるんですかß?」
オカモト：「それならこれで!!」

・2領域マルチスポット再生において短時間フーリエ変換の各周波数領域でのレベル差を比較し，レベルの大きい方をHann窓，小さい方を矩形窓とするリアルタイムフィルタリングをリアルタイム逐次再生する方式を提案

・計算機シミュレーションおよび聴取実験において提案法の有効性を確認



1-1-11

1-1-11　音声音響分類トランスフォーマーにおけるパッチ分割の影響

Investigating the Impact of Patching Methods on the Use of Transformer-based Image Classification Models for Audio Classification

☆張詩斉（早稲田大），李莉（NTT），大石康智（NTT），竹内大起（NTT）
仁泉大輔（NTT），原田登（NTT），牧野昭二（早稲田大学）

◆ Transformer model has Self-attention mechanism, which can detect the correlation of information within a sequence, making it excellent for sequence processing tasks.

◆ Most existing Transformer-based models for audio classification are constructed by the original Patching method from the image classification domain. However, whether this original patching method can work or not has yet to be validated due to the difference in physical sense between spectrograms and general images.

◆ In this paper, we reconstructed DeiT and ConViT using the time-shift approach and verified their classification performance and superiority in saving computational resources by comparing the conventional method.

Table 1 Performance and GPU memory cost (inference with batch size of 100 on ESC-50) of different patche sizes based Transformer

Model	Patch Size	Memory Cost	ESC50 fold_1	fold_2	fold_3	fold_4	fold_5	full	Speech Commands v2
DeiT	16 × 16 (conventional)	4959MB	56.50%	58.75%	58.00%	62.00%	55.75%	58.20%	86.69%
	128 × 16 (proposed)	**1751MB**	57.25%	60.00%	63.50%	66.50%	63.50%	62.15%	93.32%
	128 × 6 (proposed)	2519MB	60.25%	59.75%	64.50%	68.75%	64.00%	**63.45%**	**93.83%**
	64 × 6 (proposed)	3179MB	62.25%	57.25%	60.50%	68.50%	59.00%	61.50%	92.79%
	64 × 12 (proposed)	2455MB	61.00%	57.75%	59.50%	63.50%	57.75%	59.90%	92.23%
ConViT	16 × 16 (conventional)	10369MB	56.50%	57.00%	53.50%	64.75%	56.75%	57.70%	91.50%
	128 × 16 (proposed)	**1985MB**	63.25%	62.00%	65.50%	67.50%	62.00%	64.05%	93.40%
	128 × 6 (proposed)	3021MB	63.75%	66.50%	68.25%	71.25%	66.25%	**67.20%**	**94.65%**
	64 × 6 (proposed)	6945MB	59.25%	60.00%	61.75%	66.25%	59.75%	61.40%	94.38%
	64 × 12 (proposed)	3407MB	60.75%	61.25%	59.50%	66.75%	58.25%	61.30%	93.18%

1-1-12

1-1-12　トラフィックカウンタにより取得した学習データに基づく音による交通モニタリング

Training Sound-based Traffic Monitoring System Using Traffic Counter Data

☆高橋知宏，木下裕磨，若林佑幸，小野順貴（都立大），
△本多潤，△福馬誠士，△中川浩（ネクスコ東日本エンジニアリング）

◆我々は交通モニタリングにおいて比較的低コストで交通状態を把握可能である音響情報と深層学習を用いた手法の研究を行っている.

◆深層学習に基づく従来の交通モニタリングには人手でアノテーションされた小規模なデータセットが使用されており，十分な量のラベル付け学習データの収集が課題であった.

◆そこで我々は，高速道路に敷設されたトラフィックカウンタを用いることで交通状態のラベル付けを自動化したデータセットを構築した.

◆畳み込みニューラルネットワークの 1 つである ConvMixer を用いて学習を行い，交通状態（通過した全車両の台数，通過した大型車両の台数，速度，時間占有率）の推定性能を従来手法と比較して評価した.

◆Fig.1 は通過台数（全車両）の推定の時間遷移であり，提案法が最も正解に近い推定を行えることが確認できる.

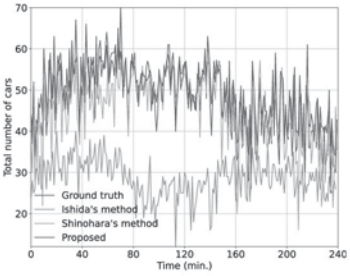


Fig.1 : Temporal change of estimated total number of cars.

1-1-13

1-1-13　DCASECaps: 単一音源に説明文を付与した環境音データセット

DCASECaps: dataset consisting of particular sources with captions

◎岡本悠希（立命館大学/日立製作所），井本桂右（同志社大学），
土肥 宏太，川口洋平（日立製作所）

◆ **単一音源**に説明文を付与したデータセットを構築
- ✓ 従来：複数の音源が含まれる音に対して説明文を付与
  - 単一の音を詳細に説明できてない
    → 混合音中からの特定の音のみの抽出は困難

◆環境音抽出実験にて本データセットの性能評価
- ✓ 二種類のモデル学習手法にて比較実験
  - Training by mixture caption:
    - 混合音に対して付与された説明文を利用したモデル学習を擬似的に再現したモデル学習方法
  - Training by particular caption:
    - 単一音源に対して付与された説明文を利用したモデル学習方法
- ✓ 単一音源に付与された説明文は環境音抽出に有効であること確認

Table 1 SDRi [dB] for extracted signals

Model training method	SNR −10 dB	−5 dB	0 dB	5 dB	10 dB
Training by mixture caption	0.71 ± 0.70	0.70 ± 0.71	0.72 ± 0.80	0.80 ± 1.10	0.87 ± 1.11
Training by particular caption	1.62 ± 1.35	1.49 ± 1.21	1.41 ± 1.13	1.25 ± 1.16	1.12 ± 1.07

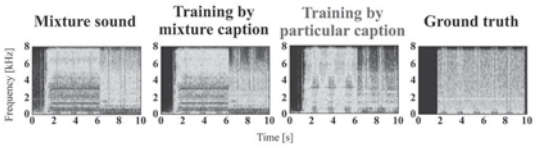


Fig. 1 Results of extracted sound each model training method

1-1-14

1-1-14　汎用音響信号表現 BYOL-A の性能解析

Performance Analysis of the General-purpose Audio Representation BYOL-A

◎仁泉大輔，竹内大起，大石康智，原田登，柏野邦夫（NTT）

◆自己教師あり学習 BYOL-A は高い性能を示した。
- ➢ 課題：学習・モデルの性能への寄与は明らかではない。

◆下記の要素を対象として実験的に分析。
- ➢ データ拡張，エンコーダ構造，BYOL フレームワーク。

◆実験により，エンコーダ構造が大きく貢献していることを確認。
- ➢ CNN エンコーダ構造により得られる特徴量が性能に大きく貢献。
- ➢ BYOL フレームワークと BYOL-A データ拡張による自己教師あり学習により，その後の重要な性能向上を実現。

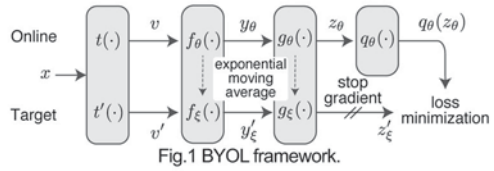


Fig.1 BYOL framework.

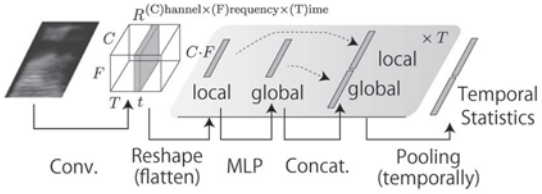


Fig. 2 Encoder CNN architecture.

1-1-15

1-1-15　特徴表現解きほぐしによる音響差分説明文の生成

Audio difference captioning with disentangling feature

◎竹内大起，大石康智，仁泉大輔，原田登，柏野邦夫（NTT）

◆音響説明文生成：音響信号からその説明文を生成するタスク
- ➢ 課題：類似した音響信号には類似した説明文が付与
  ⇒ 説明文からそれらを区別することは困難

◆本研究では，音響差分説明文生成という新しいタスクを提案
- ➢ 類似した音響信号同士の違いの説明文を生成するタスク
  ⇒ 類似した音響信号を区別する情報を付与

◆このタスクの解法として次の２つを提案
- ➢ 相互注意機構に基づく音響差分エンコーダ
- ➢ 類似部分と相違部分を考慮した特徴表現解きほぐし

◆実験より，違いを捉える機構としての相互注意機構や差分の効果と解きほぐし損失が相違部分の特徴量を強調することを確認

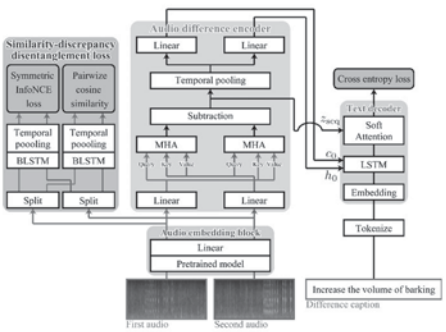


Fig.1 Illustration of proposed method

1-3-1

1-3-1　訳語対の推定と順序入れ替え操作による説明可能な End-to-end 音声翻訳

Explainable end-to-end speech translation
based on predicting bilingual word pairs and reordering

◎大町基（ヤフー），Brian Yan, Siddharth Dalmia（CMU），
藤田悠哉（ヤフー），渡部晋治（CMU）

◆音声翻訳（ST）の代表的な手法として、End-to-end 音声翻訳（E2E-ST）が提案されている。E2E-ST システムの構築においては、出力される翻訳文の品質に加えて、音声から翻訳文を生成するまでの過程の説明可能性を考慮することが望ましい。

◆提案法は、単語レベルでの翻訳の過程を説明可能な系列（Fig. 1）を E2E-ST の枠組みで推定する。本研究では、翻訳文における目的言語単語の絶対位置に基づく表現（**Abs-OP**）と相対位置に基づく表現（**Rel-OP**）の２通りの表現を検討した。
- ➢ **Abs-OP**: 直感的だが、学習データの系列長に強く依存
- ➢ **Rel-OP**: Abs-OP よりも汎化性能が高い

◆本稿では、提案する Rel-OP が説明可能な音声翻訳を実現し、E2E-ST と同等な翻訳文を生成できることを示す。さらに、ストリーミング処理において、提案法が適切な翻訳文を生成できることも示す。

(a)	I　really　need　it
(b)	Ich　brauche　das　wirklich
(c)	I　really　need　it Ich　brauche　das　wirklich
(d)	I Ich [1] really wirklich [4] need brauche [2] it das [3]
(e)	I [NO_OPS] Ich [EOP] really [SM] wirklich [EOP] need [JB] brauche [EOP] it [NO_OPS] das [EOP]

Fig.1: Target sequences for the proposed model: absolute positional operation sequence **Abs-OP** (d) and relative shift operation sequence **Rel-OP** (e). These sequences are generated from the word alignments (c) of the source transcriptions (a) and the target translations (b). Symbols in brackets are post-editing operations dictating target word insertion order.

1-3-2

1-3-2　自発的発話認識のための BERT による非流暢文生成に基づく言語モデリング

Language Modeling Based on Disfluent Sentence Generation with BERT for Spontaneous Speech Recognition.

◎堀井こはる(豊橋技科大), 太田健吾(阿南高専), 西村良太(徳島大), 小川厚徳(NTT), 北岡教英(豊橋技科大)

◆BERT で生成した大量の非流暢文の学習により，自発的発話認識に適した言語モデリングを行った。

◆非流暢ラベル（フィラー「#」，言い淀み「@」）付けされた自発的発話の書き起こしデータで fine-tuning された BERT を用い，大量の書き言葉文にラベルを挿入することで大量の非流暢文を生成。

◆言語モデルを統合しないベースラインでは文字誤り率 10.3%，文誤り率 32.8%だったが，提案手法によってそれぞれ **9.6%**，**30.5%**と有意に改善。

◆生成文のみでも自発的発話に適した言語モデリングに成功。

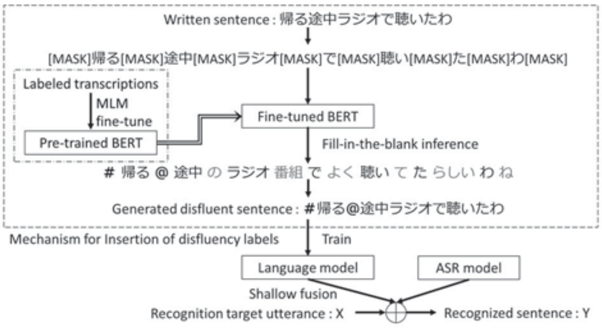


Fig.1 : Overview of proposed method.

1-3-3

1-3-3　中間層予測を用いた End-to-end ダイアライゼーション

End-to-end diarization with intermediate predictions

○藤田雄介, 小松達也, 木田祐介（LINE）, 小川哲司（早大）

◆End-to-end ダイアライゼーション(EEND)は、話者ダイアライゼーションにおける一連の課題（音声/非音声境界の識別、重複区間の検出、話者ラベルの割当）を、単一のニューラルネットワークでモデル化する。既存の EEND モデルは非自己回帰型で、全時間フレームの話者ラベルを同時に生成できるが、出力話者ラベルのフレーム間依存性を考慮した学習が難しい。

◆本研究では，フレーム間依存性を導入した新しい EEND モデルを提案する。提案手法は、中間層で話者ラベルを予測し、そのラベルを用いて上層の条件付けを行う。提案モデルは非自己回帰的に動作するが、中間層において得られる話者ラベル系列全体を参照することでフレーム間の依存性を考慮する。2 話者 CALLHOME データセットで提案手法の有効性を確認した。

Table 1: Diarization error rates (%)

	DER
EEND-NA 8-layer	9.45
+ Intermediate loss	7.34
+ Self conditioning	**7.12**

Table 2: DERs at intermediate layers.

Layer	DER (%)
1	26.05
2	12.76
3	8.93
4	7.89
5	7.17
6	7.11
7	**7.01**
8	7.12

1-3-4

1-3-4　出力の繰り返しを低減する音声認識デコード処理

Speech recognition decoding process to reduce output repetition.

☆上田哲也（早稲田大）, 岡部浩司(NEC)

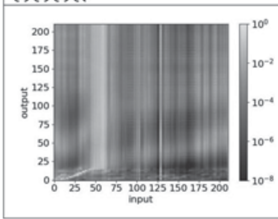
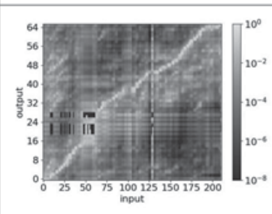
◆Transformer に基づく音声認識器を用いて音声認識を行う際に，同じ認識結果を繰り返し出力する(繰り返し問題)場合がある．従来では，この問題を解決するために，学習時の対応が必要であった．

◆本稿では，学習時の対応を行わずに繰り返し問題を低減するように，以下のように認識時の処理の改良を行った．

➢ 繰り返し問題の有無を検知する．

➢ 繰り返し問題を検知した時だけ，過去に注視した部分を繰り返し注視しないように source-target attention を修正する．

Tab.1: 繰り返し問題が発生する発話に対する認識結果と source-target attention

従来のデコード処理	提案デコード処理
奇麗にしましょうねって話そうそ、	まだ奇麗にしましょうねって話そーそーそーそーそーそーそーそーまずは、苦味を見て、さっきみたいに葉が球態になっていることをまずは、意識で、
	

1-3-5

1-3-5　講義の字幕付与のための Whisper のマルチスタイル fine-tuning

Multi-style fine-tuning of Whisper for captioning lectures

○三村正人, 河原達也（京大）

◆少量の字幕つき音声データを用いて、高精度に講演や講義の自動字幕付与を行うモデルを構築。

◆精度と書き起こしスタイルの両面で有望な Whisper を初期モデルとして、字幕つき音声データによる fine-tuning を行う。

◆書き起こしスタイルの異なる外部データで適応データを拡張し、適応の効果を改善するマルチスタイル fine-tuning の提案。

◆CSJ 講演および京大講義の字幕付与タスクにおいて有効性を示す。

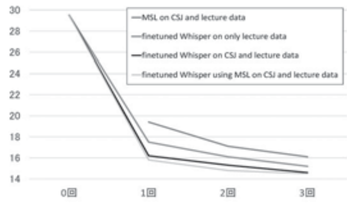


Fig.1 : Caption generation performance on Kyoto-u lecture dataset (CER(%))

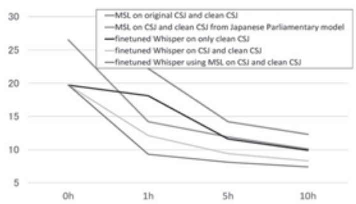


Fig.2 : Caption generation performance on clean CSJ dataset (CER(%))

1-3-6

1-3-6　学習・評価ループを用いたデータ選択による ダークデータからの音声合成

Data selection for text-to-speech synthesis with training and evaluation in the loop

☆関 健太郎, 高道 慎之介, 佐伯 高明, 猿渡 洋（東大院・情報理工）

◆ダークデータ：インターネット上に存在する大規模データ
- データ量は莫大だが，品質は玉石混合
- 本稿ではダークデータ（特に YouTube 音声）からデータ選択を行い，テキスト音声合成コーパスを構築

◆新たなデータ選択指針：学習データとしての品質の評価
- 近年のテキスト音声合成：学習データの品質劣化に頑健
- データそのものの品質と「学習データとしての品質」は乖離
- 学習・評価ループによる評価手法を提案

◆実際のダークデータを用いた実験により手法の有効性を確認

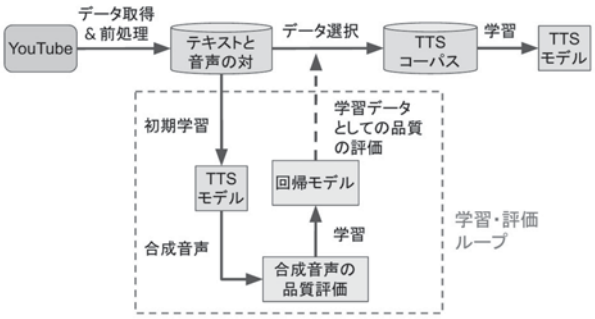


図1：提案手法の概要

1-3-7

1-3-7　Generative Spoken Language Model を用いた 劣化雑音音声の分析と他言語への適用

Analysis of Degraded Noisy Voice and Application to Other Languages Using Generative Spoken Language Model

☆朴 浚鎔, 高道 慎之介, 中村 友彦, 関 健太郎, 辛 徳泰, 猿渡 洋（東大院・情報理工）

◆Generative spoken language model（GSLM）は，自然言語で定義される文字シンボルの代わりに音声から学習されるシンボルを用いる音声分析合成系である．この系は，音声信号に対する処理を自然言語処理と同様に離散シンボルに対する処理として記述できるため，新たな枠組みで自然言語処理と接続できる音声情報処理を確立できると期待される．

◆本論文では，音声強調と音声合成においてその効果を検証する．具体的には以下の通りである．
- 雑音に対する頑健性：GSLM が 音声の劣化に対してどの程度頑健かを調査する．具体的には，種々の signal-to-noise ratio(SNR)と雑音種類を含む劣化雑音音声を入力した際，その音声成分の強調可能程度を評価する．
- 他言語での構築：強い言語依存性が予想される GSLM を，評価済みの英語以外の言語で構築し評価する．具体的には，日本語 GSLM を構築し，その分析再合成音声を評価する．

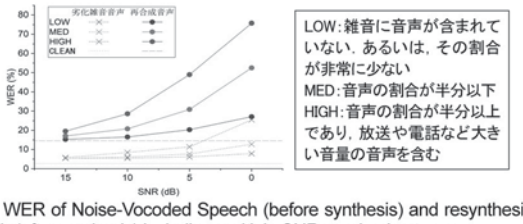


Fig. 1 WER of Noise-Vocoded Speech (before synthesis) and resynthesized speech (after synthesis) including multiple SNRs and noise types

1-3-8

1-3-8　自己教師あり学習モデルを用いた zero-shot 音声合成の検討

Zero-shot text-to-speech synthesis with self-supervised speech-representation model

◎藤田健一，芦原孝典，金川裕紀，森谷崇史，井島勇祐（NTT）

◆自己教師あり学習（Self-supervised Learning; SSL）により得られた音声表現抽出モデル（SSL モデル）を zero-shot 音声合成に活用する手法を提案

◆提案法の概要
- SSL モデルから得られる各フレームの特徴ベクトルから，音響特徴量，音素継続時間長に着目した固定長ベクトルを獲得
- 得られた固定長ベクトルを FastSpeech2 等の非自己回帰型音声合成モデルの条件付けに活用することで，zero-shot 音声合成を実現

◆従来の x-vector を用いた zero-shot 音声合成との性能評価を実施
- 客観評価の結果，提案法は対数メルスペクトログラム，音素継続時間長の生成誤差を大幅に改善することを確認
- 主観評価の結果，提案法は参照音声に対する類似性が向上することを確認

Model	MOS-naturalness	DMOS-similarity
x-vector	3.40±0.05	2.86±0.05
HuBERT(average)	3.25±0.05	3.73±0.04
HuBERT(LSTM)	3.45±0.05	3.73±0.04

Table 1：主観評価結果．

1-3-9

1-3-9　音声認識用コーパスを用いた 複数話者テキスト音声合成における 高品質化の検討

Investigation of improving quality of multi-speaker text-to-speech using speech recognition corpus

☆日田光紀(1,2)，岡本拓磨(2)，西村竜一(1)，戸田智基(3,2)，河井恒(2)
（1 和歌山大学，2 NICT，3 名古屋大学）

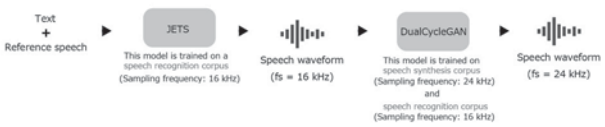
◆音声認識用コーパスを用いた複数話者テキスト音声合成の高品質化を目的とする

◆広帯域の音声合成用コーパスを使った HiFi-GAN で単純に帯域拡張しただけでは品質は改善しない (日田ら，音講論(秋)，2022)

◆本研究では，End-to-end テキスト音声合成モデル JETS を x-vector を用いて複数話者に対応させ，DualCycleGAN を用いた収録環境の異なるコーパス間の帯域拡張を行う

◆x-vector を用いた複数話者 JETS は Tacotron 2 と比較して原音と同等の音声認識率(文字誤り率)を達成
原音：3.1 %，Tacotron 2：15.1 %，JETS：3.2 %

◆JETS の合成音声と JVS コーパスで学習した DualCycleGAN により帯域拡張を実現しつつ合成音声の品質を改善できる可能性を示した



1-3-10

1-3-10　ノイズ掃き出し法とクロスデータ行列法による end-to-end 音声合成の埋め込み空間の可視化

Visualization of embedding space of end-to-end speech synthesis by noise reduction methodology and cross-data matrix methodology

○太刀岡 勇気（デンソーアイティーラボラトリ）

◆合成音声の話者性と感情性を制御することを目的として、Tacotron2 に話者タグとスタイルタグを加えた。

◆音声合成器の内部状態を分析する問題が高次元小標本データであることに着目し、ノイズ掃き出し法とクロスデータ行列法を適用した。

◆CDM は初期の集合の分割によっては主成分分析よりも妥当な結果が得られた。

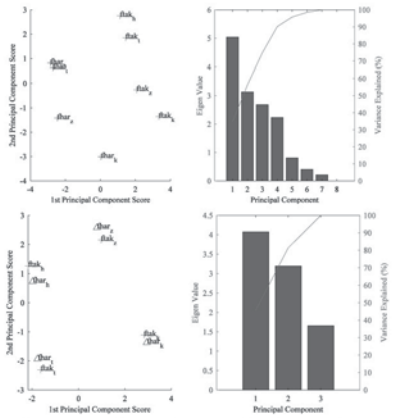


Fig.: Principal component analysis (upper) and cross data matrix methodology (lower) of embedded vectors of Tactron2 dependent on speakers and styles.

1-3-11

1-3-11　F0 パターンと声質情報を解きほぐす深層音声変換モデルの学習法

Learning disentangled representation of F0 pattern and vocal timbre information in deep voice conversion model

◎渡邊千紘，亀岡弘和（NTT コミュニケーション科学基礎研究所）

◆オンラインコミュニケーション等の場面で音声変換を行う際，元音声のニュアンスを変えずに声質のみを変換する技術が望まれる．特に，F0 パターンは音声の意図や感情に強く依存するため，本研究では F0 パターンと声質情報の個別変換に焦点を当てる．

◆近年の音声変換の主流な手法は大まかに以下の 2 種類に分類できる：(1)メルケプストラム係数を特徴とした声質変換と F0 変換を個別に行い，それらの結果からパラメトリックボコーダを用いて音声合成を行う方法．(2)メルスペクトログラムに基づく F0 と声質の同時変換を行い，その結果からニューラルボコーダを用いて音声合成を行う方法．(1)では，F0 の相対的な上下パターンを変えずに声質のみを変換可能であるが，一般的に(2)の手法よりも変換音声の音質が下がりやすい．(2)では，近年，F0 パターンで条件づけたモデルにより F0 パターンと声質情報の個別変換を可能とする手法が提案されているが，これらの手法におけるモデルの学習規準では，変換音声と指定された F0 パターンとの関係が考慮されていないため，変換音声が必ずしも指定された F0 パターンを持つとは限らないという課題があった．

◆本研究では，上記の問題を解決し，変換音声の音質と F0 パターン・声質情報の個別変換を両立するために，入力されたメルスペクトログラムから F0 パターンと声質情報の解きほぐしが可能な深層音声変換モデルの学習法を提案する．

1-3-12

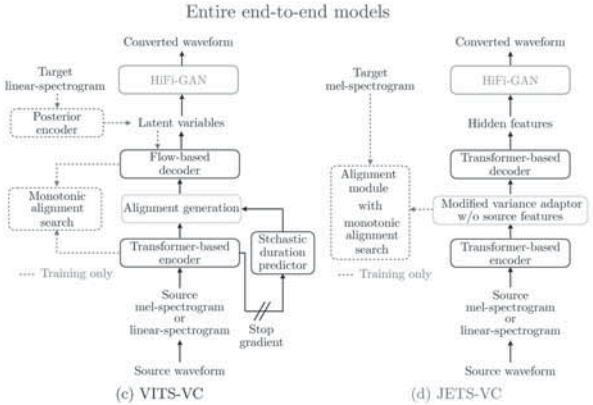
1-3-12　E2E-S2S-VC：End-to-end系列変換型声質変換

E2E-S2S-VC: End-to-end sequence-to-sequence voice conversion approach

○岡本拓磨[1]，戸田智基[2,1]，河井恒[1]
([1]情報通信研究機構，[2]名古屋大学)

・Sequence-to-sequence系列変換型声質変換
- ・フレームワイズ方式と比べて話速や韻律の制御が可能なモデル
- ・非自己回帰型モデルは提案されているが波形生成部とは独立学習

・提案法：End-to-end Sequence-to-sequence型声質変換
- ・End-to-end型テキスト音声合成モデルを声質変換に適用
- ・従来のパイプライン方式よりも高品質な変換を実現



1-3-13

1-3-13　End-to-End 系列変換型声質変換の高速化およびノンネイティブ話者変換の検討

Acceleration of end-to-end sequence-to-sequence voice conversion and investigation of non-native speaker voice vonversion

◎山下陽生(1, 2)，岡本拓磨(2)，高島遼一(1)，滝口哲也(1)，戸田智基(3,2)，河井恒(2)（1 神戸大学，2 情報通信研究機構，3 名古屋大学）

◆ L2 話者から L1 話者への声質変換において以下のことを確認した
- ➢ JETS-VC を用いて End-to-End の有効性を確認
- ➢ ボコーダー部に MultiStream HiFi-GAN、iSTFTNet、FC-HiFi-GAN に変更し、声質変換分野における HiFi-GAN の高速化の適用

◆以下に実験結果を示す
- ➢ L2 話者から L1 話者への変換において、End-to-End 化によって CER が大幅に改善
- ➢ End-to-End 化によってより人間らしい音声を出力可能
- ➢ 高速 HiFi-GAN では RTF の改善はされたが CER は低下

Table.1：Objective Evaluation

	L1 female (slt) → L1 male (bdl)				L1 male (bdl) →L1 female (slt)			
	RTF	MCD [dB]	$\log f_0$ RMSE	CER [%]	MCD [dB]	$\log f_0$ RMSE	CER [%]
original	N/A	N/A	N/A	0.5	N/A	N/A	1.2
CFS2+PWG	3.44	5.76	0.20	3.5	5.23	0.20	3.0
JETS-VC(HiFi-GAN)	0.78	5.80	**0.19**	3.9	5.26	0.18	2.3
JETS-VC(MS HiFi-GAN)	0.50	**5.75**	0.21	4.2	**5.04**	**0.17**	**0.7**
JETS-VC(iSTFTNet)	0.50	9.16	0.23	4.6	5.47	0.18	1.6
JETS-VC(FC-HiFi-GAN)	**0.48**	5.87	0.20	**2.8**	5.56	0.18	1.6

	L2 female (NJS) → L1 male (bdl)				L2 female (NJS) → L1 female (slt)			
	RTF	MCD [dB]	$\log f_0$ RMSE	CER [%]	MCD [dB]	$\log f_0$ RMSE	CER [%]
original	N/A	N/A	N/A	0.5	N/A	N/A	1.2
CFS2+PWG	3.44	7.47	0.22	44.2	6.54	0.20	46.8
JETS-VC(HiFi-GAN)	0.78	6.77	**0.21**	**20.9**	6.04	0.20	**16.9**
JETS-VC(MS HiFi-GAN)	0.50	**6.56**	0.22	21.2	**5.90**	**0.17**	21.8
JETS-VC(iSTFTNet)	0.50	6.93	0.23	25.0	6.61	0.23	25.5
JETS-VC(FC-HiFi-GAN)	**0.48**	6.92	0.26	24.4	6.39	0.18	21.8

1-3-14

1-3-14　ストリーミング処理にむけた Sequence-to-sequence 音声変換モデルの知識蒸留

Knowledge distillation
For Streamable Sequence-To-Sequence Voice Conversion

○田中宏，亀岡弘和，金子卓弘，関翔吾（NTT）

近年，より豊かなコミュニケーションを実現する技術として，音声の各特徴量を所望の特徴量へと変換する音声変換が注目されている．深層学習をベースとしたモデルが広く用いられ，注目機構を含むエンコーダ・デコーダ構造により入出力系列の長期依存性を反映した変換規則の学習が可能となる Sequence-to-sequence (S2S) 学習を用いたモデル学習が高い効果を示している．しかし，エンコーダには非因果層があるため，文全体を取得した後に変換処理を実行することが可能である．実用化を考えると，日常生活において発話終わりを待つことは，円滑な音声コミュニケーションに対する大きな障壁となる．

本稿では，高品質な S2S 音声変換をストリーミング処理可能にするため，教師-生徒フレームワークとして知られる知識蒸留アプローチを提案する．

1-3-15

1-3-15　自己教師あり学習による特徴抽出と coarse-fine変換を用いた少量学習データによる歌声変換

Singing voice conversion with a small amount of training data using feature extraction by self-supervised learning and coarse-fine conversion

☆田中琉聖，△山田淳司，高木信二，橋本佳，南角吉彦，徳田恵一
(名工大)

- 本稿では，歌声変換のための大規模な日本語歌声データベースを用いた自己教師あり学習による特徴抽出，話者表現入力による変換と音響特徴量間の変換を組み合わせたcoarse-fine変換を提案する．
- 主観評価実験の結果より，自己教師あり学習による特徴抽出の利用，coarse-fine変換の有効性が示された．

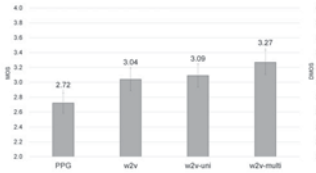


Fig. 1: MOS

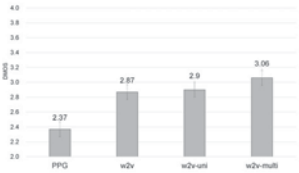


Fig. 2: DMOS

1-5-1

1-5-1　一般的なスピーカを用いた任意の指向性音源構成の試み -全指向性音源に関する検討-

An attempt of constructing arbitrary directional sound source - A study of omnidirectional sound sources -

☆加藤真木，岩見貴弘，尾本章（九大・芸工）

◆一般的なスピーカを多方向へ向けて，球面調和展開係数が全指向性となるようフィッティングを行うことで全指向性音源を構築した。

◆使用するスピーカの指向性が低周波数では全指向性に近いことを利用し，最適化問題を解く周波数ビンを削減した。

◆実際に無響室の複数地点で方向別室内インパルス応答を測定し，全指向性を目標に室内インパルス応答 (RIR) を合成した。

◆提案法は，12 面体スピーカで測定した RIR よりもパワースペクトルの分散が小さく，より平坦な特性に近いことを確認した。

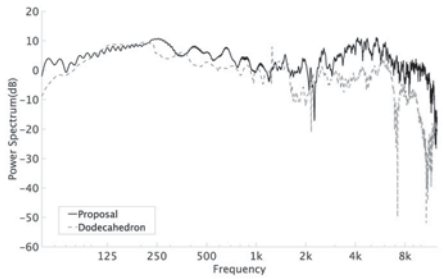


Fig.1　Power Spectrum of Proposal Method and Dodecahedron Speaker. All levels are normalized to 0 dB at 1 kHz.

1-5-2

1-5-2　次数制限付きの方向拡散性 ―定義と音場モデル―

Limited-degree directional diffuseness of a sound field: Its definition and a sound field model

○田中 達宏，大谷 真（京都大院・工学研）

背景　**音場の方向拡散性**を定量的に評価するために，音場の拡散性の**度合い**を適切に定量化したい．

目的　時間領域において音場の方向拡散性を**方向分解能の制限**付きで定式化し，この拡散性を成立させる平面波音場を方向拡散音場モデルとして構築する．

上のように定式化された拡散性を次数制限付きの方向拡散性と名付ける（方向拡散性の**次数**が拡散性の**度合い**に対応する）．

方法　時間領域の音場の球面調和展開係数の分散・共分散を用いて音場の拡散性を定式化する．

時間振幅と到来方向をそれぞれ適切な多変量正規型確率変数と球面デザインで与えた平面波の重ね合わせによって，ある次数までの方向拡散性を満足する音場を構成し，これを方向拡散音場モデルとする．

結果　上の方法でモデル化した音場が適切に方向拡散性を模擬できることを，数理統計的に導くことができた．

簡単な数値実験でも，その妥当性を確かめた(Fig. 1)．

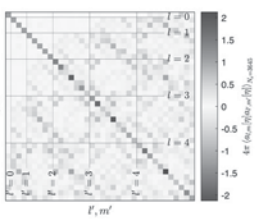


Fig. 1: Sample covariance matrix of a numerically synthesized directional diffuse sound field of 2nd degree.

1-5-3

1-5-3　音響モードを考慮した矩形残響室内音場の減衰特性について

Decay characteristics of rectangular rooms considering Eigenmode

☆花山諒平（京大工），高野靖（京大院）

◆矩形残響室において，音響モードによる減衰率の違いにより，一部の音響モードが支配的となり，残響曲線の折れ曲がりが起こる場合がある。

◆低周波領域では，音の波動性が無視できず，音響モードの影響で残響特性が変化することがある。

◆本研究では，矩形残響室の音響モードを１自由度自由減衰振動モデルとして扱うことで，残響室の共鳴特性および減衰特性をモード合成法（MS 法）によりモデル化した。

◆MS 法による残響曲線では，乱反射率を導入した先行研究と似た折れ曲がりが現れた。また，低周波領域での残響時間に音響モードが影響することを確認した。

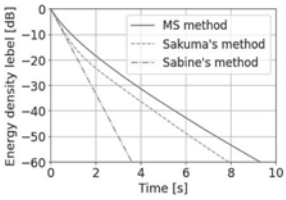


Fig.1 : Energy decay curve of proposed method, Sakuma's method with scattering coefficient s = 0.01, Sabine's method at 500 Hz

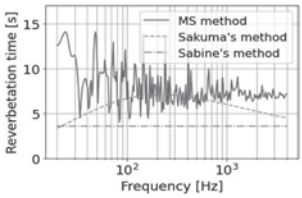


Fig.2 : Frequency characteristics of reverberation time by proposed method, Sakuma's method with s = 0.01, Sabine's method.

1-5-4

1-5-4　声楽者の曲仕上げ過程と室内音響条件の関係 ―三次元音場再現システム“音響樽”を用いた実験的検討―

A research for room acoustic effects on the practice process of solo singer - Experiments using 3D sound field simulation system “Sound Cask” -

☆永野洋介，松尾綾子，上野佳奈子（明治大），伊勢史郎（東京電機大）

◆声楽者の曲仕上げ過程における練習内容と室内音響条件の関係を明らかにすることを目的に，三次元音場再現システム“音響樽”を用いて演奏実験を行った。まず声楽者へのインタビュー調査を元に曲仕上げ過程の検討を行い，その過程に沿って３種類（スタジオ，小ホール，大ホール）の響きの異なる音場のもとで練習を行ってもらい，各音場条件の評価を求めた。

◆演奏実験後の結果から，Ⅰ段階（譜読み・音取り）で行う練習ではあまり響かない音場（スタジオ）が好まれ，Ⅱ段階（表現練習）で行う練習ではやや響く音場（小ホール）が好まれた。また，Ⅲ段階（本番を意識）に行われる練習ではよく響く音場 (大ホール) が好まれた。

Fig.1: Scene of experiment

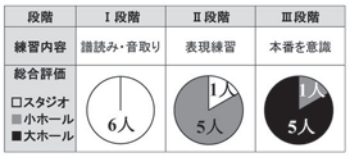


Fig.2 : Experimental results by practice stage

1-5-5

1-5-5　音場支援システムにおけるカラーレーション評価とその低減手法 その３ ―ディレイ変調適用時の聴感印象への影響

Objective detection method of sound coloration in electroacoustic enhancement system – Part 3: Effect on auditory impression when delay modulation is applied.

○渡辺隆行，橋本悌、宮崎秀生（ヤマハ）

◆音場支援システム（AFC）の音響帰還系に起因するカラーレーションに対しディレイ変調による時変制御を用い抑制を試みた。

◆ディレイ変調の聴感印象への影響を主観評価実験により検討した。

◆実使用を想定したシステムでは変調パラメータが変調周波数 ω_m=0.5-1 Hz (変調幅 Δ_r=1 ms)であれば音質劣化を感じさせないことを確認した。

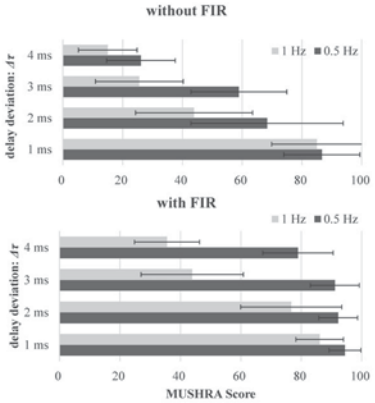


Fig. 6 : Average results of subjective listening tests for delay modulation in a sound field.

1-5-6

1-5-6　先行音効果避難誘導システムの VR 平面空間での音の方向感特性

Characteristics of sound direction in VR plane space of an evacuation guidance system using the precedence effect

☆会田征太郎（日大院・理工），大隅歩，伊藤洋一（日大・理工）

◆先行音効果を利用した音声避難誘導システムの研究を行っている．本システムを実装する場合，基本音響設計に加え，現場での調整が必要となる．

◆このような課題に対して，対象となる建物空間に本システムを設置するときの各種音響条件の設定において，VR の利用を提案している．

◆本報告では，VR 平面空間における先行音効果を利用したときの各位置での音の方向感特性について検証した．

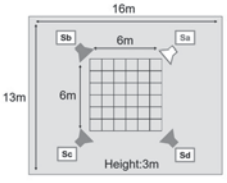


Fig. 1. Schematic of the experiment.

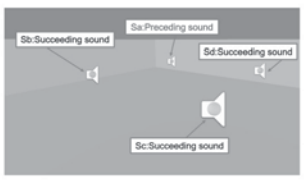


Fig. 2. Visual information in VR.

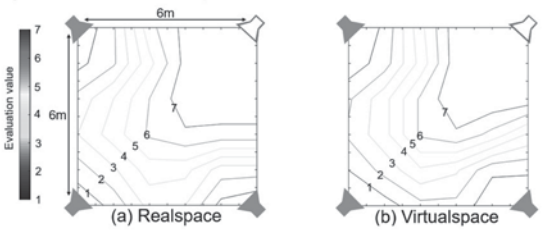


Fig. 3. Relationship between evaluation value and position.

1-5-7

1-5-7　高音域の音圧低下による子音の聴き取りにくさに関する基礎検討

Basic study on listening difficulty of consonants by lowering sound pressure in the high frequency range.

○渡辺茂幸（都産技研），宮入徹（都産技研）

◆高音域の遮音性能確保がある程度可能と考えられる天井が開放されているセミクローズ型ブースを想定して，1kHz 以上の音圧を低下させることでどの程度子音が聴き取りにくくなるのかを検討した。

◆提示音圧はほぼ同等（約 56dB）で 1 kHz 以上の音圧のみを低下させた提示音を用いて単語了解度試験を実施した。音圧低下は Fig.1 に示す特性のローパスフィルタ（LPF1～LPF4）により処理した。

◆今回の検討（Fig.2）では，提示音圧がほぼ同等の提示音でも 1 kHz 以上の音圧低下により単語了解度が低下することが分かった。しかし，高親密度 4 の単語では了解度の低下が低親密度 1 に比べて小さく，子音部分が不明瞭でも類推して回答しているものと考えられる。

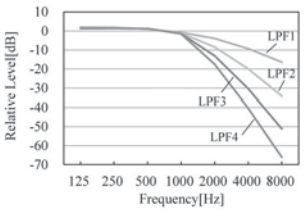


Fig.1：Frequency response of LPF.

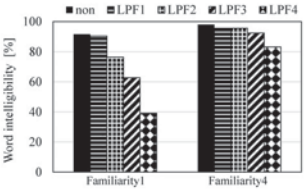


Fig.2：Word intelligibility.

1-5-8

1-5-8　Voice Handicap Index を用いた学校教員の声や喉への負担評価

An investigation into Voice Associated Problems amongst School Teachers, using Voice Handicap Index

○エバンズ直子（TOA（株）/大阪大・基礎工），△金子美樹（大阪大・基礎工），
△重松大輝（大阪大・基礎工），△阪本浩一（大阪公大・医），△清野健（大阪大・基礎工）

◆学校教員は日常的に長時間声を使う職業であるため、声や喉の不調の発生リスクは高いとされる。一方、授業中の教員の明瞭な声での発話は、教育効果を高めるために不可欠である。

◆国内の学校教員の声や喉の負担や障害に関する実態を明らかにするため、埼玉県内の学校教員を対象として自覚評価指標 VHI を用いたアンケート調査を実施した。

◆合計 830 名の学校教員（性別は男性 431 名、女性 390 名、その他または無回答 9 名）から得られた回答を分析した結果を報告する。

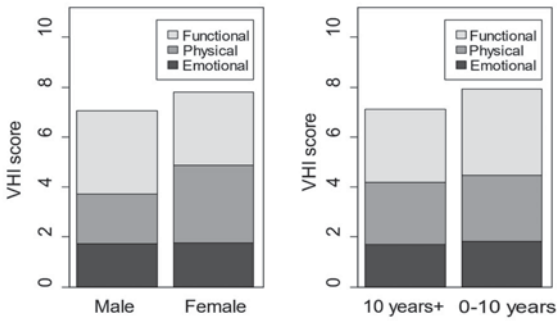


Fig.1：VHI score comparison, between gender (left) and regarding years of experience as a school teacher (right)

1-5-9

1-5-9　外部音場解析解を用いた FDTD 法への音源の放射指向性の導入

Introduction of directional sound sources to the FDTD method based on the analytical expression of exterior sound field.

☆小川晃史，大谷 真（京都大院・工学研）

◆FDTD 法に球面調和関数で表される指向性音源を導入する手法を提案する．提案法では，音源を囲む閉領域（音源領域）を定義することで，外部音場の解析解を音源領域外での時間発展演算に導入する．

◆提案法で表現可能な指向性には，**周波数，球面調和次数，音源領域の大きさの関係に依存した制限**があると考えられ，低周波，高次数の指向性の導入には大きな音源領域を確保する必要があると考えられる．

◆オープンデータとして公開されているギターの指向性（4 次まで，実測値）を導入した場合の相対誤差を Fig.1 に示す．結果より，提案法により**−15 dB 以下の相対誤差**で音源指向性を表現可能であることが分かる．

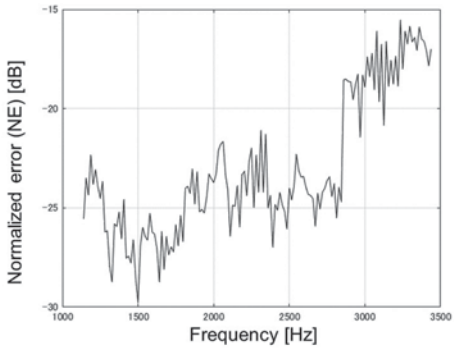


Fig. 1 Normalized error between analytical solution and FDTD

1-5-10

1-5-10　FMBEM 指向性解析に基づく音場の拡散性指標の計算

Calculation of sound-field diffusion indices based on FMBEM incidence directivity analysis.

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

◆残響室内の音響測定では音場の拡散性による影響が報告されており，指標化によって拡散性を把握する試みが様々に行われている。

◆本報では筆者らが提案したHF-FMBEMに基づく入射指向性解析を利用し，直方体室内の定常音場において２種類の拡散性指標を計算した。

◆指向拡散度 d_θ と等方性指標 d_ϵ を計算し，d_θ の計算で行う観測系の方向分解能による基準化を d_ϵ の計算においても新たに導入した。

◆周波数の上昇に伴い d_θ は概ね一定の値を取るのに対し，d_ϵ は低下するという異なる傾向を確認した。

◆両指標値ともに，反射性の音場ほど値が大きくなることや，吸音偏在により固有周波数付近でディップが生じることを観察した。

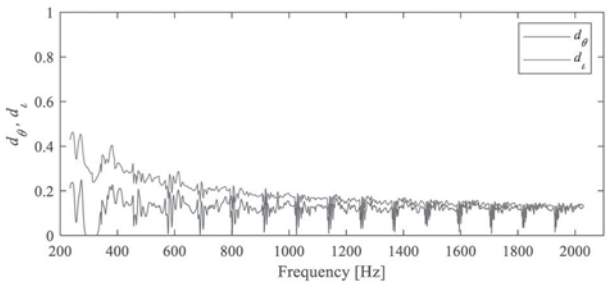


Fig. 1 Diffusion indices in a room.

1-5-11

1-5-11 音響エネルギ密度による室内の残響特性の測定に関する数値解析的研究

Numerical investigation on measurement of reverberation characteristics in a room using acoustic energy density.

☆木村太紀（大分大院・工学研），富来 礼次，岡本 則子，
大鶴 徹（大分大・理工），金澤篤秀（大分大院・工学研）

◆ 小空間における低周波数域の測定では，音圧に比べ，音響エネルギ密度の測定値のばらつきが小さい．本報では，音響エネルギ密度による残響時間の測定値に生じる不確かさに関する検討を行った．

◆ Fig.1に示す34m³ 程度の矩形室を対象に，境界要素法を用い，多点の音圧および音響エネルギ密度の時間応答から減衰特性を求め，残響時間の標準偏差や平面分布を算出した．さらに測定値の出現頻度を定量化した．

◆ Fig.2 に位置エネルギ密度 E_p，運動エネルギ密度 E_k，音響エネルギ密度 E より算出した全受音点の残響時間の標準偏差を示す．E_pと比較し，E_k，Eの標準偏差が小さいことが確認された．

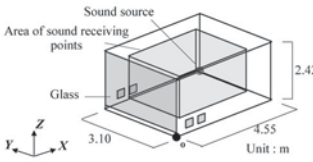


Fig. 1 Analyzed sound fields (RCroom)

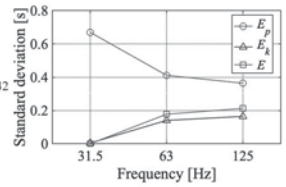


Fig. 2 Standard deviation of T_{30} based on E_p, E_k, E

1-5-12

1-5-12 外周壁から室内への低音域音響透過に関する数値解析的検討

Numerical analysis on low-frequency sound transmission from façades into room

◎劉金雨（東大・エ），井上尚久（前橋工大），佐久間哲哉（東大・エ）

◆本報では窓面と外壁に板要素を適用した有限要素法解析により，二重窓と外壁重量化の防音効果を検討した。

◆窓剛性の影響：外壁を剛条件，一重窓・二重窓を膜・板要素（剛性有無）として内外レベル差を算出した。一重窓に対する二重窓の改善量 ΔD_{free} を Fig. 1 に示す。二重窓共鳴周波数 50 Hz 帯域及び窓間 2 次元音場の固有周波数 100 Hz 付近で，窓ガラス剛性により ΔD_{free} は顕著に上昇した。

◆外壁質量・剛性の影響：一重窓・二重窓（剛性あり）設置時に，外壁質量 25, 55 kg/m² の 2 条件に対して膜・板要素（剛性有無）を用いて，内外レベル差を算出した。外壁重量化による改善量 ΔD_{free} を Fig. 2 に示す。一重窓の場合は窓からの透過の寄与が支配的になるためと考えられる。一方，二重窓の場合は一重窓の場合と異なり，100Hz 付近の低周波音対策として外壁の重量化が有効となる可能性が示唆される。

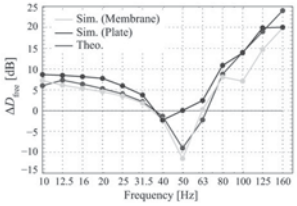


Fig. 1 Effect of double window on single window (with membrane/plate element).

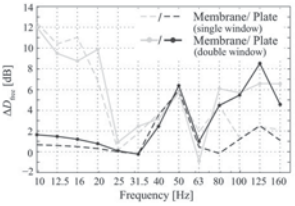


Fig. 2 Effect of weighting outer walls with a single/double window.

1-5-13

1-5-13 3次元位置での音圧を反映した CE-FDTD 法による放射特性の再現

Reproduction of radiation characteristics by CE-FDTD method applying sound pressure at 3D positions

◎大久保翔太，堀内俊治（KDDI 総合研）

◆CE-FDTD 法における入力音圧，非伝播領域を３次元位置に配置する手法を提案し，音源の放射特性についての再現を試みた．

◆音響特徴量のうち音圧レベル差，到達時間差，スペクトル重心，LSD (Log-Spectral Distance)を用いて定量評価を行った．

◆音圧レベル差（Fig. 1）から提案手法により全ての評価位置で誤差が軽減されることが確認された．

◆スペクトル重心の結果から，評価位置により異なる周波数重心の傾向が提案手法により類似する傾向が得られた．LSD の結果から，平面における入力手法において誤差の大きい 180 度方向において，全ての周波数帯で誤差を 6dB 以下に軽減された．

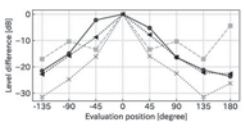


Fig.1 Result of level difference

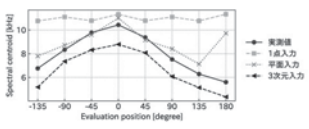


Fig.3 Result of spectral centroid 180 degrees

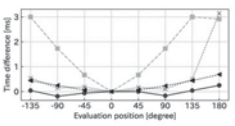


Fig.2 Result of time difference

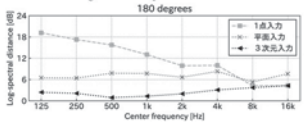


Fig.4 Result of log spectral distance

1-5-14

1-5-14 圧縮加工に由来するかさ密度分布を有する繊維系多孔質材料の吸音率の有限要素解析

Finite element analysis on the sound absorption coefficient of a fibrous porous material with a bulk density distribution derived by the compression forming

○井上尚久（前橋工科大学），上田康介，佐波慧也(HOWA)

◆圧縮加工に由来するかさ密度分布を有する多孔質材料内部の音場の有限要素解析に関する考察を行い，その一実装を提案する。

◆以下本稿で提案する計算方法の概略を示す。

- 圧縮前後の面密度の不変性より圧縮後のかさ密度を算出する
- 詳細な点群データから解析用の有限要素へマッピングを行う
- 各有限要素に対して局所的なかさ密度を与え，多孔質材料内部の音響特性は Kato モデルにより実効密度・複素体積弾性率を算出する

◆粗毛フェルトを圧縮成形したテーパー型の多孔質材の吸音率の解析例を示し，物性値が分布する影響を検証した。

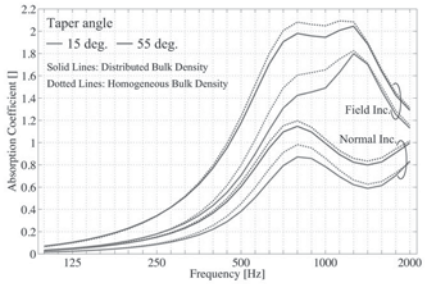


Fig. 1 The field- and normal-incidence absorption coefficients calculated for two different tapered felts.

1-6-1

1-6-1　新幹線鉄道騒音の推計モデルによる試算

An attempt to estimate
Shinkansen railway noise using the estimation model

☆遠藤路悠(近大院)，原田和典(岡山県立大)，
菅原彬子，平栗靖浩(近大)，横島潤紀(神奈川県環境科学センター)

◆キーワード：ノイズマップ、騒音推計、騒音実測調査
高速鉄道、推計モデル

◆新幹線鉄道騒音は環境基準が定められている。現行の実測ベースの評価に対して、環境省を主体に、新幹線鉄道騒音の推計による評価を行うための推計モデルの検討が進められている。

◆本研究では、建物群内における新幹線鉄道騒音の推計値$L_{AE,\mathrm{calc}}$ [dB]の精度検証を目的として、新幹線鉄道騒音の実測値$L_{AE,\mathrm{meas}}$ [dB]と比較し検討した。

◆Fig. 1 は、新幹線鉄道騒音の推計値と実測値の関係である。推計値は実測値より 5 dB 程度小さくなった。Fig. 2 は、トンネル抗口からの距離と、推計値から実測値を引いた差の関係である。トンネル抗口からの距離が近いほどトンネル抗口音の影響で推計精度が低下していた。

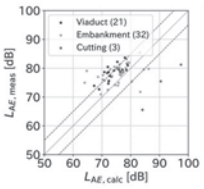


Fig.1 Relationship between measured and calcuratedated noise level

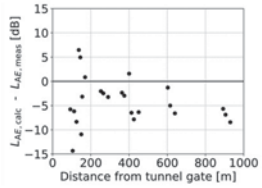


Fig.2 Relationship between distance from tunnel gate and noise level difference

1-6-2

1-6-2　交通騒音による睡眠影響　その6 体動と心拍の併用実験に関する予備検討

Sleep effects on transportation noise Part 6

Preliminary study on combination experiment of body movement and heart rate

☆青木大成，高橋達樹，上田麻理(神奈川工科大)，廣江正明(小林理研)

◆我々は 2019 年以降，実験協力者宅内で交通騒音を曝露した睡眠(介入)実験を行い，質問紙による主観評価と手腕型体動計による客観評価を用いて睡眠影響を評価した.

◆新たな客観的データとして体動以外に心拍を加えた睡眠調査を行うことを考えた．本研究ではその第一段階として体動と心拍を併用した睡眠/覚醒判定の妥当性を確認するための予備的な実験を行った.

◆アクチグラフの睡眠/覚醒の判定を基準とし，覚醒判定のうち心拍計も覚醒と判定した割合（感度）と，睡眠判定のうち心拍計も睡眠と判定した割合(特異度)を求めた.

◆Fig.1 に航空機騒音 $L_{Aeq,1h}$=32.5 dB の場合の中途覚醒（アクチグラフ）と LF/HF（心拍計），騒音レベルの時系列データの分析例を示す.

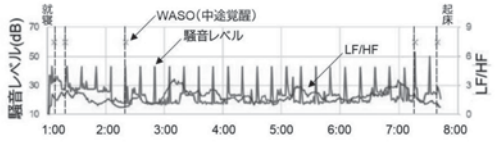


Fig.1 中途覚醒(アクチグラフ)・LF/HF(心拍計)・騒音レベルの時系列データの分析例
－航空機騒音 $L_{Aeq,1h}$=32.5 dB－

1-6-3

1-6-3　道路交通振動を対象とした住宅床の複合振動評価に関する検討

Study on evaluation of combined vibration of housing floors for road traffic vibration

○松田貫(日大院・理工)，冨田隆太(日大・理工)

◆「建築物の振動に関する居住性能評価規準・同解説」は、鉛直振動と水平振動がそれぞれ独立して知覚閾を超える単独振動の評価方法については記載されているが、鉛直振動と水平振動の複数軸が知覚閾を超える場合の評価方法については記載されていない。そこで、本報では、バス・トラック・車等によって発生する道路交通振動を対象に複合振動の感覚評価法について検討を行った。

◆検討の結果、複合振動(鉛直振動≧水平振動)の場合、人の知覚度合は、知覚閾を超える振動方向数の影響を受け、鉛直振動に加えて水平振動も加味して評価していることがわかった。

◆道路交通振動を対象とした複合振動の知覚度合の評価は、下記の鉛直振動と水平振動の合成式である$L_{WC,P}$を用いることで対応は高まることを示した。また、本報の範囲である鉛直振動と水平振動の相対レベルが 10dB の場合には、複合振動補正係数k_{RL}は 0.2 であることを示した。

$$L_{WC,P} = 10\log\left(10^{L_{WZ,P}/10} + 10^{P_h/10}\right) \quad (1)$$

$$P_h = 10\log\left(10^{L_{WF\&A,P}/10} + 10^{L_{WL,P}/10}\right) + C_{RL} \quad (2)$$

$$C_{RL} = 0.2 \times RL \quad (3)$$

$L_{WC,P}$：複合振動評価レベル (dB)、C_{RL}：複合振動補正レベル (dB)、k_{RL}：複合振動補正係数、RL：鉛直振動と水平振動の相対レベル (dB)

1-6-4

1-6-4　住宅床に伝搬する交通振動を対象とした複合振動評価法の一検討

A study of combined vibration evaluation method for traffic vibration propagated to housing floors

○松田貫(日大院・理工)，冨田隆太(日大・理工)

◆本報では、バス・トラック・車等の道路交通による振動と新幹線による振動を対象に鉛直振動と水平振動の複合振動(鉛直振動≧水平振動)の感覚評価法について検討を行った。

◆検討の結果、鉛直振動と水平振動の相対レベルが 10dB 以上と 10dB 未満で異なる複合振動補正係数k_{RL}を用いることで知覚度合との対応は高まることを示した。

◆その複合振動補正係数は下記の Fig. 1 の通りである。

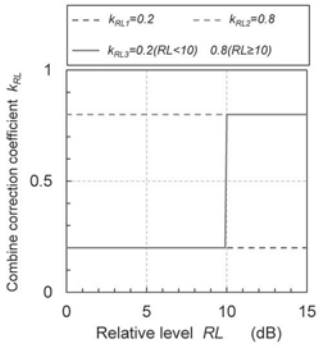


Fig. 1 Relationship between combine correction coefficient k_{RL} and relative level RL

1-6-5

1-6-5　身の回りの VHF 音 －屋外のVHF音の音源探査－

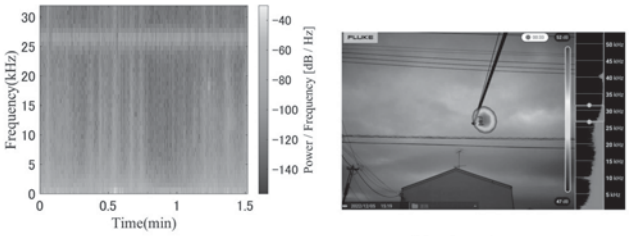
VHF sounds around us － A survey on outdoor VHF sound sources －

☆春澤恒輝，△鈴木勇人（神奈川工科大），廣江正明（小林理研），
△神崎晶（東京医療センター），上田麻理（神奈川工科大）

◆近年、家電製品等の我々の身の回りの生活音に超高周波音（VHF 音：16 kHz~32 kHz の音）を含むことが多数報告されている。

◆先行研究では、主に家電製品から発生する VHF 音に着目していたが、本報では、屋外の VHF 音源に着目し、交通騒音や自然の音などを対象に調査を実施した。

◆以下の図は、調査の中で確認された VHF 音源の１つの(a)スペクトログラムと発生源を(b)音カメラで撮影した写真である。超音波を使用し、交通量や感応式の信号機の切り替え等を行う機械である「超音波式車両感知器の送受器」が VHF 音源であった。



(a)　Sound spectrogram　　(b)　Sound camera

Fig.1 Ultrasonic vehicle sensory transmitter/receiver

1-6-6

1-6-6　分散音響計測ネットワークを用いた 音源探査手法における反射面の影響

Influence of reflective surfaces in Sound Source Identification Using Distributed Acoustic Measurement Networks

☆武内蓮，池見樹（近畿大院），原田和典（岡山県立大），
菅原彬子，平栗靖浩（近畿大）

◆本研究では一般的なマイクロホンアレイを使用することでは困難な広範囲での音源探査が可能なシステムの開発を目指す。

◆これまでの研究で同システムは自由音場では 20mm 以内の誤差での推定が可能である。本報告では無響室内に反射面として、パーテーションを設置し(Fig.1)半自由空間の状態で音源位置推定を行い同システムの有効性の検討を行った。

◆壁１の位置に反射面を設置した場合は高い精度で推定することが確認されたが、壁２の位置では大きな誤差が生じた。(Table 1)この誤差を修正する方法の検討が今後の課題である。

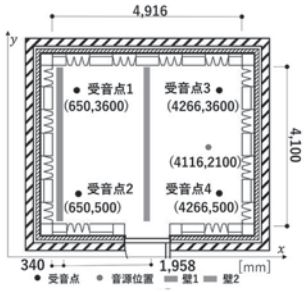


Fig.1:Layout drawing

単位は[mm]		x	y
音源位置		4,116	2,100
壁１	推定位置	4,088	2,080
	推定誤差	28	20
壁２	推定位置	5,179	2,123
	推定誤差	1,063	23

Table 1:Estimation result

1-6-7

1-6-7　振動情報を利用した ドローン衝突回避のための音源探知

Sound source detection for drone collision avoidance using acoustic vibration information

☆星野瑛彗（中央大院），△白石英孝，戸井武司（中央大）

◆ドローンが普及し，複数の機体が飛行すると機体同士の衝突の可能性が高く，衝突回避が必須である。そこで，ドローン活用の更なる発展のためには，長距離飛行と夜間飛行時を含めた衝突回避が求められる。

◆研究の目的は，ドローンが飛行中に発生させる音圧の高い動作音を利用した自身の機体（以下，自機）から衝突の脅威となる機体（以下，脅威機）の探知である。音源探知による衝突回避には，センサの小型適用が期待される。

◆自機のモータ付近に設置した加速度センサから取得する振動情報を利用して自機から発生する動作音を推定および抑制することで脅威機の動作音を推定する。

◆2 機のドローンの動作音が混在する Fig. 1 より,自機の動作音を抑制し,Fig. 2 のように音源探知に必要な脅威機の動作音が推定できた。

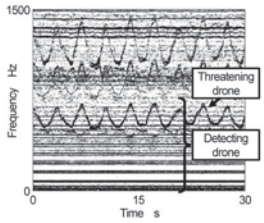


Fig. 1　Spectrogram of mixed flight data

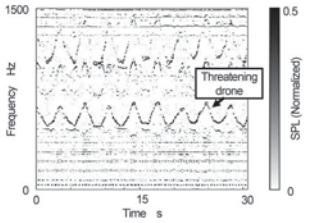


Fig. 2　Spectrogram after suppression

1-6-8

1-6-8　移動型ロボットの音響計測その 1

Sound measurements of autonomous robots. Part-1

☆増渕 優太（神奈川工科大），△有馬 祐杜，中島 康貴（リオン），廣江 正明（小林理研），△脇田 敏裕，上田 麻理（神奈川工科大）

◆人に近い場所で活躍する自律移動ロボットが増えると騒音が社会課題になり，測定に関する研究が求められる可能性がある．そこで本研究では，今後波及拡大が見込まれる移動型ロボットの音響計測方法確立のための第一段階として騒音測定を行った一例を紹介する．

◆測定対象はヴイストン製の megarover v2, v3 を神奈川工科大学で改造した車両で，モータ駆動で自律やリモート制御できるものである．

◆室内でロボットを通過させ，騒音を高さが異なる３本のマイクロホンで騒音測定を行い，床面の反射による干渉と距離減衰によるレベル差を比較検討した．

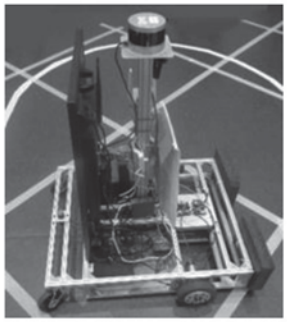


Fig.1 : Autonomous robots (megarover-V2)

1-6-9

1-6-9　非接触音響探査法を用いた欠陥探査に関する研究 −共振判定を用いた欠陥位置の推定精度向上に関する検討 2−

○中川裕，杉本恒美，杉本和子，上地樹（桐蔭横浜大　大学院工学研究科）

◆現在、我々は音波照射加振とレーザドップラ振動計を用いた非接触音響探査法により構造物等に生じる剥離や亀裂欠陥の探査について研究を行っている。

◆測定された信号に含まれる欠陥部の共振による信号とそれ以外のノイズを判別する共振判定処理を用いる事で欠陥位置の推定精度の向上に関する検討を行った。

◆測定対象は内部に剥離欠陥を模擬した 0.5 mm 厚の PET 樹脂板と空洞欠陥を模擬した発泡スチロールが配置されているコンクリート供試体を使用した。

◆Fig.1 は共振判定の有無による振動エネルギー分布である。共振判定を用いることで欠陥部以外の高い振動エネルギーを示していた箇所の改善が確認された。

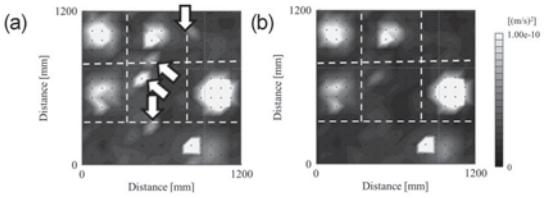


Fig.1 : Vibration energy distribution.
(a) Without resonance judgment, (b) With resonance judgment

1-6-10

1-6-10　受信信号に対する畳み込み積分による符号変調を用いた Lamb 波のパルス圧縮

Pulse Compression of Lamb Waves Using Code Modulation by Convolution of Received Signals

◎清水鏡介（日大院理工），大隅歩，伊藤洋一（日大理工）

◆空中超音波フェーズドアレイを用いた弾性波源走査法による非破壊検査について研究を行っている.

◆本報告では，受信信号に対する畳み込み積分による符号変調を用いた Lamb 波パルス圧縮方法を提案し，検証した.

◆Fig. 1 にパルス圧縮を適用する前の(50, 150)における時間波形を，Fig. 2 に提案手法によりパルス圧縮を適用した(50, 150)における時間波形を示す.

◆結果より，パルス圧縮の適用によって急峻な波面の立ち上がりが可視化されており，提案手法の有効性が確認された.

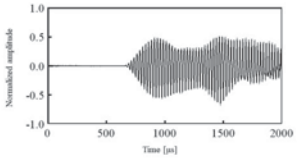


Fig. 1 Guided wave propagation waveform (Over all).

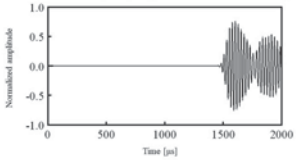


Fig. 2 Guided wave propagation waveform (After pulse compression).

1-6-11

1-6-11　高分子ゲルファントムの機械的バースト加振によるせん断波伝搬特性の周波数分散性

Frequency dispersion of shear wave propagation characteristics by mechanical burst excitation in the polyacrylamide gel phantom

☆齋藤慎一郎，△黒川要，△菅幹生，吉田憲司，
山口匡（千葉大），蜂屋弘之（東工大），平田慎之介（千葉大）

◆せん断波の伝搬波面を可視化すること，およびせん断波伝搬速度の周波数分散性を評価することを目的とした機械的バースト加振 Shear Wave Elastography (MBESWE) を提案している.

◆高分子ゲルで作成した粘弾性ファントムに 60, 90, 112.5, 150, 225 Hz のバースト波で機械的な加振を与え，せん断波伝搬波面を可視化した. (Fig1)

◆せん断波伝搬波面から位相勾配を求めることでせん断波伝搬速度の計測を行った. (Fig2)

◆MBESWE によりせん断波伝搬波面の可視化と周波数分散性の評価が可能.

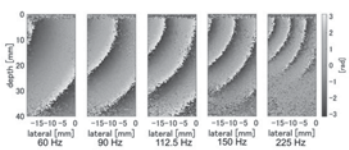


Fig1. Shear wave propagation wavefront at 60, 90, 112.5, 150 and 225 Hz

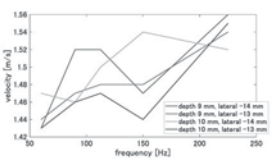


Fig2. Shear wave velocities at 60, 90, 112.5, 150, 150 and 225 Hz

1-6-12

1-6-12　強散乱媒質におけるアニュラアレイによる後方散乱特性解析の精度検証

Accuracy verification of backscattering coefficient analysis using annular array in strongly scattering tissue

☆崔廷宅（千葉大・工），八木風音，沓沢駿人，比賀大成（千葉大・院融合），
伊藤一陽（東京農工大），平田慎之介，吉田憲司，山口匡（千葉大・CFME）

◆弱散乱源が均質に分布している生体組織で安定的な後方散乱係数（BSC）評価が可能とされている平面波送受信を軸とし，アニュラアレイセンサによる送受信での BSC 評価の精度について比較検討した.

◆直径 5 μm と 10 μm の散乱球をそれぞれ体積分率 5%含む寒天ファントムについて，リニアアレイプローブ（平面波）とアニュラアレイセンサ（収束波）でエコーデータを取得し，リファレンスファントム法で BSC を評価した.

◆参照媒質と評価対象が同じケースでは，評価媒質の BSC が理論値に近似したが，参照媒質と評価媒質が異なるケースでは，周波数が高くなるにつれて理論値からの乖離が大きくなる傾向を示した.

◆特に，アニュラアレイセンサでは散乱減衰が BSC 評価に強く影響することが確認された.

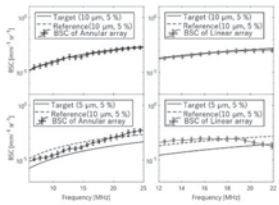


Fig.1 : BSC referenced by 10 μm, 5 % (left: Annular, right: linear)

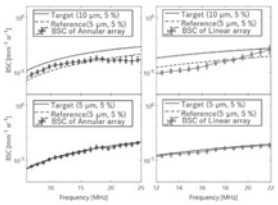


Fig. 2 : BSC referenced by 5 μm, 5 % (left: Annular, right: linear)

1-6-13

1-6-13　平面波送受信による多種散乱源の振幅包絡特性弁別評価の基礎検討

Fundamental evaluation of relationship between amplitude envelope statistic characteristic and muilt-component scatterer distribution in plane wave ultrasound

☆ZHANG TINGZHEN, 氏原裕貴, 比賀大成(千葉大･院融合), 大村眞朗, 長谷川英之(富山大･工), 平田慎之介, 吉田憲司, 山口匡(千葉大･CFME)

◆Amplitude envelope statistic with Double-Nakagami (DN) model was used to evaluate the relationship between scatterer distribution and echo signal statistic characteristics.

◆Echo data of agar-based phantoms consisted with two kinds of scatterer (diameters of 10 μm and 20 μm) were acquired by ultrasound equipment (Vantage 256, Verasonics) and a linear array probe (9L-D, GE Healthcare) in coherent plane-wave (CPW) method.

◆In CPW ultrasound imaging, DN model can distinguish the independent signals from different kinds of scatterer and evaluate fatty mixture ratio in early stage of fatty liver (volume fraction of 20 μm scatterer is less than 0.1%). (Fig.1)

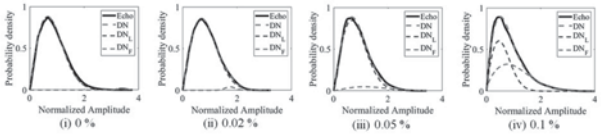


Fig.1 : Probability density function of original echo signal from four phantoms with different 20 μm scatterer volume fractions.

1-7-1

1-7-1　高周波数超音波凹面トランスデューサにおけるエッジ波に関する基礎検討

A basic investigation on edge wave of a concave transducer in high frequency ultrasound measurement

◎長岡亮, 長谷川英之(富山大), 村上舞, 磯辺悠斗(豊橋技科大), 大村眞朗(富山大), 穂積直裕(豊橋技科大), 小林和人(本多電子)

◆単一の凹面トランスデューサを用いた高周波数超音波計測において, エッジ波などの回折の影響をシミュレーション実験によって検討した結果を報告する. また, シミュレーション実験結果とワイヤーファントムを用いた実測結果とを比較した.

◆結果として, シミュレーション実験で得られた送信音圧に対してガウス窓を適用しない場合の点拡がり関数(Point Spread Function: PSF)と実測で得られたPSF との形状がよく類似していた. 一方で, エッジ波の影響を低減するためにガウス窓を適用した場合の PSF とは異なっていた. 以上の結果より, 凹面トランスデューサを用いた高周波数超音波計測においても, エッジ波などの回折の影響を受ける可能性が示唆された.

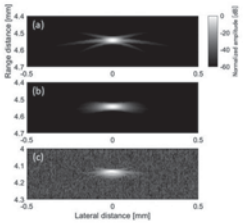


Fig.1 : Simulated PSFs obtained without Tx apodization (a) and with Tx apodization. (c) Measured PSF.

1-7-2

1-7-2　ラット肝臓内組織と音速の三次元評価の基礎検討

Basic study of three-dimensional evaluation of tissue and sound velocity in rat liver

☆本郷玄太(千葉大･工), 瀬戸駿(千葉大院･院融合), 田村和輝(浜松医科大), 吉田憲司, 平田慎之介, 山口匡(千葉大･CFME)

◆摘出臓器の病理像と音速像の対応関係を高分解能で三次元的に理解することを目的として, NASH 肝の薄切試料を対象に, 中心周波数 300 MHz の振動子を用いて音速を評価した.

◆三次元レジストレーションにより, 病理像と音速像で肝臓内のミクロな線維と脂肪滴の分布の対応関係をリンクさせて理解可能であることを確認した（Fig. 1）.

◆試料全体の音速の平均と標準偏差に比して, 線維部分の音速が速く, 標準偏差が小さいことから, 特定の組織を安定して評価できていると想定される（Fig. 2）.

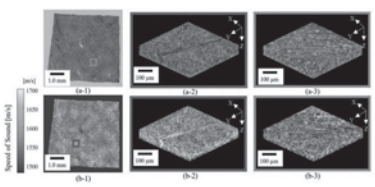


Fig .1 Three-dimensional images of pathology and speed of sound of NASH rat liver.

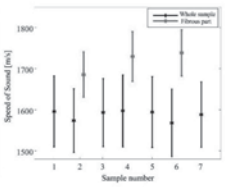


Fig. 2 Speed of sound of NASH rat livers.

1-7-3

1-7-3 リファレンスファントム法による減衰係数評価における誤差要因についての考察

Consideration of error factors in evaluation of attenuation coefficients by reference phantom method

☆森川紗帆(千葉大･工), 氏原裕貴, 沓沢駿人(千葉大･院融合)
平田慎之介, 吉田憲司, 山口匡(千葉大･CFME)

◆リニアアレイプローブを用いて, 送信フォーカス深度 40, 80 mm, 送信素子数 32, 64, 96, 受信素子数 192 の各条件での収束波と, 送受信素子数 192, ステアリング角±5°, ±10°, ±15°の平面波で均質ファントムのデータを取得し, 参照媒質と評価対象を組み合わせて, リファレンスファントム法で減衰係数を算出し比較した.

◆参照媒質と評価対象のフォーカス深度が異なる場合には評価値と真値の乖離が大きく, 等しい場合には真値に近い値を取ることが確認された. 開口が広いほどその傾向が強くなる結果となった（Fig. 1）.

◆収束波でもフォーカス深度が深く評価領域の音場特性が平面波に近しい場合では評価精度が高いが, 複雑な構造を有する対象での実用を想定すると, 音圧の空間分布がフラットな平面波での送受信が有効であることが示唆された.

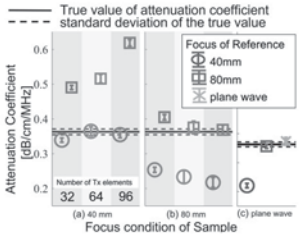


Fig.1 : Attenuation coefficient for multiple transmission conditions

1-7-4

講演取消

1-7-5

1-7-5　頸動脈内腔表面粗さの超音波 *in vivo* 計測結果の空間周波数解析

Spatial frequency analysis of surface roughness on carotid artery measured by ultrasound

☆山根綾太[1]，森翔平[2]，荒川元孝[1,2]，金井浩[2,1]

([1]東北大院・医工，[2]東北大院・工)

本研究グループは，動脈硬化症の極早期診断を目標に，頸動脈後壁の内膜中膜複合体の血管長軸方向変位を利用して，音速不均一領域の影響を受けずに頸動脈内腔表面の微小な粗さを超音波計測するための手法を開発してきた．健常者群の表面粗さの空間周波数スペクトルからの逸脱度により，動脈硬化症を検出できる可能性がある．そのための基礎検討として，本報告では 20 代健常者 3 名の表面粗さを本手法により計測し，健常者の表面粗さの空間周波数特性の傾向を確認した．その結果，いずれの被検者においても同様な傾向の空間周波数スペクトルが得られ，0~2 mm^{-1}の帯域に8~20 μmの高さ成分が見られた．

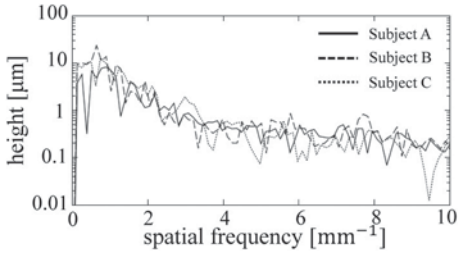


Fig.1 : Spatial frequency spectrum of surface roughness.
(Solid line: Subject A, Dashed line: Subject B, Dotted line: Subject C)

1-7-6

1-7-6　超音波加振における変位測定の安定化に関する検討

Stabilization in measurement of displacement generated by ultrasonic excitation

☆小松祐也[1]，森翔平[1]，荒川元孝[2,1]，金井浩[1,2]([1]東北大院・工)([2]東北大院・医工)

本研究グループでは，対象物に超音波を照射して局所的に加振し，その粘弾性特性を求める方法を検討してきた．本手法においては，超音波加振によって生じる動的変位を測定する必要があるが，発生変位は数μmと微小であるため安定に計測することが難しい．本報では，動的な微小変位を安定に計測することを目的に，ダイナミックレンジが広く，微弱な超音波信号を高精度に測定可能なネットワークアナライザを用いた動的変位計測法を検討した．

点集束超音波トランスデューサから対象物に超音波を照射する．このときの S パラメータ $S_{11}(f)$（f:周波数）をネットワークアナライザで計測し，信号処理により対象物からの1回目の超音波反射の周波数特性$S_{11}^{(1)}(f)$を求め，変位発生前に対する変位$\Delta L(t)$発生後の位相差$\Delta\angle S_{11}^{(1)}(f)$から変位$\Delta L(t)$を推定した．その結果，Figs. 1，2 に示すように，対象物をΔf = 60 Hzで振動させた場合，推定変位にも 60 Hz の周期性が見られ，レーザー変位計の測定値ともほとんど一致した．また，その他の周波数においても同様の結果を示した(Fig. 3)．これらの結果より，本変位計測法によって，周波数Δfで振動する対象物の動的変位を推定できるといえる．

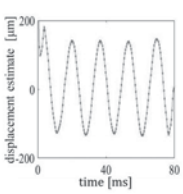


Fig. 1. displacement estimation result (Δf = 60 Hz).

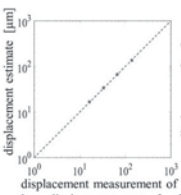


Fig. 2. Comparison with measured value of laser displacement meter (Δf = 60 Hz).

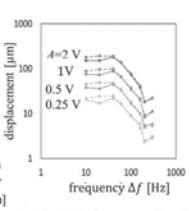


Fig. 3. Displacement estimation at Δf = 10 − 300 Hz. Solid line: estimation result, dotted line: laser displacement meter.

1-8-1

1-8-1　聴覚障害の臨床と聴覚心理学

Hearing Impairment Clinical and Auditory psychology

○竹内京子(順天堂大)，青木直史(北大)，荒井隆行(上智大)，△鈴木恵子(北里大)，世木秀明(千葉工大)，△秦若菜(北里大)，安啓一(筑波技術大)

◆言語聴覚士は，ことばのリハビリを行う職業である。その養成校では，聴覚心理学が必修科目となっている。

◆先行研究の現役言語聴覚士対象の調査によると，聴覚心理学は授業内容の記憶がほとんどなく，聴覚検査，補聴器・人工内耳の調整のような聴覚障害の臨床との関わりが分からないという声が多かった。

◆本発表では，ほぼ同じ内容の調査を別の集団の現役言語聴覚士に行った。その結果と比較し，報告する。

◆先行研究の結果と比較したところ，今回の調査では，聴覚障害の臨床と聴覚心理学の関係があると考える者が多かった。また，養成校時代，聴覚障害関連の科目がはっきりと好きという訳ではないが，嫌いではない集団であった。

◆しかしながら，今回の調査の参加者のような「聴覚心理学」興味がある者は，言語聴覚士全体のうちのほんの僅かでしかないだろう。それを考えると，今回の調査の参加者の回答が例外で，どちらかというと，前調査の内容の方が現状を反映している可能性が高い。

◆今回の調査の収穫として，「聴覚障害の臨床」に興味のある集団には，「聴覚心理学」との関わりが見えるという可能性が示唆された。

◆今後は，言語聴覚士への「聴覚心理学」の啓蒙活動を行うとともに，その活動が，今後の変化の糸口になることを期待したい。

1-8-2

1-8-2　発症から長期間経過した一側性難聴者が示す残響下のモノラル受聴における文章了解度の改善と方向性マスキング解除について

Those with longstanding unilateral hearing loss indicate some improvement in release from masking of reverberation and spatial release from masking under reverberation for speech recognition comparing monaural listeners with normal hearing

☆辻慎也，荒井隆行（上智大・理工）

◆片側耳は正常だが，対耳側に難聴がある状態を一側性難聴という．

◆一側性難聴により，以下３つの場面で困難な状況が生じる．

１）難聴側聴取，２）騒音下聴取，３）音源定位

◆本稿では一側性難聴者・健聴者のモノラル受聴・両耳聴を対象に，残響下での文章了解度テストを行い，次の結果を得た (Fig. 1).

１）残響下での音声聴取のために，一側性難聴者は両耳聴と比べて，大きな音声の音圧を必要とした．

２）健聴者のモノラル受聴と比べて，一側性難聴者は閾値が改善し，より大きな方向性マスキング解除の効果を示した．

３）発症から長期間経過した一側性難聴者のモノラル受聴におけるキューへの適応が示唆された．

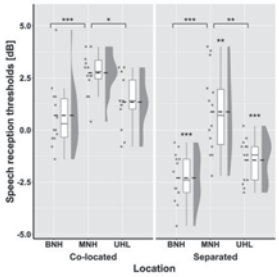


Fig. 1 The speech reception thresholds (SRT) for 3 hearing conditions; BNH (binaural normal haring, n=14), MNH (monaural normal hearing, n=14), and UHL (unilateral hearing loss, n=13). The solid and broken horizontal lines indicate the median and mean, respectively. SRT were measured in 2 conditions; target speech and masker were co-located (left) and separated (right). ***p < 0.001, **p < 0.01, *p < 0.05

1-8-3

1-8-3　終助詞「ね」のイントネーション ー「そうですね」聴覚的な女性性ー

Intonation of Sentence-Ending Particle Ne: Perceived Femininity of Soudesune

○越智景子（京大），△小林マヤ（昭和女子大）

◆背景と目的：本稿では、トランスジェンダー、とくにトランスジェンダー女性(Male-to-Female: MTF)の音声治療への応用を目指し、女性的な聴覚印象を持つ音声特徴を探索する

◆方法：『日本語日常会話コーパス』から抽出した「そうですね」の末尾の終助詞「ね」の韻律的特徴と、聴覚印象での女性性との関係を調べた

◆被験者は 33 名の大学生の女性で「そうですね」の音声 23 個を聴取して visual analogue scale (VAS)で女性性の度合いを回答した。

◆結果：「ね」の対数 F0 の中央値が相対的に高いほうが女性的である聴覚印象であることが示された。したがって、上昇調で急激に F0 が上昇するとより女性的に聞こえることが示唆される。また、推定値と実測測値との相関係数が男女二元論的質問項目で高く、男女二元論的質問評価項目のほうがより極端に声の男性性・女性性を韻律から判断していることが示唆された

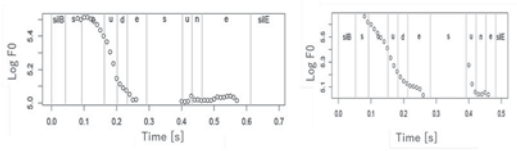


Fig. 1 Examples of log F0 patterns of utterances.

1-8-4

1-8-4　声帯ポリープの切除による発声への影響

Effect of removing a polyp on oscillation properties of the vocal fold physical model

☆植村泰佑（立命館大），△上田可奈子（立命館大），徳田功（立命館大）

◆声帯の酷使などが原因の声帯ポリープを模擬する物理モデルを構築した。

◆2 種類の大きさ（3mm と 5mm）、2 種類の硬さの合計 4 種類の物理モデルで実験を行った。

◆声帯ポリープモデル作成し、吹鳴実験を通して、各モデルの声門下圧、音声、高速度動画を計測し、その後、ポリープを切り取り、再度同じ計測を行った。

◆大きいポリープを切り取ることで最大声門面積が大きくなった。また、大きく硬いポリープを切り取ることでオンセット圧が低下した。

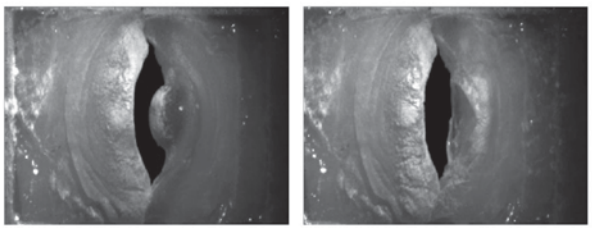


Fig.1 Physical model of the vocal fold having a polyp(left) and removing a polyp(right)

1-8-5

1-8-5　対話のことばとモダリティ

Modality of Dialogue Language

○市川　熹（千葉大／早大）

日常対話は、利用可能な感覚モダリティにより健聴者は聴覚言語（音声）で、また聴覚障碍者は視覚言語（手話）で、全盲ろうの方は触覚言語（指点字や触手話など）により行っている。多くの共通性を持ち、音声や手話の L1 の日常対話では精神的負担も軽い。夫々のモダリティに対応する「対話のことば」（メディア）の利点は活用され、欠点は補償する構造が存在する。

「対話のことば」が必要とする「要求条件」と、その条件を満たすための「実現条件」を「制約条件」として想定し、「対話のことば」の実態事例との関係から、多くの共通性を生む要因はなにか、「要求条件」はモダリティの違いにどの様に反映されているのかという例について、モダリティ横断的に考察を試みる。

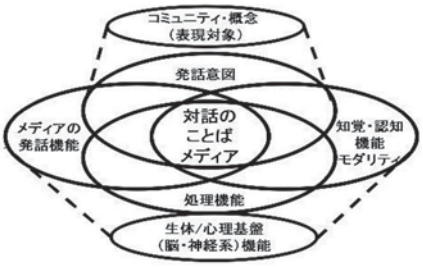


Fig.1: Areas of Establishment of Dialogue Language

1-8-6

1-8-6　認知言語学から見たコミュニケーションのマルチモーダル性

○ 岡久 太郎 (静岡大学)

◆本発表では、理論言語学のアプローチの１つである認知言語学の分野において、言語使用に見られるマルチモーダル性がどのように位置付けられ、いかなる研究が行われているかを論じる。

◆2000 年頃から認知言語学の分野において言語使用のマルチモーダル性に着目する研究が多く見られるようになった。このような認知言語学の展開は、認知言語学が当初から掲げている理論的前提である使用基盤モデルと記号的言語観に由来するものである。

◆認知言語学がマルチモーダル要素を言語的要素と同等のものとして想定するのであれば、語彙項目に類するマルチモーダル要素の分析に留まらず、言語研究において明らかにされてきた文法的な構文を言語以外のモダリティにおいても想定し、研究を行う必要がある。

◆認知言語学が言語の実際の使用を重視するのであれば、言語以外の１つのモダリティに着目したバイモーダル研究に留まらず、真のマルチモーダル研究を行う必要がある。

1-8-7

1-8-7　電話リレーサービスにおけるコミュニケーションのモダリティ

Example of an abstract sheet (16pt)

○井上正之（筑波技術大）

◆電話リレーサービスは，世界で広く利用されている「音声電話」のネットワークを聴覚障害者が利用可能にするサービスであり，聴覚障害者の社会参加を促進する上で非常に重要な意義を持っている。日本でも，2021 年 7 月から公的サービスとしての電話リレーサービスが開始され，徐々にユーザ数・利用が増大してきている。

◆電話リレーサービスは，音声ベースのモダリティに基づく一般の電話ユーザと，文字や動画など視覚ベースのモダリティに基づく聴覚障害者との間でメディア変換（文字～音声もしくは手話～音声）を行うことにより，一般の電話ユーザと聴覚障害者の間でのコミュニケーションを可能にしており，モダリティに大きく依存したサービスと言える。

◆本稿では，現状の電話リレーサービスを音声電話サービスと比較した機能的等価性の観点から見た現状と課題について，コミュニケーションのモダリティに着目して概観する。

1-8-8

1-8-8　体性感覚信号の音声知覚およびその調整への影響

Somatosensory contribution to speech perception and its recalibration

○伊藤貴之（GIPSA-lab, CNRS）

音声知覚処理は、音声知覚の運動理論に代表されるように発話生成過程の関与が古くから考えられている。しかしながら、発話生成過程には運動指令と体性感覚信号の両方が関与することから、どちらがどのように音声知覚への感覚運動統合に機能しているかはよくわかっていない。本発表では、体性感覚信号に着目し、運動情報に関連した体性感覚入力の音声知覚処理への影響および、機能的な役割について考える。ここでは、顔の皮膚伸長が発話運動情報の取得に貢献していることを元に開発された皮膚感覚刺激装置を用いて、いくつかの心理物理実験を行った。まず皮膚刺激により変調された体性感覚信号が音声知覚処理を変調すること母音同定課題および単語分割課題において示した。そして、発話運動学習および音声知覚学習課題を用いて、この聴覚―体性感覚連関が音声知覚表現の調整において重要であることを示した。これらの結果から、体性感覚信号は運動指令とは独立に音声知覚処理およびその調整に貢献していると考えられる。

1-8-9

1-8-9　日本語における促音の知覚とモダリティ

Perception of geminate consonants and modality in Japanese

○荒井隆行（上智大・理工）

◆日本語の促音：音響的に子音の持続時間が伸長
＋先行母音末尾にフォルマント遷移.

◆先行研究：聴覚刺激が/a/+pause+/ta/，視覚刺激が/atta/の場合，促音回答率が上昇（Arai *et al.*，JASA-EL，2017）.

◆本稿：聴覚刺激のみに加え，Fig. 1(a)の視覚刺激と聴覚刺激，あるいはFig. 1(b)の触覚刺激と聴覚刺激で同様の実験を行った別の先行研究に対して，新たな考察を議論.

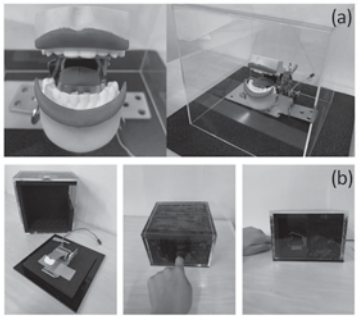


Fig. 1: (a) Physical models of upper/lower jaws and tongue used for visual stimuli; (b) the same robot-hand unit was used for tactile stimuli (Arai, INTERSPEECH, 2022).

1-8-10

1-8-10　話者の音声と表情に基づくマルチモーダル興味認識

Multimodal Interest Detection Based on Speaker's Voice and Facial Expression

○山田仰(ドコモ)

人対人のコミュニケーションの中で対話相手の「興味」を認識する研究はまだ数が少なく，特に日本人を対象とした大規模（100 名の話者等）でのマルチモーダルなモデル構築・評価の前例はない．本研究では興味認識のための大規模なマルチモーダルデータセットを構築し，機械学習モデルの構築と評価を行なった．

1-9-1

1-9-1　歌唱者の過剰な筋活動と聴取者の生理指標や評価への影響

The effect of excessive muscle activity in singers on evaluation and physiological indicators of listeners.

☆宮川桃，饗庭絵里子(電通大)

◆本研究は喉しめ声による発声の聴取時に聴取者の筋活動や評価に与える影響から新たな歌唱評価指標を開発することを目的とする．

◆プロ歌手２名，アマチュア４名に対して楽な発声と喉しめ声による歌唱聴取時の筋活動及び呼吸活動の差異を計測した．

◆聴取刺激による筋活動の差異に一貫性は見られなかった．

◆喉しめ声による歌唱は特に抑揚・声質の好み・聴き心地の三項目の評価結果に影響を与えることがわかった．

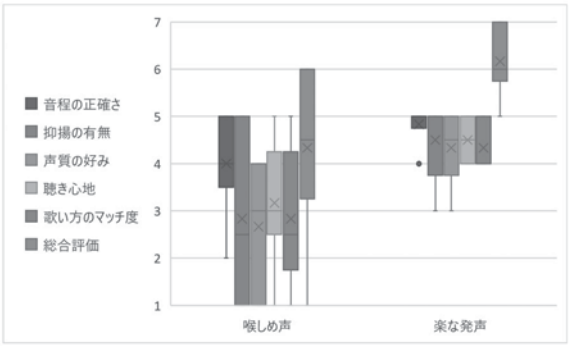


図１ 評価結果の差異

1-9-2

1-9-2　縞投影法を用いた不均一張力ドラムヘッドのチューニング可視化

Tuning visualization of a drumhead under non-uniform tension using Fourier transform profilometry

☆橋本涼汰(早大理工)，矢田部浩平(農工大)，及川靖広(早大理工)

◆実際のチューニング過程を想定した反復測定の提案

1．(1,1)モード形状の歪みから，張力の小さい領域を推定

2．推定領域のボルトを締めて再可視化

3．ボルトの再調整を繰り返し，張力分布の変化を追う

◆Fig. 1 各図はチューニング条件と２つの(1,1)モード

- 左図：数字が大きいほどボルトが強く締まっている
- 右図：縮退の解けた２つの(1,1)モード
- 均一に近づけられた条件とそうでないものが存在

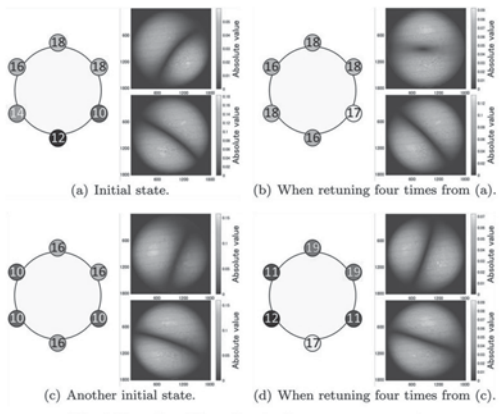


Fig.1 Results of iterative tuning measurement.

1-9-3

1-9-3　音叉の振動継続時間における温度の影響：共振周波数の異なる音叉間の比較

Effects of the temperature on the vibration duration of tuning forks: Comparison between tuning forks with different resonance frequencies

☆野水響太，大塚翔，中川誠司（千葉大）

◆音叉には，共振周波数の正確さに加えて，基音の十分な持続，および基音とともに発生する高調波成分の速やかな減衰が求められる．

◆音叉の製作工程においては周波数の調整はなされるものの，音の持続の評価は行われていない．前報で対象としたアルミニウム，ステンレス製音叉に炭素鋼製音叉を加え，それぞれの材料において基音周波数が異なる２種の音叉の振動継続時間に対する温度の影響を比較した．

◆基音の振動継続時間への温度の影響は材料によって異なり，アルミニウム，ステンレス製音叉においては多くの個体で温度の上昇によって短縮したが，炭素鋼製音叉では延長した (Fig.1)．また，アルミニウム，ステンレス製音叉については，温度の影響に個体差が見られた．

◆材料間で温度の影響が大きく異なったことから，材料間における比熱，熱膨張率といった温度関連パラメータの差異が振動継続時間に大きな影響を及ぼすことが示唆される．

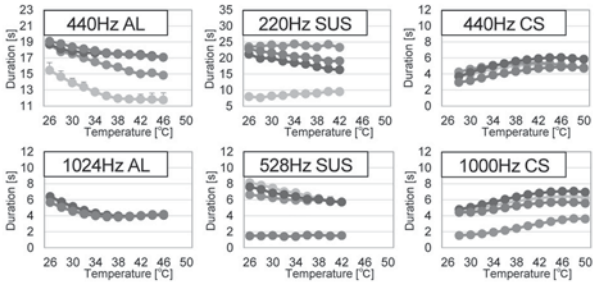


Fig.1 Vibration durations of fundamental tone as functions of the temperature.

1-9-4

1-9-4　インクルーシブ教育に資する打楽器演奏システムの開発 ～卓上型打楽器演奏ロボットと演奏支援装置の開発～

Development of interactive robotic percussion systems contributing to inclusive education : development of desktop percussion playing robots and their performance supporting equipment

○庄司英一（福井大院工）

本研究は、特別支援が必要な児童生徒が手軽に打楽器演奏を楽しめる手段を実現し、インクルーシブ教育を工学的に支援する打楽器演奏システムの開発を目的とする。この打楽器演奏システムは、Teacher2(T2)としての打楽器演奏ロボットと、インクルーシブな観点から発案される個の特性に応じた演奏支援装置やソフトウェアから成る。今回の発表では、先ずこの打楽器演奏システムについて紹介する。本構想では様々な演奏機器の創造がカギとなるが、今回の発表では、開発中の機器の中からの話題として、音楽室にある打楽器を脱着式で取り付ける機構を備えた卓上型打楽器演奏ロボット(Fig.1A)と連携して使う重度肢体不自由者（脳性麻痺を含む）が打楽器演奏を楽しめる演奏支援装置(Fig.1B)について、それらのパフォーマンスについて紹介する予定である。

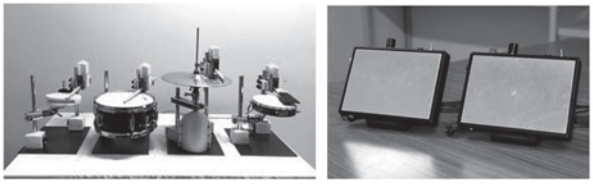


Fig.1 Desktop interactive robotic percussion systems (IROPS-8):A(left), and performance supporting equipment (built-in rhythm performance system):B(right).

1-9-5

1-9-5　モーションキャプチャを用いた複数のチェンバロ奏者の演奏動作解析*

*Performance Analysis of Multiple Harpsichord Players Using Motion Capture

○三戸勇気（日大・芸術），松崎桃子，山口達也（日大・理工），
駒澤大介，川上央（日大・芸術），三浦雅展（国立音大），篠田之孝（日大・理工）

我々は，モーションキャプチャシステムを用いた鍵盤楽器の演奏動作の計測ならびに解析を行ってきた．複数のチェンバロ奏者の演奏動作を取得することで傾向が把握できるのではないかと考え，チェンバロ奏者6名の感情表現に対する演奏動作を計測し，その演奏動作の解析を行った．今回，チェンバロ奏者の左右の手首，肘，肩の関節部位の軌跡長の平均化により，感情表現に対するチェンバロ演奏動作について検討を行った．Fig.1 は感情平面上に，6 名の演奏者の各関節に取り付けたマーカの単位時間当たりの移動量の平均値の結果を表示している．

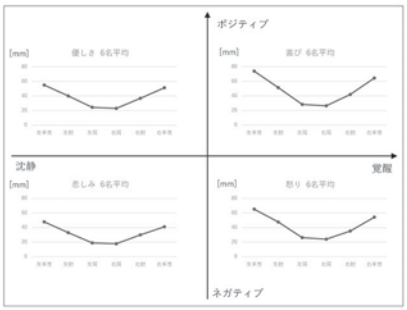


Fig. 1 Trajectory length per unit time during performance on 2D plane of emotion

1-9-6

1-9-6　構成音のピッチの有無・ラウドネス・減衰がリズム同期に与える影響

Effects of pitch, loudness, and attenuation of the component tone on rhythmic synchronization

☆山中律輝，松井淑恵（豊橋技科大）

◆リズム判断の手がかりが低音にあるかどうか，またその手がかりの強さが個々の音の減衰時間とラウドネスバランスによって変化するかを確かめるタッピング実験を行った．また，これらの効果が明確なピッチを持たない音でも生じるかをバンドノイズを用いて確かめた．

◆高音域と低音域のピアノ音 2 音を一定間隔で同時呈示し，時折，片方のタイミングを逸脱させた．この刺激に対してタッピングさせることで逸脱に対する感度，すなわちリズム手がかりの強さを計測した．

◆減衰時間を短くする事でリズム判断の手がかりとなりやすくなる効果が低音にのみ見られた．

◆低音が高音に対して 9 dB 減衰したとしても低音にはリズム知覚を生む効果があった．それに対して高音が低音に対して 3 dB 減衰した場合には高音のリズム知覚を生む効果は小さくなっていた．

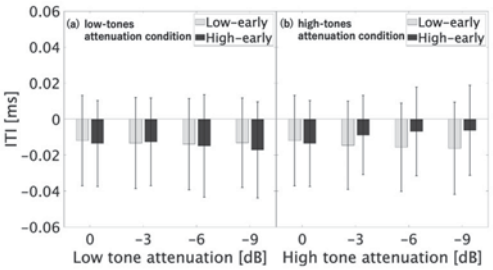


Fig.1 : ITI mean and standard deviation for each attenuation condition

1-9-7

1-9-7　音楽的期待に基づく和音進行聴取時の印象評価と脳波との関連

Relation between impression and ERP from expectations for chord progressions

☆末崎真実（電通大），高橋範行（愛知県立大），森本智志（慶応大），饗庭絵里子（電通大）

◆多くの西洋音楽は，機能和声のような規則に従って構成されているため，和音の予測がおおよそ可能となり，音楽的な期待が生じる．期待から逸脱した音を聴取すると mismatch negativity(MMN)と呼ばれる脳波の振幅等に変化が現れることが報告されている。

◆しかしながら，音楽的期待には，典型的な和音進行から逸脱したことによる「違和感」だけでなく，予想外の「良い意外性」をもつ進行も存在する。

◆本研究では，「違和感」や「良い意外性」に対する主観評価が極端であった和音進行刺激に対して予備的な脳波計測を行い，主観評価との対応関係を検証した。

◆その結果，音楽理論と脳波との関連性は示されたが，主観評価との関連性はあまり見られなかった。

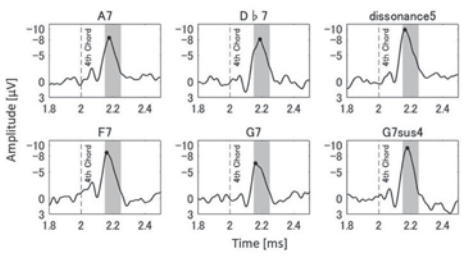


Fig.1: ERP Waveform for various chord progression

1-9-8

1-9-8　和音の印象特徴量の計算モデルの再検討および拡張

Reexamination and Extension of the Chord Impression Feature Calculation Model.

☆千葉湧河(筑波大), 寺澤洋子(筑波大), 松井 淑恵(豊橋技科大)

本研究では，和音進行や楽曲の分析，印象解析への貢献を目的とし，それぞれの和音が持つ感性的特徴を明らかにする．先行研究として，単一の和音の感性的特徴を，その物理的性質から理論的に算出する和音性モデルが挙げられるが，このモデルは理論的な裏付けがあるものの実証実験による検証が不十分であるため，本研究では和音の聴取実験を行い印象評価とモデルの対応関係を探った．また，3 音からなる和音（三和音）を対象に定義された和音性モデルを，4 音からなる和音（四和音）へ拡張することを試みた．

聴取実験による検証をベースに，実際の和音の印象評価と和音性モデルには対応関係があるということを明らかにした．また四和音の印象評価は三和音の印象評価と同様の構造があることも判明したため，四和音についての和音性モデルも同様に定義できた．今後，より音数の多い和音へのさらなる拡張が期待できる．和音の和音性モデル算出の際に，音の組み合わせごとに重みを変えたり，人の聴覚特性や個人性モデルを考慮することでさらに有用なモデルへと発展させられるだろう．

1-9-9

1-9-9　演奏内容と演奏環境が演奏音の印象に与える影響に関する検証

Verification of the influence of performance content and performance environment to the impression of performance sound

◎桶本まどか（国立音大・院），三浦雅展（国立音大）

◆演奏音の印象には演奏内容と演奏環境のいずれもが重要と言えるものの，演奏内容の印象と演奏環境に起因する印象の比較はこれまで検討されていない．本報告では演奏内容の印象と演奏環境に起因する印象を SD 法により得られた因子スコアによって比較し，演奏空間に起因する印象の重要性を調査する

◆印象評価においては，26 通りの 7 段階の両端尺度による SD 法（Semantic Differential Method）を用いた．

◆得られた印象評価結果の内，19 通りの評価軸を用いた因子分析により，第 1 因子に鮮明因子，第 2 因子に明暗因子，第 3 因子に厚み因子を得た．なお，因子寄与率は第 1 因子が 0.45，第 2 因子が 0.24，第 3 因子が 0.10，累積寄与率は 0.79 であることより，3 つの因子で模擬演奏音の印象が説明可能であることが示唆された．

◆第 1 因子と第 2 因子における各模擬演奏音の因子スコアより，第 1 因子は演奏環境における反響の長さ，第 2 因子は演奏内容として用いた楽器のピッチとの関係が示唆され，さらに，第 1 因子の因子寄与率が第 2 因子の因子寄与率より高いことから，演奏内容に比べ演奏環境の方が演奏の印象に与える影響が大きい可能性が示唆された．

1-9-10

1-9-10　波動方程式に基づく深層学習モデルによる金管楽器の音響解析

Acoustic analysis of brass instruments by deep learning model based on wave equation

○横田 和哉，藤本 一毅，阿部 雅二朗（長岡技術科学大学）

◆楽器の設計において、音響数値解析は重要な指標の一つである。

◆近年の機械学習の進歩により、ニューラルネットワークを用いて数値解析を行う手法が提案されている。その一つである Physics-Informed Neural Networks (PINNs) は、支配方程式による拘束を損失関数に導入することで、物理的に満足する解を得る手法である。

◆非線形な方程式にも適応可能であり、ニューラルネットワークを用いる手法であることから、最適設計などの逆問題に応用が期待される。

◆本稿では、金管楽器の音響解析への応用を目指し、基礎的検討として、1 次元波動方程式に基づく PINNs を構築した。単純化した吹奏条件下での解析を行い、波動方程式の解を計算可能であることを示した。

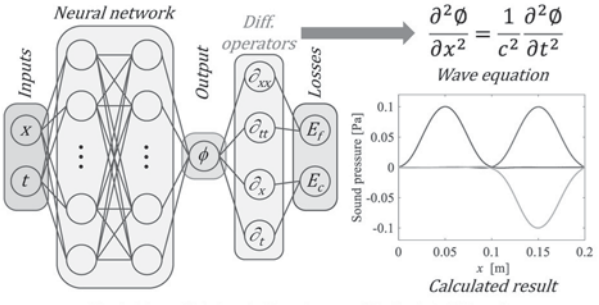


Fig.1 : Neural Network Structure and Calculated Result.

1-9-11

講演取消

1-9-12

1-9-12 尺八による奏法の違いによる音響特徴量の変化

Changes in Acoustic Features Due to Differences in How to Play the Shakuhachi

○佐々木 柊哉, 石垣優弥, 横山真男(明星大)

◆尺八は奏法を変化させることで音高を変化させたり、ビブラートをかけることができる。メリ/カリは音高が増減する際にどれだけ音が変化しているのか、ユリはビブラートがかかる周期と強さがどの程度か、といった様なことは明らかになっていない。本研究ではメリ/カリではピッチと倍音の変化、ユリではビブラートの周期や強さについてスペクトログラム解析を行い、どれだけ変化したのかを定量的に分析した。

◆本研究ではあらかじめ録音したデータを Audacity と Praat を用いて解析し、メリ/カリ/ユリの音響的特徴の分析を行う。また最近作られたアルミニウム製のメタル尺八でも同じ条件で分析し、竹尺八との比較も行い、メタル尺八の音響特性を理解する。

◆奏法によってスペクトログラム、パワースペクトルに違いが見られ、メリ/カリ/ユリの音響的特徴を捉えることができた。また、メタル尺八と竹尺八との違いも見られ、メタル尺八の音響特性も捉えることができた。

1-9-13

1-9-13 湾曲振動板からの音の放射分析

Sound radiation analysis from curved diaphragm

○半田帆峻, 石垣優弥, 横山真男（明星大）

◆通常、スピーカーの振動板は円錐形が一般的であるが、ミライスピーカーという振動板を湾曲させて音を放射しているものもある。湾曲させた状態の振動板をスピーカーに取り付けて音を放射した場合、湾曲の有無による放射音への影響は実際に生じるのか検証を行った。

◆市販のスピーカーにプラスチックや金属の振動板を取り付け、音の放射を行い、音響的特徴である指向性と音量について焦点を当て、解析を行った。

◆振動板を湾曲させることで基音と高周波数時の倍音の音圧が上昇し、スピーカーの放射の音をより大きく感じられるのではないかと考察する。

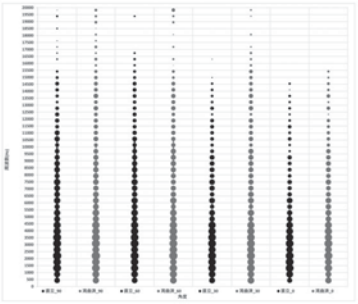


Fig.1：直立と湾曲済での角度による音圧の比較

1-9-14

1-9-14 箏爪の素材と奏法による箏の音色の音響分析

Acoustic analysis of koto tone by koto pick material and technique

○大堀航輝, 石垣優弥, 横山真男(明星大)

◆箏爪の素材によって箏の音色が違うと言われているが、どのように違うのかを検証した事例は少ない。本研究は、スペクトル解析やアンケート等により、箏爪の素材や奏法による音色の違いの検証を行った。

◆実験は、素材ごとに A4 の音高で単音を録音しスペクトル解析、余韻の長さの算出を行った。解析結果と音色の印象についてのアンケートの結果を比較し素材ごとの特徴を考察した。

◆箏爪の素材、奏法ごとにパワースペクトルや余韻に違いがあり、音色が違う原因のいくつかを明らかにすることができた。素材の硬さが余韻に影響し、奏法は音の立ち上がりに影響すると思われる。

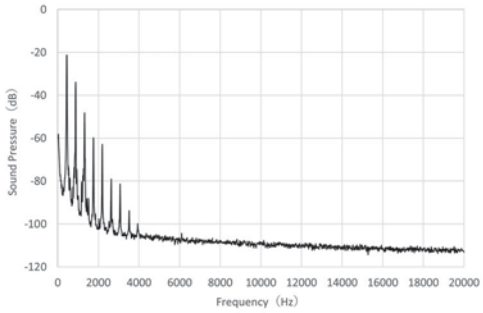


Fig.1：Power spectrum of ivory koto pick.
0.7 seconds from the timing of playing.

1-9-15

1-9-15 ドラムパターンを用いた楽曲ジャンル分類

Music Genre Classification using Drum Patterns

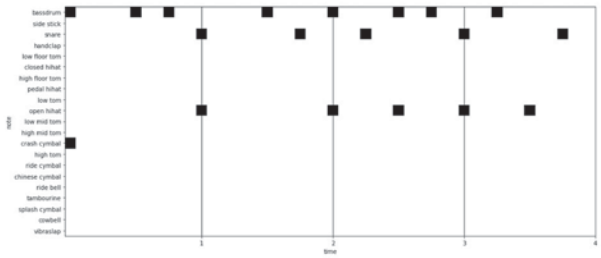
○荒　尭志、青木　直史、土橋　宜典(北大)

◆この研究では、ドラムのみの楽曲データを用いて、それらの楽曲ジャンル(ロック、ジャズなど)を分類することが目標である。

◆楽曲において、リズムを司る役割を持つドラムは楽器として占める重要度が高いと考えられ、ドラムのみを用いても他の楽器を含む場合と変わらずに分類を行えるのではないかと考えた。

◆この研究によって、楽曲データ全体だけではなく特定の楽器だけでも分類が行え、また楽曲検索キーとして活用が行える可能性を示す。

◆分類はニューラルネットワークを用い、訓練データには下図のような縦軸を楽器の種類、横軸を時間とした画像(ドラムロール画像)を用いた。



図：分類に用いたドラムロール画像の例

1-9-16

1-9-16 打撃音の時間スケールを保持したタイムストレッチング

Improved time stretching method preserving time scale of percussive sound

◎赤石夏輝（早大理工），矢田部浩平（農工大），及川靖広（早大理工）

タイムストレッチング
＝ ピッチを変えずに時間スケールを伸長する処理

位相ボコーダ（PV）を用いたタイムストレッチングの課題
＝ ドラムなどの**打撃音の劣化**

ストレッチ後の打撃音の時間スケールを
ストレッチ前のまま保持するための前処理を提案

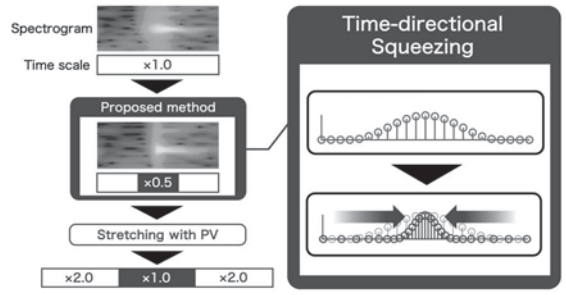


Fig.1: Overview of proposed method

1-9-17

1-9-17 禁則違反回数の少ない和声自動生成のためのデータ拡張手法の比較検討

Comparison of data augmentation methods for harmony generation with fewer parallel errors

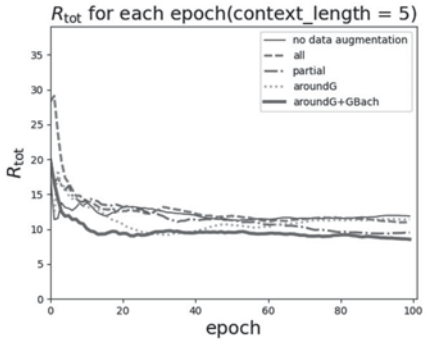
○唐津裕貴，深山覚（産総研）

◆目的：深層学習によるコラール自動生成結果の品質向上

◆課題：和声学の初歩的な禁則の違反回数が多い
- 初歩的な禁則：連続　並達など
- 禁則違反を減らすための学習データ数の不足

◆方法：データ拡張手法の検討
- 移調(all, partial, aroundG)
- 自動生成データ(aroundG+GBach)

◆結果：コラールの調性に合わせた自動生成データによる拡張が有効
- R_tot (epoch)：禁則違反音符数の割合(生成 epoch までの平均値)



1-9-18

1-9-18 和音推定のためのスパース時間周波数表現のオンライン推定

Online estimation of sparse time-frequency representation for chord estimation.

◎伊藤陸人，赤石夏輝（早大理工），矢田部浩平（農工大），及川靖広（早大理工）

- 目的：
  FifthNet をリアルタイム推定に応用できる形にする
- FifthNet：
  和音推定 DNN の state-of-the-art
  入力特徴量の計算に再割り当て法を用いる
  →入力の全時間分の情報を必要とするためリアルタイム化できない
- 提案法：
  Synchrosqueezing を用いたスパース周波数表現を利用
  → 1 つの時間セグメントのみを用いてスパースな表現が得られる
  → FifthNet の**リアルタイム化**への応用が期待できる

＋

再割り当て法を用いたものよりも和音推定に有効だった

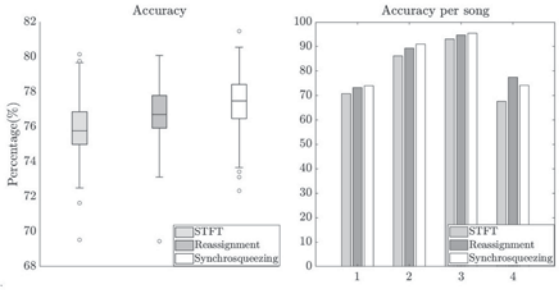


Fig1: Comparison of accuracy between conventional and proposed method

1-9-19

1-9-19 ボーカル音源を考慮した楽曲のコード認識の検討

A study of chord recognition for songs considering singing voices

☆北堀和佳，小坂哲夫（山形大）

◆従来、楽曲のコード認識においては，ボーカルや伴奏などが混合された音源（MIX 音源）を認識対象とするのが一般的であった．

◆一方，単一楽器やボーカルでは同時に一音しか出力されないため，ある時間幅の連続的なメロディから分散和音（アルペジオなど）を抽出する必要がある．以上より，同時和音と分散和音では異なる識別機構が必要と考えられる．

◆本研究では，同時和音と分散和音両方を学習データとするマルチコンディション学習により両者の認識を可能とするシステムの構築を試みた．

◆従来法ではボーカル音源に対しては低認識率だったが（48.33%），提案法により 71.27%まで向上した．また MIX 音源に対しても従来法と同等の認識率が得られることがわかった．

Table.1：Recognition results for each model（%）. (MIX：original sources, INST：instruments sources, VOCAL：vocal sources)

train_data＼test_data	MIX	INST	VOCAL
MIX	78.4	77.47	48.33
INST	78.35	**78.39**	36.21
VOCAL	76.55	75.79	70.44
MIX+VOCAL(proposal)	**78.93**	77.05	**71.27**

1-9-20

1-9-20　CNMF を用いたサウンドコラージュの楽器性の向上

Improvement of instrumentality for sound collage using CNMF

☆宮口聡良，小坂直敏（東京電機大）

◆デジタル音エフェクトの一つであるサウンドコラージュの合成法について、新たに最適な推定法が報告された CNMF アルゴリズムを使用してより楽器性の向上を目指した。

◆サウンドコラージュは、ある音波形（目的音）を別の音波形（要素音）を用いて再合成するエフェクトのことである。

例）セミの鳴き声をバイオリンで表現する。

◆ 主観評価による結果を示す(抜粋)

◆ 合成手法による平均評価値を再現性と楽器性で比較した結果を下図に示す。

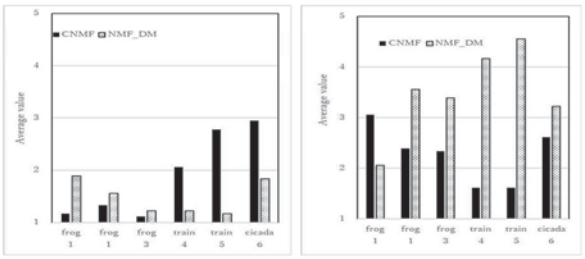


Fig1.　各条件の再現性(左)及び楽器性(右)に対する平均評価値

- 再現性はどれだけ目的音が知覚できるか。
- 楽器性はどれだけ要素音らしさを知覚できるか。

1-12-1

講演取消

1-12-2

1-12-2　BoSC システムを用いたオーディオルームの聴感評価 ―拡散体の違いによる聴き分けの検討―

Auditory evaluation of an audio room using the BoSC system
-Experimental study on distinguish between diffuser sounds-

☆伊藤幸一　伊勢史郎（東京電機大）

◆本研究では，BoSC システムを用いてオーディオルームの再現を行いその聴感評価を行った．拡散体の設置条件の違いによる聴き分けが可能かについて検討を行った．

◆全壁面に柱状拡散体がある場合（AGS），カーテンで全ての AGS を覆った場合（DRY），DRY の条件で簡易型拡散体（SYLVAN）を２台設置した場合（SYL）の３つの条件で実験を行った．

◆逆システム法として打ち切り特異値分解法（実験１）と正則化パラメータ法（実験２）について検討した．

◆被験者 14 名について聴感実験を行った結果，正則化パラメータ法による逆システムを用いた場合には総合的な判断を必要とする「音質の良さ」に関して AGS と SYL は DRY に比較して有意に評価が高いことが示された（図１右）．

Fig.1 : Results of t-test about sound quality

1-12-3

1-12-3　レンダリング条件によるオブジェクトベース音響コンテンツのラウドネスレベル

Loudness levels of object-based audio contents depending on rendering conditions

◎岩崎泰士，久保弘樹，大出訓史（NHK）

◆異なるレンダリング条件を用いて、オブジェクトベース音響（OBA）コンテンツのラウドネスレベルの測定実験と主観評価実験を行った。

◆31 個の音源を用いた測定実験の結果，スピーカ配置（6 種類）とレンダラ（4 種類）によるラウドネスレベルの差は，それぞれ最大 3.5 LU（平均 0.6 – 1.2 LU）と最大 3.5 LU（平均 0.1 – 1.1 LU）であった。

◆主観評価実験の結果，測定値（ラウドネスレベルの差）と主観値（基準音と評価音の大きさを合わせた調節値）の誤差は最大 2 – 4 LU（平均 0.8 – 2.9 LU）程度であった。同一コンテンツに対するスピーカ配置とレンダラによる測定値と主観値の誤差はそれぞれ最大 3 – 4 LU（平均 1.2 – 1.5 LU）と最大 4 LU（平均 1.1 – 1.3 LU）程度であった。

◆OBA コンテンツのラウドネスレベルは，レンダリング条件によっては最大 3.5 LU 程度の差が生じるが，聴感上の誤差は従来のチャンネルベース音響と同程度であることが分かった。

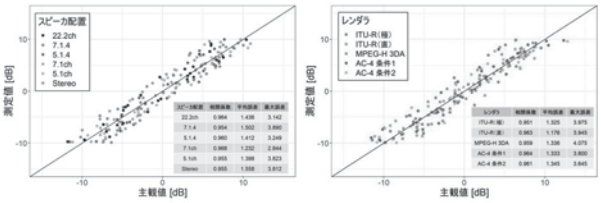


Fig.1 : Relations between Subjective (Horizontal Axis) and Measured (Vertical Axis) Loudness Values

1-12-4

1-12-4　異なる残響量が混在するリズム再生時における空間同一性知覚の許容範囲の検討

Investigation of the acceptable range of spatial identity perception during the playback of rhythmic patterns with mixed different DR energy ratio

○伊藤彰教(東京藝術大／東京工科大)，亀川徹，丸井淳史(東京藝術大)

◆高域・低域の周波数帯を組み合わせた、打楽器二重奏を模したリズムパターンを音楽刺激として、異なる残響量が混在する際の空間同一性知覚につき、ロジスティック回帰分析を用いて分析した

◆高域パートではDR 比増加部に多くの統計的有意性が認められ、空間同一性知覚の推定許容範囲DR 比はmean=2.9dB, SD=1.2 という結果を得た。

◆低域パートではDR 比減少部に多くの統計的有意性が認められ、空間同一性知覚の推定許容範囲DR 比はmean=-4.7dB, SD=1.7 となった。

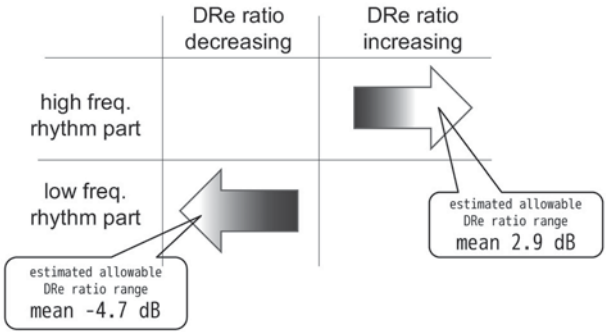


Fig.1: Estimated allowable DRe ratio range for special identity perception with mixed DRe ratio

1-12-5

1-12-5　3D オーディオの収音マイクロホンはどれくらい減らせるか？

How much can the microphones for 3D audio recording be reduced?

○亀川徹，丸井淳史(東京芸大)

◆3Dオーディオのひとつである22.2 マルチチャンネル音響の収音をより簡便にするため，MS 方式と直接音と拡散音を分離する手法を用いて 9 あるいは７本のマイクロホンから 20 チャンネル(ch)にアップミックスする手法を提案した。

◆聴取実験の結果から，提案手法は 20ch の原音と比較して違いが少ないと評価された。また，上層のマイクは前方よりも拡散性の高い後方のマイクを使用した方が，より効果的であることが示された。

◆ホールの残響などの拡散音の収音には中層４ch，上層４または 2ch の計９（あるいは７）ch で収音したものを変換することで 20ch の収音と同等の効果が得られることが期待できる。

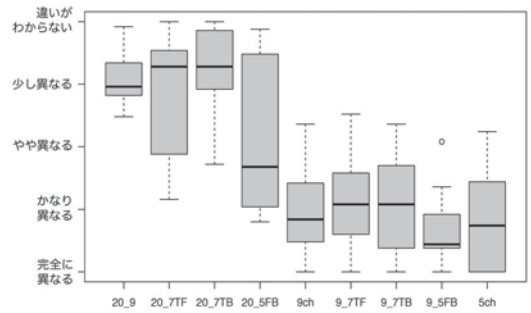


Fig.1:Results of the listening experiment.

1-12-6

1-12-6　遠隔合奏を実現する SYNCROOM とその技術

SYNCROOM and its technology for remote ensembles

○原貴洋(ヤマハ)

◆ヤマハ株式会社は，遠隔地間の音楽合奏を実現する SYNCROOM というアプリケーションを提供している。

◆SYNCROOM は，インターネット回線を介し，極力少ない遅延で双方向に音声データをやり取りすることで，遠隔地間での音楽合奏を可能としている。

◆Covid-19 の影響により，さまざまな音楽活動への制約が発生した。そのような状況下において，多くの音楽演奏者に SYNCROOM を活用いただいた。

◆本講演では，SYNCROOM の概要と利用環境(Fig.1)，開発に至る経緯，さらにその技術要素について紹介する。

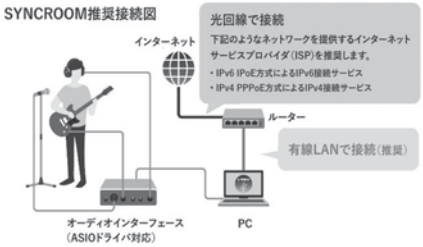


Fig.1 A recommended connection diagram for SYNCROOM

1-12-7

1-12-7　ハイレゾ音響信号のライブ中継を実現する高音質コーデックシステム

Audio CODEC systems for live transmission of high-resolution audio

○野口賢一（NTT）

◆2010 年代中頃より，ハイレゾオーディオが注目され，IP 通信環境の拡充とともに，ライブビューイングや放送・配信におけるライブ中継の高音質化が可能となった.

◆安定的なライブ中継伝送には，ハードウェアコーデック装置による音響符号化が求められ，音響符号化は，音質を犠牲することのないロスレス符号化が有効である.

◆ハイレゾ音響信号にも対応し，高音質なライブ中継伝送を可能とするコーデック装置の開発と活用事例について紹介する.

Fig.1:An example of Audio CODEC device

1-12-8

1-12-8　高音質インターネット動画配信システム「Live Extreme」

"Live Extreme" – a live video streaming system with the highest audio quality ever in the industry

○大石耕史（コルグ）

◆最大 4K の高解像度映像とともに、高音質なロスレス・オーディオをはじめ、さらに高品位なハイレゾ・オーディオをライブ配信できるインターネット動画配信システム

◆主な特長

- ソフトウェア・ベースのライブ・エンコーダー
  - ✧ 外部 ASIO オーディオ・インターフェイスからの入力
  - ✧ オーディオ・クロックを軸とした映像との同期
  - ✧ HLS / MPEG-DASH ダイレクト・アップロード
- ロスレス／ハイレゾ音声への対応
  - ✧ FLAC および Apple Lossless コーデックを採用
  - ✧ 最大 PCM 384kHz/24bit, DSD 5.6448MHz の配信が可能
  - ✧ 最大 7.1ch のサラウンド音声や副音声の配信に対応
- ウェブ・ブラウザでの再生に対応
- 60 公演を超えるライブやイベントでの配信実績
  - ✧ スマート・ファクトリーなど音楽用途以外の活用事例も

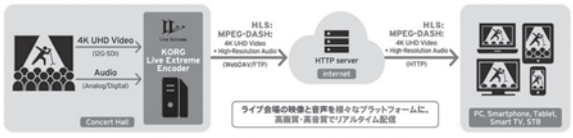


Fig.1: Live Extremeシステム概略図

1-12-9

1-12-9　（招待講演）WOWOW のハイレゾ・3D オーディオ配信の試み

Trials of High-Resolution and 3D Audio Distribution by WOWOW

○入交英雄（株式会社　WOWOW）

◆高臨場感音声はイマーシブオーディオとしてサービス化が始まった。

◆しかしパッケージメディアを除くと、圧縮音声サービスがほとんどであり、高音質についての議論は後回しになっている。

◆WOWOW では、音質を担保した 3D オーディオ制作手法を工夫すると共に、テレビ業界において、なおざりになっている音声の高度化について検討を進め、AURO 3D を用いたハイレゾ対応のωプレイヤーを試作し、コンセプト検証（PoC）を実施した。

◆その結果、３Dオーディオにおいては、ヘッドホンよりもスピーカ再生の顧客体験度が高いこと、ハイレゾマスター音源と通常ロスレスの音質が聴き分けられていることなどが判った。

◆また高音質、より高臨場感の音声に対するニーズがあることが確かめられた。

◆本発表では、高臨場感、高音質について、技術的要件に加え、PoC 参加者のアンケート調査を盛り込んで問題提起したい。

1-12-10

1-12-10　名古屋芸術大学学生によるオーケストラコンサートのバイノーラルライブ配信について

Binaural live streaming of orchestral concerts by students

○長江和哉（名古屋芸大）

◆名古屋芸術大学サウンドメディア・コンポジションコースはこれまで 15 年にわたり、本学コンサートのライブ録音を行ってきた。2020 年、コロナ禍となり、この「録音」に「映像」を組み合わせることによりライブ配信が行えるのではと考え、2020 年 10 月より行うこととなった。そして、このライブ配信に「何かもう一つ新しいことを」と考えたのがバイノーラルライブ配信であった。本来であればマルチチャンネルの 3D 音声でライブ配信を行いたいが、まだ一般的ではない。そこで、本学では、3D Audio の制作ツールである、Dolby Atmos Renderer、360 Reality Audio Creative Suite、AURO-3D Creative Tools Suite などのヘッドフォン出力から得られたバイノーラル音声を用いてライブ配信を行った。本発表では、どのようにバイノーラル処理やマイクアレンジを研究し、実際に配信を行ったかを解説し、音楽伝達の未来と今後の可能性について考察する。

Fig.1: Concert at Aichi Arts Center

Fig.2: Backstage

1-1P-1

1-1P-1　音声強調の処理歪みが音声認識に与える影響の分析:データ拡張が及ぼす影響の調査

Effect Analysis of Processing Distortion of Speech Enhancement on Speech Recognition: Investigating Effect of Data Augmentation

◎岩本 一真(同志社大院), 落合 翼, デルクロア マーク, 池下 林太郎, 佐藤 宏, 荒木 章子(NTT), 片桐 滋(同志社大)

◆背景と目的

- 音声強調の処理歪みが音声認識に与える影響の分析
  - ⇨ 音声認識の学習データの拡張が及ぼす影響の検証

◆実施要項

- 学習データに含まれる処理歪みの種類やスケールの多様性が音声認識に与える影響を分析
- 検討した学習データの拡張
  - (a) 強調信号の数（処理歪みの数）を増やす
  - (b) 処理歪みのスケールの多様性を増やす

◆評価実験

- 強調信号の数を増やしても, 原音付加が有効（i.e., WER : $\hat{s} > \bar{s}$）
- 処理歪みへの頑健性の獲得は不十分であることを確認

(a)

training data		evaluated signal		
# of y	# of $\hat{s}$	y	$\hat{s}$	$\bar{s}$
30k	–	15.9	16.6	13.0
30k	30k	15.5	13.7	12.3
30k	40k	15.7	12.6	11.9
30k	50k	15.8	12.7	12.2
30k	60k	16.2	12.6	12.2
30k	70k	16.4	12.6	12.1

(b)

training data				evaluated signal		
y	$\hat{s}$	$\hat{s}'$	α	y	$\hat{s}$	$\bar{s}$
✓	✓	–	–	15.7	12.6	11.9
✓	–	✓	-5	15.3	13.8	12.1
✓	–	✓	-10	15.4	14.8	12.4
✓	–	✓	-10～0	15.6	13.3	12.0
✓	✓	✓	-5	16.1	13.1	12.2
✓	✓	✓	-10	16.5	13.2	12.4
✓	✓	✓	-10～0	16.2	12.9	12.2

Fig.1 : Results of (a) increase in type of artifact error and (b) adjustment of artifact error scale. (WER [%] (lower is better))

1-1P-2

1-1P-2　大規模雑音混入音声データを利用した DNN 音声強調学習の効果

Effectiveness of the DNN-based speech enhancement using a large-sized noisy dataset

◎藤村拓弥，戸田智基（名古屋大学）

◆Noisy-target Training (NyTT) の枠組みでは，音声強調の学習に，容易かつ大量に入手可能な雑音混入音声を利用できることが示された．

◆本稿では，大規模な雑音混入音声データを用いて，NyTT とその改良法 IterNyTT に関する評価実験を実施する．

◆実験では，(i) より多くのデータを用いることで NyTT と IterNyTT がクリーン音声を用いる手法 (CTT) の性能を上回ること (Table 1)，(ii) クリーン音声に大量の雑音混入音声を追加して用いることで，音声強調性能の向上が可能であること (Table 2) が示された．

Table 1：Evaluation using a large-sized noisy dataset.
SI-SDR and PESQ of the input signals were 10.69 and 1.53.

Speech datasets	Method	SI-SDR	PESQ
Libri1	CTT	17.00	2.46
Libri2-CHiME1	NyTT	16.62	2.43
Libri2-CHiME1	IterNyTT	17.51	2.47
Libri2-DEMAND1	NyTT	17.27	2.62
Libri2-DEMAND1	IterNyTT	**17.79**	**2.64**
Libri2	CTT	17.96	2.73

Table 2：Evaluation using both clean and noisy targets.
SI-SDR and PESQ of the input signals were 10.69 and 1.53.

Speech datasets	enhanced?	SI-SDR	PESQ
Libri1	–	17.00	2.46
Libri1, Libri2-CHiME1	No	17.03	2.47
Libri1, Libri2-CHiME1	Yes	17.86	2.62
Libri1, Libri2-DEMAND1	No	17.57	2.59
Libri1, Libri2-DEMAND1	Yes	**17.87**	**2.67**
Libri1, Libri2	–	18.05	2.74

1-1P-3

1-1P-3 継続学習による MRI 騒音の変化に対応した被験者音声検出法

A method of subject voice detection on various MRI noises by continual learning

☆若松奏，武藤憲司（芝浦工大），小林洋介（室蘭工大）

◆MRI 検査時に被験者の発話音声を検出し，医師や技師に伝えることでコミュニケーションを可能にすることが目的である．MRI 駆動音は撮像法や被験者ごとに設定するパラメータによって，駆動音が変化する．学習に用いた MRI 駆動音と実際の MRI 駆動音が異なる場合には検出が困難であるという課題があった．また，様々な MRI 駆動音に対応するためには，過去の学習を維持しながら新規の MRI 駆動音を学習する必要がある．

◆本報告では，CNN(Convolutional Neural Network)を用いた継続学習においてリプライするデータの選択を，マルチコンディション学習で使用した重畳音声の SNR に着目して比較し検討した．

◆過去データをリプレイすることで F 値が向上し，全ての SNR のデータを均等にリプレイする手法(All)で F 値が最大であった．

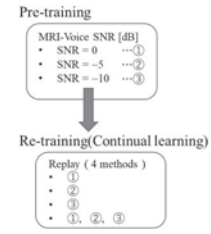


Fig.1 Method of selecting data for replay

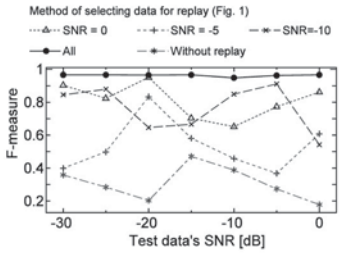


Fig.2 F-measure of models (Past data : EPI)

1-1P-4

1-1P-4　音響イベントのゼロショット学習における属性情報の拡張

Extending attribute information for zero-shot learning of sound events

☆LIN Yi-Han，高島遼一，滝口哲也（神戸大）

◆音響イベント分類において，イベントによっては学習データが入手困難な課題がある。

◆我々の先行研究では，属性とスペクトログラムの局所情報との関連を学習する手法として，Attribute Prototype Network を使用した音響イベントのゼロショット学習手法を提案し，学習データが一切存在しないイベントクラスを分類可能であることを確認した。

◆本研究では，属性情報を用いるアプローチをベースに，オノマトペを用いて属性情報を拡張することで，ゼロショット分類性能をさらに改善する手法を提案する。

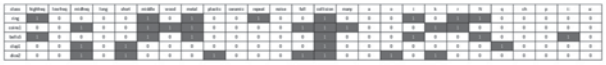


Fig.1：Example of binary attribute information of some classes in our experiments.

	bowl	clock2	kara	maracas	ring	tambouri	Average
previous [11]	83.9	15.4	21.0	99.3	64.1	55.1	56.6
proposed	99.5	0.0	34.5	99.3	86.1	91.5	67.1

Table 1：Recognition accuracies [%] of six unknown acoustic events

1-1P-5

1-1P-5 自動採譜に向けた複数楽器の楽曲における複数音高推定モデルの検討

Study on a multi-pitch estimation model for multi-instrument music aims at automatic music transcription

☆川凌司，坂野秀樹，旭健作（名城大）

◆背景・目的
- 自動採譜において複数音高推定は重要な役割を担う
- 複数楽器の楽曲において複数音高推定は未だ実用的でない
- 複数音高推定の深層学習モデルの検討を行う

◆実験
- Transformer と CNN を組み合わせた Conformer モデルの有効性を検討する
- 実験は 5 分割交差検証で行い、評価は Transformer と比較する

◆結果
- Conformer は少ないパラメータ数で Transformer の平均適合率を上回った
- Transformer と CNN の組み合わせが複数音高推定においても有効であると確認した

Table 1:Experimental results of 5-fold cross-validation.
(P: Precision, R: Recall, F: F-score, AP: Average Precision)

Model	Param.	Metrics (%)			
	$\times 10^6$	P	R	F	AP
Conformer	25.7	**63.4** ± 0.8	55.1 ± 3.8	**56.1** ± 2.4	**64.3** ± 0.7
Transformer	51.0	56.6 ± 0.1	**55.2** ± 3.6	52.6 ± 2.2	62.1 ± 0.6

1-1P-6

1-1P-6　手の動きに合わせた仮想音源の速度と音量の変化による感じ方の違い

Sensation of focused source being moved by hand-tracking with varying speed and loudness

☆廣橋美帆，羽田陽一（電通大）

◆手でつまんで動かすことができる焦点音源を生成
- モーションキャプチャで手の位置とつまむ動作を取得

◆焦点音源の移動速度と音量を手の動きに合わせて変化
- ものを動かしている感じや重量感を再現

◆有意差は見られないものの，聴覚以外の感覚を提示できる可能性
- 手の動きに合わせた移動と音量変化→ものを動かしている感じ
- 焦点音源の移動の滑らかさ→重さの感じ

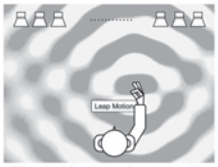
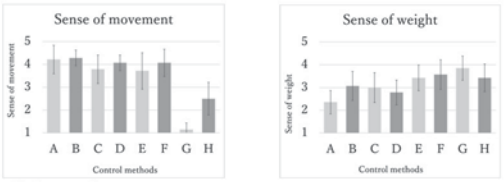


Fig. 1 Pinch and move the focused source.

Control method	Moving speed	Volume change
A	Fast	No
B	Fast	Yes
C	Normal	No
D	Normal	Yes
E	Slow	No
F	Slow	Yes
G	No movement	No
H	No movement	Yes



(a)Sense of movement　(b) Sense of weight

Fig. 2 Results of listening test.

1-1P-7

1-1P-7　進化的アルゴリズムによる帯域分割型多点制御法

Band-splitting Multi-point Control Method by Evolutionary Algorithm.

○辻村侑弥(龍谷大院・理工学研)，安枝和哉(東京医療保健大)，片岡章俊(龍谷大・先端理工)

◆本研究では 多点制御法に用いる制御点配置を周波数帯域毎に進化的アルゴリズムを用いて決定した。

◆計算機シミュレーションにより 再生エリアと抑圧エリア間の音圧レベルを比較し 制御点配置と周波数帯域の関係及びエリア再生の性能を確認した。

◆低周波数帯域(177Hz 以下)では抑圧制御点の配置されない箇所を有する配置、中周波数帯域 (708~2828 Hz)では従来法に近いＬ字型の配置等、周波数帯域毎に傾向の異なる制御点配置の得ることができた。

◆各制御点配置を用いることでエリア再生の性能が平均で 2.227dB 向上した。

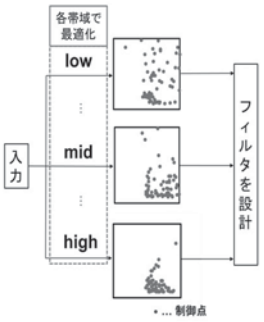


Fig1: 帯域分割型多点制御法のイメージ

Table1: 各周波数帯域の適合度*

周波数帯域[Hz]	従来法[dB]	提案手法[dB]
100 ~ 177	15.403	17.667
177 ~ 354	18.838	21.156
354 ~ 708	26.311	27.767
708 ~ 1414	31.907	34.285
1414 ~ 2828	33.618	37.338
2828 ~ 3400	30.993	31.702
100 ~ 3400	30.436	33.183

*適合度：再生エリアと抑圧エリアのパワー

1-1P-8

1-1P-8　音源の向きを変更したバイノーラル合成と音像知覚の検討

Binaural synthesis and sound image perception with different sound source orientations

◎西山織絵(KDDI 総合研／千葉工大・院)，大久保翔太，堀内俊治(KDDI 総合研)

◆ 本稿では，聴取者から見た発話者の位置(方向 1)と，発話者の向き（方向 2）が変化したときに聴取者はその変化を知覚できるのかを検証するため，発話者の放射特性と，聴取者から見た発話者の方向を反映した音源を用いて主観評価実験を行った．

◆その結果，聴取者に向かって話している，あるいは聴取者に背を向けて話している場合の変化は知覚できる傾向にあった．また，聴取者の前方よりも 90°，135°の正答率が高かった．

Table 1 Correct answer rate

方向1(°)	方向2(°)	正答率(%)
0	0	40.0
	90	20.0
	180	23.3
45	0	33.3
	90	16.7
	180	13.3
90	0	66.7
	90	50.0
	180	53.3
135	0	73.3
	90	50.0
	180	56.7
180	0	26.7
	90	10.0
	180	50.0

1-1P-9

1-1P-9　ユーザ自身の位相調整に基づく骨伝導による騒音抑制システムの検討

Investigation of noise reduction system by bone conduction based on user's own phase adjustment

☆酒井賢人(静大・院・総合科技研)，立蔵洋介(静大・院・総合科技研)

◆目的：0.9 ~ 1.0 kHz における提案手法による騒音抑制効果を評価

◆提案手法：使用者自身が骨伝導（二次音源）の位相・振幅調整を行い，外部騒音の抑制を試みる

◆調査内容：伝達関数の次数を変化させながら 0.9 ~ 1.0 kHz の抑制を試みた

◆結果：伝達関数の次数を適切に選択することで 0.9 ~ 1.0 kHz の抑制を達成した

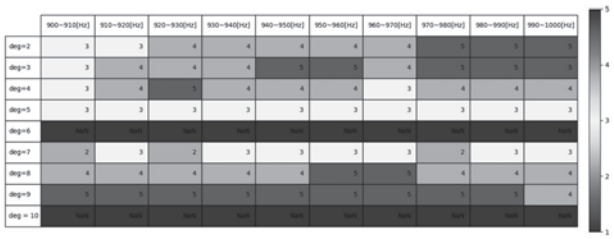


Fig.1 Evaluation of noise reduction performance in each order of transfer function. (The horizontal axis is the frequency band in which the evaluation

1-1P-10

1-1P-10 複数誤差マイクの時空間関係を考慮したバーチャルセンシング型能動騒音制御

A method of ANC system with virtual sensing technique assuming spatio-temporal shiftable relationship

☆吉松亨真,信夫直樹,羽田陽一(電通大),
伊藤弘章,鎌土記良,小塚詩穂里(NTT)

◆ 耳元近傍に複数の誤差マイクを等間隔に設置する耳元の音圧予測および ANC の手法を提案する.

◆ 提案手法は,複数の誤差マイク間で観測される音圧の時空間関係は,波長に比べて短い距離では平行移動可能であると仮定し,その時空間関係を音圧予測フィルタとして推定し(Fig. 1 (a)α_1, α_2),推定した予測フィルタを用いて計算した耳元での音圧(Fig. 2 (b))に基づいて騒音制御フィルタを逐次更新する.

◆ 騒音源位置の変化に対するロバスト性,壁や頭部の反射による収束性能への影響を計算機シミュレーションで検証した結果について報告する.

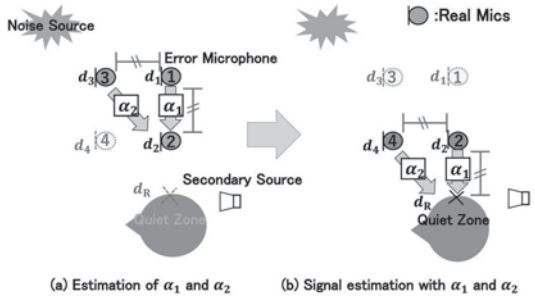


Fig. 1：Basic concept of proposed method

1-1Q-1

1-1Q-1　残響除去を目的としたConv-TasNet の学習法

Training method of Conv-TasNet for dereverberation

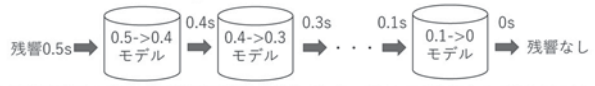
☆岡田健佑(龍谷大院・理工学研), 片岡章俊(龍谷大・先端理工)

◆Conv-TasNet による残響除去では, 除去後の音声に歪みが発生することや, 高周波数帯域の成分があまり考慮されず音声が籠る問題点がある.

◆本研究では Conv-TasNet のモデル構成を用いて, 学習時の学習データに対して高域強調フィルタの適用や損失関数の変更, 残響除去の構成の 3 点から有効な学習法を検討した.

◆高域強調フィルタはプリエンファシスフィルタを用い, 損失関数は従来の時間領域ではなく時間周波数領域での計算, 残響除去の構成では一括で除去するのではなく段階的に除去する構成を検討した.

Fig.1 段階的な残響除去の構成



◆客観評価実験と主観評価実験の結果から, 従来の学習法より音声の品質を下げることなく, 音声の歪みを抑えた残響除去ができた.

Table.1 客観評価実験結果

	PESQ	STOI	Si-SDR
処理前	1.898	0.399	-7.41
Si-SDR Loss	2.234	0.528	10.06
プリエンファシス	2.250	0.529	10.18
MSE Loss	2.317	0.537	10.85
段階的な残響除去	2.349	0.552	10.40

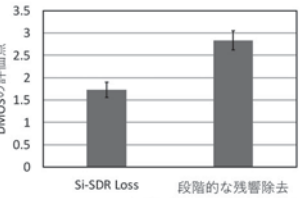


Fig.2 DMOSの平均点の95%信頼区間

1-1Q-2

1-1Q-2　Double-DIP に基づく教師なし雑音除去の分析と性能改善

Analysis and performance improvement of Double-DIP based unsupervised noise reduction

☆大中緋慧, 宮崎亮一（徳山高専）

◆与えられた雑音混入信号１つに対して, 尖度の低い信号（雑音）と尖度の高い信号(クリーン信号)に分けて生成されるように２つの DNN を学習させ, 最終的にクリーン信号を推定する手法を以前に提案した.

◆この方法の適用可能な条件は, ホワイトノイズなどの尖度が低い雑音の場合に限定され, それ以外では雑音が残る問題がある. (Fig. 2 (c))

◆本稿では, 上記の問題を解決するために信号を分割した局所領域での尖度を損失関数に導入する. また, 音声歪みとミュージカルノイズを抑制するためのバッチ学習と平均処理を導入した.

◆客観評価より, 以前提案した手法と比較して改善を確認した. (Fig. 1)

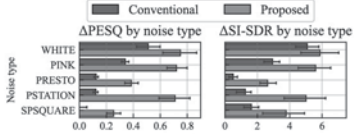


Fig.1: Result of objective evaluation.

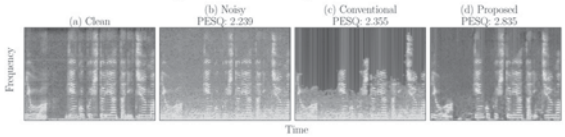


Fig.2: Spectrogram of processing results for a PSTATION noise at SNR 7.5 [dB].

1-1Q-3

1-1Q-3　マルチモーダル DNN を用いた船舶検出におけるノイズ対策のための信頼度推定ユニット

Confidence Estimation Unit for Boat Detection using Multimodal DNN in Noisy Environment

☆庭山耕平, 武藤憲司(芝浦工大), 小林洋介(室蘭工大)

◆マルチモーダル DNN を用いた船舶検出において, 強風により生じる風雑音や夜間や雨天時の視界不良など, 悪天候時に環境音や映像に混入するノイズに弱いという問題があった.

◆本報告では, クリーンなデータである度合いとして信頼度を定義し, 信頼度の高いモーダルの情報をより重視して船舶検出を行う手法を提案した.

◆信頼度を用いたモーダル間の重み付けを行うことでノイズ環境下での船舶検出性能がＦ値で最大 0.06 ポイント向上した.

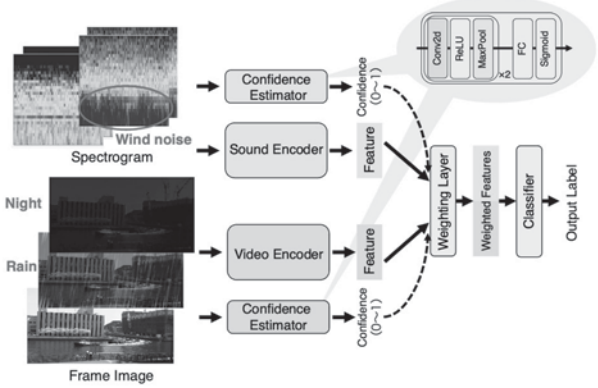


Fig.1 Multimodal boat detection using confidence levels

1-1Q-4

1-1Q-4　ビジュアルマイクロホンを用いた鉄琴のキー識別

Key classification of glockenspiel sound using Visual Microphone.

☆髙橋瑞貴，立蔵洋介(静岡大学)

◆ 識別実験：無音の映像から音を抽出するビジュアルマイクロホンを用いて鉄琴音 6 音を収音し，識別を実施

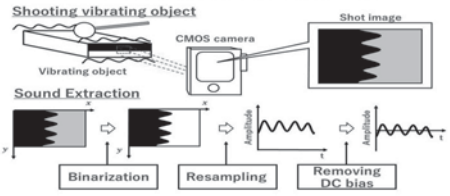


Fig. 1 Overview of glockenspiel sound extraction via CMOS camera.

◆ 結果：正解率が約 98.9 %となった．また，抽出音の特徴を確認すると同一音でも異なる特徴を持つデータが確認された (Fig. 2)

◆ 追加実験：シャッタースピードを変更させ，抽出音の特徴の変化を調査

◆ 結果：シャッタースピードの変化によって抽出音の特徴が変化することを確認 (Fig. 3)

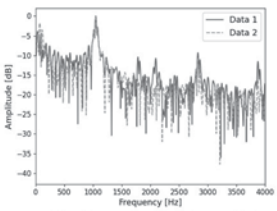


Fig. 2 Amplitude spectrum of C6.

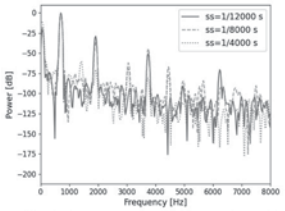


Fig. 3 Power spectrum of extracted sound at different shutter speed.

1-1Q-5

1-1Q-5　22.2 マルチチャネル音響を用いた高次アンビソニックスに基づく放射音場再現

Radiated Sound Field Reproduction for 22.2 Multichannel Audio Based on Higher-order Ambisonics

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

◆22.2 マルチチャネル音響における従来の仮想音源構築手法では，仮想音源位置を制御できるが，音源の放射指向特性は考慮されていない．

◆放射指向特性は，人間の音声や楽器などの音源毎に異なるため，音の聞こえ方は音源の向きにより異なる．そのため，音源の放射指向特性を考慮することでより実音源に近い仮想音源が構築可能となる．

◆本稿では，所望の放射指向特性をもつ音源による放射音場を再現することで，より実音源に近い仮想音源の構築を目指す．計算機シミュレーションの結果，22.2 マルチチャネル音響で所望の放射指向特性をもつ音源による放射音場が再現され，提案手法の有効性が示された．

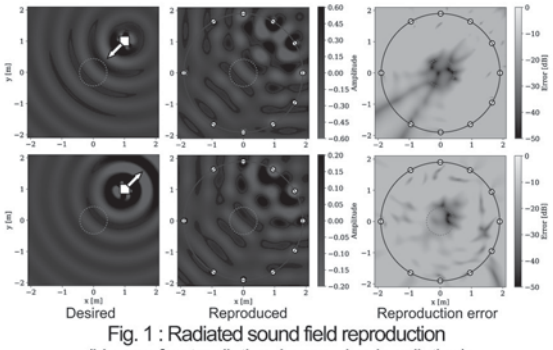


Fig. 1 : Radiated sound field reproduction
(Upper : front radiation, Lower : back radiation)

1-1Q-6

1-1Q-6　投動作によって投げられたように移動する焦点音源の受聴について

A Study on Perception of Focused Sound Source Moving with Throwing Motion

☆山崎萌恵，羽田陽一（電通大）

◆投動作に合わせて移動する焦点音源を生成
- ➢ モーションキャプチャで投動作を認識
- ➢ 32 ch 直線スピーカアレイによる波面合成で焦点音源を生成
- ➢ 非同期処理と骨格位置の線形補間による低遅延再生

◆焦点音源の移動で投げ感覚が得られるかと，知覚に必要な要素を調査
- ➢ 投げ感覚，腕の長さ知覚，楽しさ，移動速度などを主観評価
- ➢ 焦点音源位置の制御方法
  - ✧ 身体動作に関わらず受聴位置に呈示される焦点音源（A）
  - ✧ 手のひらの位置に追従する焦点音源（B）
  - ✧ 投動作に合わせて移動する焦点音源（C，D，E）

◆投げ感覚向上には，音像位置のわかりやすさと移動速度の向上が必要

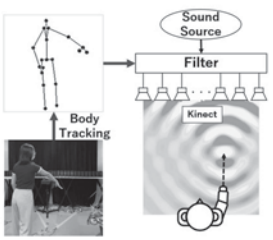


Fig.1 : System Overview

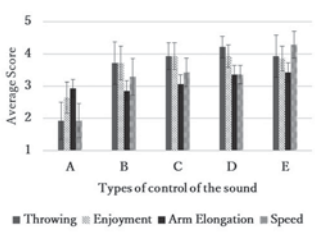


Fig.2 : Subjective evaluation scores

1-1Q-7

1-1Q-7　頭部伝達関数の球面調和展開に関する一検討

A study on spherical harmonic expansion of head-related transfer functions

〇松井健太郎，伊藤敦郎(NHK)，坂本修一(東北大)

◆球状マイクロホンアレイと頭部伝達関数を用いたヘッドホンベースの収音・再生法として，球面調和領域での積和により駆動信号を導出，レンダリングする方法が提案されている。

◆この球面調和領域でのレンダリングにあたり，頭部伝達関数の球面調和展開に求められる要件を検討し，低次の基底関数で頭部伝達関数を再合成した場合にも歪の拡大を抑制する漸近展開法を提案する。

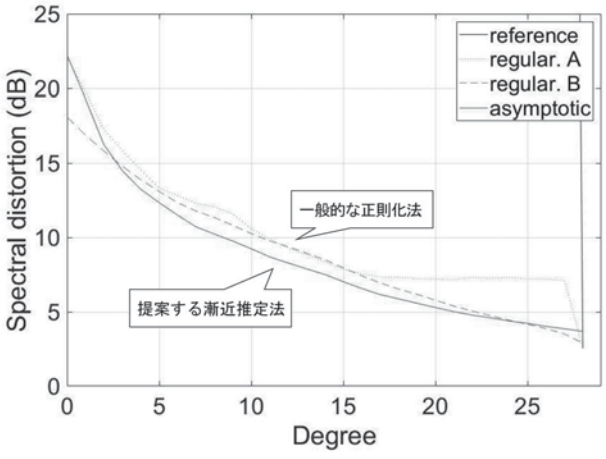


Fig. 1: Spectral distortion when truncated degree of expansion

1-1Q-8

ANC システムにおけるリモートマイクロホン法のコヒーレンス調整フィルタによる性能改善に関する検討

A study on performance improvement of RMVS ANC system with coherence adjustment filter

☆謝天雨，豊岡祥太，梶川嘉延（関西大学）

◆ANC システムは騒音に対して同振幅，逆位相の擬似騒音を生成し，騒音と重ね合わせることで騒音低減を行うシステムである．

◆所望地点にマイクロホンを設置することなく騒音低減を可能とするバーチャルセンシング技術の一つにリモートマイクロホン(RMVS)法がある．RMVS は誤差マイクロホンとバーチャルマイクロホン間のコヒーレンスが低い帯域で，騒音低減性能が劣化する問題がある．

◆本稿では RMVS にコヒーレンス調整フィルタ(CAF)を導入することで，マイクロホン間のコヒーレンスを向上させる．

◆Fig.1 にバーチャルマイクロホン地点の周波数スペクトルを示す．従来法で騒音低減性能が大きく劣化している帯域において，提案法では騒音低減効果が向上していることがわかる．

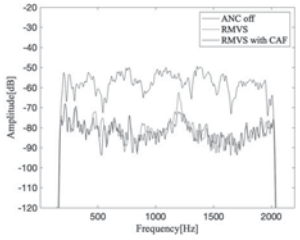


Fig.1：Comparison of noise reduction effect

1-1Q-9

1-1Q-9　頭部近傍マイクと剛球モデルを用いた音圧推定における頭部形状の影響

Investigation of Sound Prediction at Ear of Human Head Shape Based on Microphones near Head and Rigid-Sphere Model

☆信夫直樹（電通大），伊藤弘章（NTT），鎌土記良（NTT），小塚詩穂里（NTT），羽田陽一（電通大）

◆耳元から離れたマイク，スピーカによる ANC の実現に向けて，頭部を剛球とみなす剛球モデルに基づく耳元の音圧推定手法を検討した．頭部形状や耳介の影響に対して，剛球モデルによる実頭部近似の有効性を調査した．

◆マイクの配置方法に着目し，これまでに検証した球面上の等間隔配置のみならず，より実用性を考慮した半球面状の非等間隔配置による耳元の音圧推定も検討した．

◆無響室に騒音源用スピーカおよび騒音観測装置を設置し，装置内部に剛球またはダミーヘッドを挿入して測定実験を行った．騒音到来方向を変えながら，耳元の音圧を周波数領域で推定し，その精度を評価した．

◆ダミーヘッドの耳元の音圧推定精度は，剛球の場合と同程度であることが分かった．このことから，頭部形状に関する個人性を考慮せずとも，剛球モデルを ANC に適用することで人の耳元の騒音抑制が期待される．

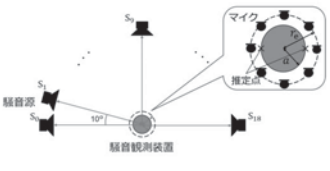


Fig.1：騒音源と騒音観測装置の位置関係

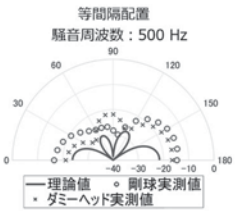


Fig.2：騒音到来方向に対する耳元音圧の推定誤差

1-3P-1

1-3P-1　くずし字翻刻のための古文用音声認識システムの検討

A study on ancient texts speech recognition system for Kuzushiji transcription

☆張宇涛，戸塚史織（立命館大院），西浦敬信，赤間亮（立命館大）

◆本研究では，くずし字翻刻に音声認識システムを導入するため，高精度な古文認識を実現できる音声認識システムの構築について検討する．Fig. 1 に古文用音声認識システムを導入したくずし字翻刻作業工程を示す．

◆古文に対する認識性能を向上させるため，江戸時代洒落本を用い，古文用の言語モデルと発音辞書を作成した．

◆評価実験の結果により，古文用言語モデルの構築が古文の認識精度を改善できることがわかった．くずし字翻刻に音声認識システムを導入する可能性が示された．

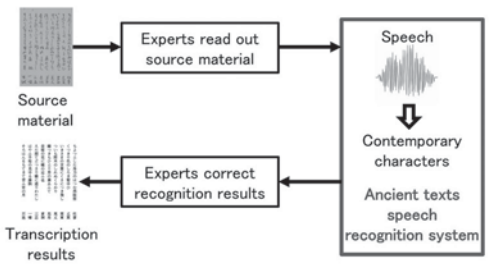


Fig. 1 Kuzushiji transcription work process with ancient texts speech recognition system.

1-3P-2

1-3P-2　wav2vec 2.0 によるラベル無し複数患者音声を用いた脳性麻痺患者の音声認識

Speech recognition of cerebral palsy patients using unlabeled speech of multiple patients with wav2vec 2.0.

☆松坂勇樹，高島遼一，滝口哲也（神戸大）

◆脳性麻痺患者の支援として，音声認識による支援システムが期待されるが，構音障害(Dysarthria)により健常者と発話特徴が異なるため，健常者音声で学習した従来の音声認識システムでの認識は困難である．

◆そのため，患者本人の音声で音声認識モデルを学習する必要があるが，学習のための音声データが十分量用意できないという問題がある．

◆本研究では，より多く収集が可能なラベル無し音声に着目し，活用法として wav2vec 2.0 による自己教師あり学習を検討する．

◆Fig.1 のように，大量の健常者音声による自己教師あり学習の後，ラベル無し患者音声による自己教師あり学習を再度行い，最後に少量のラベル付き音声でファインチューニングをする．

◆論文では，ラベル無し患者音声として特定患者に限らず，複数患者の音声で学習した場合の実験を行い，有効性を調査した．

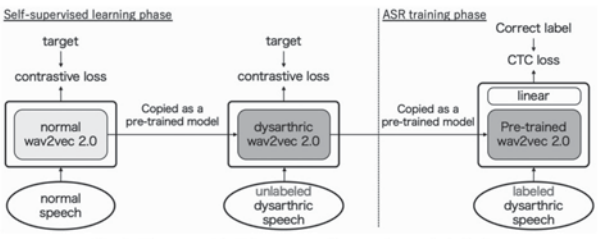


Fig.1：Proposed training procedure using unlabeled dysarthric speech based on wav2vec 2.0.

1-3P-3

1-3P-3　深層学習による口唇情報を用いた雑音下での音声認識精度の改善

Improving Speech Recognition Accuracy in Noisy Situations Using Deep Learning with Lip Information

☆伊達龍斗（徳島大），西村良太（徳島大），
太田健吾（阿南高専），北岡教英（豊橋技科大）

◆顔画像から得られた特徴点の座標変位を用いて特徴量を抽出し，既存の音声認識モデルの出力と組み合わせて後段の LSTM によって統合するマルチモーダル音声認識器の構築を行った。

◆音声データについてはSNR が 10 dB から 0 dB になるように段階的にホワイトノイズを重畳し，実験に使用した。

◆使用するモダリティを変化させて，ノイズによる CER の変化や，誤認識率が高い音素について評価した。

◆ノイズが大きい環境下で誤認識率が高くなった音素は，正解の音素と口形素が同じであるものや，口唇の動きが小さいものであった。

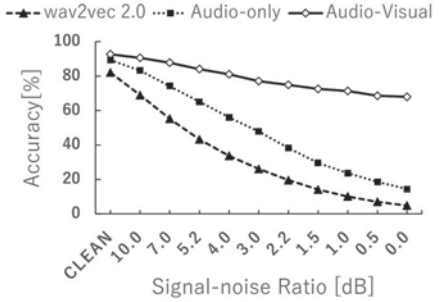


Fig.1：Recognition rate for each SNR

1-3P-4

1-3P-4　音声認識における日本語向け自己教師あり学習の言語依存性の検討

Exploration of Language Dependency for Japanese Self-supervised Speech Representation Models.

◎芦原孝典，森谷崇史，松浦孝平，田中智大（日本電信電話株式会社）

◆音声分野においても自己教師あり学習 (SSL) が注目されている．

◆最近では大規模な英語/マルチリンガル音声で学習された SSL モデルを用い，クロスリンガル ASR モデルとして構築することで，低リソース言語タスクで大いに成功を収めている (Fig. 1 のモデル(a))．

◆一方，同一言語で SSL 並びにファンチューニングするモノリンガルモデル (Fig. 1 のモデル(b)) と比較して，どの程度効果的なのかはまだあまり分かっていない．

◆そこで本稿は，ターゲット言語の未ラベルデータが一定量収集可能である場合，モノリンガルモデルはクロスリンガルモデルより高い精度を達成できるのか，またもしそうであるならターゲット言語の未ラベルデータはどの程度必要なのか，その２点を実験的に検証した．

◆最後に，日本語 SSL ベースの ASR モデルを網羅的に探索し，いくつかのタスクで state-of-the-art の精度を達成した (Table 1)．

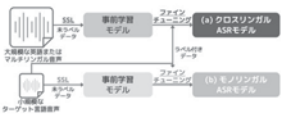


Fig. 1：SSL-based models for target languages.

Table 1：Evaluation results with different unlabeled data, encoder, and masking strategy. The table shows only CSJ results due to space limitations.

	CSJ							
ID	Unlabeled data	Enc.	Mask	CER			
				dev	eval1	eval2	eval3
B1	-	T	pre	4.2	5.3	3.8	4.1
B2	-	C	pre	4.1	5.0	3.8	3.9
M1	CSJ	T	post	3.4	4.6	3.2	3.7
M2	CSJ+LTVS	T	post	3.5	4.5	3.2	3.7
M3	CSJ+LTVS	T	pre	3.5	4.2	3.3	3.7
M4	CSJ+LTVS	C	post	3.3	4.2	3.0	3.5
M5	CSJ+LTVS	C	pre	**3.1**	**3.9**	**2.9**	**3.2**
M6	CSJ+LTVS+LS	C	pre	3.4	4.3	3.1	3.4
C1	XLSR-53 [8]		post	3.8	4.7	3.6	3.8
C2	XLS-R-128 [9]		post	3.7	4.7	3.5	3.8
E	Karita et al. [23]			**3.1**	4.1	3.2	3.5

1-3P-5

1-3P-5　異常検知による非母国語話者の日本語発話誤り検出における基本周波数の導入

Fundamental frequency in error detection of Japanese pronunciation for non-native speakers using anomaly detection techniques

☆Zhou Qihang，實廣 貴敏（愛知工科大）

◆本研究では，DNN と異常検知技術を用いた発話誤り区間を検出するアルゴリズムを提案する。日本語母国語話者による日本語発話のみで学習した Autoencoder により，非母国語話者の発話を評価，誤りを検出する。また，MFCC ベースの音声特徴量に基本周波数を導入し，複数フレームを連結した特徴量を用いることで精度向上を図る。

◆評価実験により，誤棄却率（FAR）と誤受理率（FRR）を Fig. 1 に示す。（MFCC＋対数パワー）＋Δ＋ΔΔ の前後 10 フレーム結合したものに，fo＋Δ＋ΔΔ の前後 15 フレームを結合し，音声特徴量とした。等価エラー率は 17.3%が得られ，精度は単フレームでの MFCC 特徴量より 7.2%改善した。

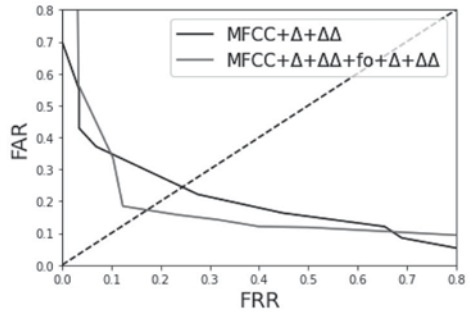


Fig.1：ROC curve.

1-3P-6

1-3P-6　音声認識DNN内部処理可視化方法の検討

Study of internal processing visualization method for automatic speech recognition DNN

○金寺　登（石川高専）

◆音声認識 DNN 内部で何を根拠に判別しているかが不明である。

◆そこで，DNN 内部の処理を可視化する方法として，ネットワーク出力を各ユニット出力で偏微分した値の変動によって，識別の根拠となる重要なユニットを特定する方法などを提案した。

◆本研究では DNN の一部区間を AI 手法としてよく用いられるランダムフォレスト等で置換し，内部動作を可視化する方法を提案する。

◆DNN とほぼ同等の性能が得られることが確認された。

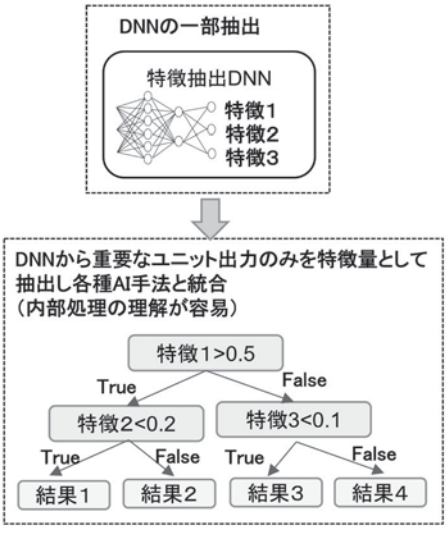


Fig.1 A system that replaces part of the DNN with AI methods

1-3P-7

1-3P-7　音声モーフィングにおける時間軸方向の対応点数が品質に与える影響

Influence of number of anchor points in temporal axis on sound quality of speech morphing

☆堀部貴紀，森勢将雅(明治大)，河原英紀(和歌山大)

◆ 音声モーフィングにおける対応点の時間軸方向に着目し，各音素区間における対応点数が品質にどのような影響を与えるか評価した．

◆ 音素境界の対応点のみで合成されたモーフィング音声をリファレンス，各音素区間の対応点数が 2 点，3 点，5 点，10 点のモーフィング音声を音刺激として，より人間らしく自然であるかという基準で 5 段階評価した．

◆ CMOS で評価したところ，時間軸方向での増加は品質改善には寄与せず，対応点数の増加に伴い，品質の低下が確認された．

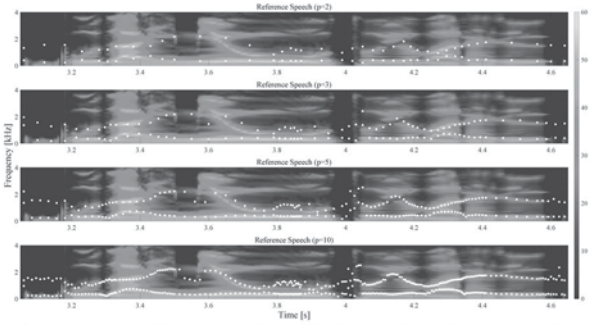


Fig.1: Spectrograms with anchor points (p is number of anchor points each phoneme interval). The circles indicate F1 and F2 set as anchor points.

1-3P-8

1-3P-8　CycleGAN に基づく音声・歌声の同時変換に関する検討

A study on speech and singing voice conversion based on CycleGAN

☆府川 省吾，能勢 隆，今井 柊平，伊藤 彰則（東北大）

◆本稿では，音声と歌声を単一のモデルで同時に学習・変換可能な声質変換手法を提案する．

◆メルスペクトログラムを変換する CycleGAN-VC2 に音声と歌声を同時に学習させると，歌声の音程が入力の音程をオクターブの整数倍だけシフトしたものから外れる箇所が発生する．

◆対数 F0 のシフト幅をオクターブ単位に制約する octave F0 loss を用いると，音声の対数 F0 のシフト幅が話者間の対数 F0 の平均の差ではなくなり，話者性が低下する．

◆対数 F0 のシフト幅を音声か歌声かに応じて切り替える adaptive F0 loss を導入することで，推論時は暗黙的に対数 F0 のシフト幅が切り替わるような変換を実現した．

◆adaptive F0 loss を導入した提案法 Adaptive により，音声と歌声の両方に対して，従来法 Baseline より高い品質，制御性および話者性を得られることを実験により示した．

Table 1　MOS for quality

	Speech	Singing	Average
Target	4.62 ± 0.12	4.57 ± 0.13	4.60 ± 0.09
Baseline	2.76 ± 0.16	2.51 ± 0.22	2.64 ± 0.14
Octave	2.90 ± 0.18	**3.53 ± 0.17**	3.22 ± 0.13
Adaptive	**3.97 ± 0.16**	3.38 ± 0.16	**3.68 ± 0.12**

1-3P-9

1-3P-9　発話音声を用いたオペラ歌唱音声合成の初期検討

Initial study of opera-singing voice synthesis using speaking voice

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

([1]神戸大学，[2]メック株式会社)

◆本研究では，声質変換手法を用いて**オペラ歌唱未経験ユーザーの発話音声からオペラ歌唱音声を合成可能なシステム**の実現を目的とする。

◆声質変換で変換先音声として用いるオペラ歌唱未経験ユーザーの発話音声をオペラ歌唱音声とのパラレルデータとすることで変換精度の向上を検討する。

◆**変換先音声をパラレルデータにすることで，**通常発話音声を用いるよりも**変換精度が向上した。**

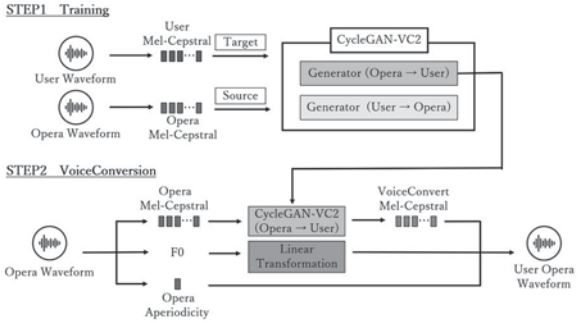


Fig.1 Training procedure using CycleGAN-VC2-based voice conversion

1-3P-10

1-3P-10 Towards Expressive Speech Conversion based on StarGANv2

Shangyang Mou[1], Jinhui Chen[2], Ryoichi Takashima[1], Tetsuya Takiguchi[1]

([1]Kobe Univ, [2]Prefectural Univ of Hiroshima)

◆Expressive Speech Conversion is a technique that aims to convert the emotional state of the utterance from one to another while preserving the linguistic information and speaker identity. Emotional expression is essential in our daily life for conveying intentions and social attitude. Therefore, expressive speech conversion is an important objective in voice conversion field. In addition, it is difficult for people with dysarthria to express themselves clearly. The expressive speech conversion can be useful as their auxiliary expressions, allowing them to express as well as people without dysarthria.

◆The overview of the method is shown in Figure 1

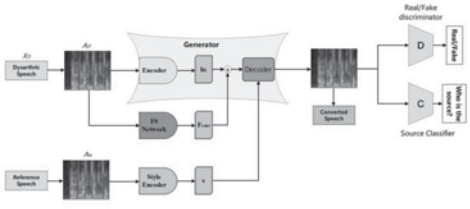


Fig. 1 X_D is neutral dysarthric speech, and h_x, F_{conv} and s denote the latent feature of the source, the F0 pitch features and the emotional style code of the reference speaker, respectively.

1-3P-11

1-3P-11　Speechsplitを用いたイントネーション・リズム・発音の矯正による外国語アクセント変換

Foreign accent conversion by correcting intonation, rhythm and pronunciation using Speechsplit

〇許　誠, 岸田　拓也, 中鹿　亘（電通大）

◆外国語アクセント変換とは、非ネイティブ話者の発話からアクセントを取り除いて、ネイティブ話者のような発話へ変換する技術である。
◆これまでの外国語アクセント変換では、アクセントをグローバルなスタイルとして表現されたが、アクセントが話者依存だけではなく発話依存でもあり、アクセントがイントネーション・リズム・発音などの情報にも含まれていると考えられる。
◆Speechsplitは発話をイントネーション・リズム・発音三つの埋め込み表現に分解し、発話を再構築する手法である。これを用いて、非ネイティブ話者の埋め込み表現とネイティブ話者の埋め込みを学習し、さらに、ネイティブ話者の埋め込み表現に近づくようにファインチューニングすることによって、イントネーション・リズム・発音を矯正するモデルを提案する。

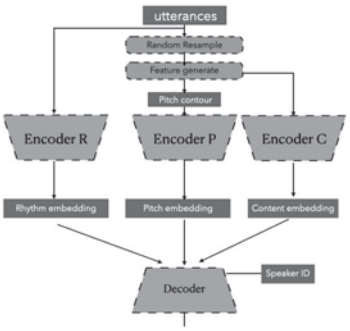


Fig.1：Proposal Method (Stage 1)

1-3P-12

1-3P-12　Correlation analysis between L2 learners' while-listening facial expressions and their listening dysfluency

☆Haokun Yang, Nobuaki Minematsu, Daisuke Saito (UTokyo)

◆In language learning, acquisition of a good listening skill is important. It is very often that learners **misidentify** given utterance due to a lack of good listening ability. In this situation, learners are faced with listening troubles, called **listening dysfluencies**.
◆In our previous work, a method of **online and acoustic** observation of evaluators' listening dysfluency was proposed [Zhu+,2021]. It was developed with speech technology called **shadowing** which has evaluators conduct oral dictation with as short delay as possible.
◆In this work, we pay special attention to how **facial expressions** of shadowers or listeners can tell about their **listening dysfluency**. We calculated correlations between **Action Unit** which is an expression coding system (Fig.1) and the listening dysfluency. If facial expressions can work as good descriptor of listening dysfluency, it can be detected not with speech technology but with **computer vision technology**.

Fig. 1: Action Units

◆The results showed a **moderate** correlation between the two kinds of observations. We also conducted an experiment to predict shadowing based listening dysfluencies only from listeners' facial expressions.

1-3Q-1

1-3Q-1　ストリーミング End-to-End 音声認識のための RNN Transducer の最小遅延学習

Minimum Latency Training of RNN Transducers for Streaming End-to-End Speech Recognition

〇篠原雄介（ヤフー），渡部晋治（CMU）

◆End-to-End 音声認識，とくに RNN Transducer (RNN-T) の研究が盛んに行われており，プロダクトへの導入も進んでいる．
◆ストリーミング音声認識では高い認識精度だけでなく，遅延を小さく抑えることが重要だが，RNN-T の学習では負の対数尤度を最小化（≒精度を最大化）するため，遅延が大きくなる課題があった．
◆本発表では，負の対数尤度と「期待遅延」の重み付き和を最小化する「最小遅延学習」を提案する．また期待遅延の勾配をフォワードバックワード法で効率的に計算できることを示す．これにより精度と遅延を同時に最適化できる．
◆英語音声認識（WSJ）の実験において，遅延が 220 ミリ秒から 27 ミリ秒に削減すること，また従来法（アラインメント制約学習 110 ミリ秒，FastEmit 67 ミリ秒）よりも優れていることを確認した．

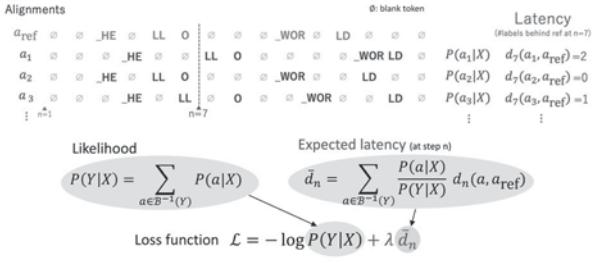


Fig. 1: Minimum latency training jointly minimizes negative log-likelihood and expected latency.

1-3Q-2

1-3Q-2　器質性構音障害者音声認識のための簡易ラベルによる中間層ロスの導入

Intermediate loss based on simple labels for speech recognition of organic dysarthria

☆冨士原健斗，高島遼一(神戸大)，杉山千尋，田中信和，
野原幹司，野崎一徳(大阪大)，滝口哲也(神戸大)

◆器質性構音障害者は発音スタイルが健常者と異なるため，健常者に比べて音声認識の精度が低くなる．
◆本研究では，話者ごとの発音スタイルを教師ラベルに反映させることで，学習しやすく簡略化された「簡易ラベル」を作成する．
◆まず特定話者音素認識モデルを学習し，学習データに対する誤認識パターンを元に簡易ラベルを作成した後、改めてモデルを学習する．
◆新規モデルの学習時は，最終層と中間層の二箇所で学習用の損失を計算する．中間層の損失計算に簡易ラベルを用いることで，モデルが誤認識パターンの情報を考慮して学習することが期待できる．
◆実験の結果，話者によって手法の有効性は異なっていた．話者ごとの相性は，誤認識パターンの傾向に依存している可能性が示唆された．

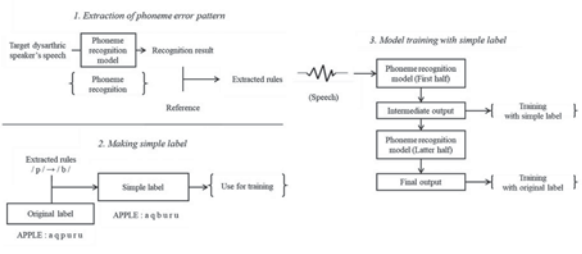


Fig.1：Overview of our proposed method.

1-3Q-3

1-3Q-3　複数 Encoder–Decoder 音声認識モデルの統合による頑健な音声認識システムの構築

Constructing Robust Speech Recognition System by Integrating Multiple Encoder–Decoder Speech Recognition Models.

☆北條圭悟，森大輝，若林佑幸（豊橋技科大），
小川厚徳（NTT），北岡教英（豊橋技科大）

◆複数の音声認識システムを統合することは，認識精度向上に有効である一方で，End-to-End 音声認識での検討は少ない。

◆End-to-End 音声認識において，精度向上の方法として言語モデルの併用があり，近年 Density Ratio Approach（DRA）が登場した。

◆我々は複数の音声認識システムの統合と DRA の考え方を利用し，新たなドメインにおける効率的な音声認識システムの構築を提案する。

◆実験の結果，提案手法は 1 つの音声認識モデルを使用する従来の統合手法の認識精度を上回ることが確認された。

◆また，本手法は大規模なデータセットがない場合において，それと同等の性能をもつ音声認識システムを構築する場合に有効である。

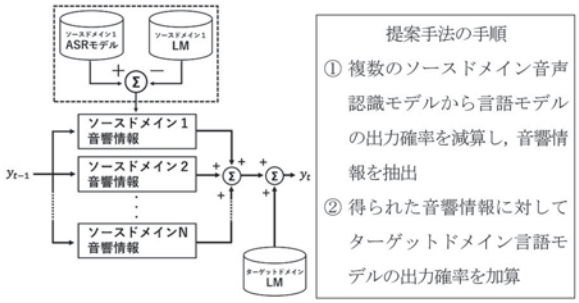


Fig.1：Overview of the proposed method.

1-3Q-4

講演取消

1-3Q-5

1-3Q-5　A Streaming RNN-Transducer Speech Recognition System for Japanese

☆ Wen Shen TEO, Yasuhiro MINAMI (UEC)

◆We trained the first streaming Transducer Japanese ASR on the Next-Generational Kaldi (NGK) framework.

◆We implemented a new CSJ tag parsing mechanism built into the recipe.

◆We designed a new validation dataset that better reflects the speech setting and speaker gender distribution of the training dataset.

◆We managed to obtain better results than Kaldi and WeNet.

◆As a streaming, one-pass model, our results are not enough to rival the CERs achieved and published by ESPnet.

Framework	eval1	eval2	eval3
Icefall	5.08	3.80	4.07
Kaldi	7.59	6.44	7.02
WeNet	7.28	4.81	5.44
ESPnet	4.5	3.3	3.6

Figure 1: Best CERs obtained for each framework

1-3Q-6

1-3Q-6　脊髄性筋萎縮症者音声合成における明瞭性および話者性を考慮した適応手法の検討

A study of an adaptive method considering intelligibility and individuality on speech synthesis for a person with spinal muscular atrophy

☆吉本拓真，高島遼一（神戸大），
△佐々木千穂（熊本保健科学大），滝口哲也（神戸大）

◆本研究では，構音障害を持つ重度の脊髄性筋萎縮症（SMA）者を対象に，健常者 TTS モデルを SMA 者データで適応することにより，明瞭かつ SMA 者本人らしさのある音声を合成するシステムを検討する。

◆先行研究では，モデル適応の際に健常者音声認識の損失を考慮することで明瞭性を確保したが，話者性の点で不十分であった。

◆提案手法では，モデル適応の際に合成音声の話者埋め込み（本研究では x-vector [Snyder *et al.*, 2018] を使用）が SMA 者の収録音声から得られる話者埋め込みに近づくような損失をさらに加える。(Fig. 1)

◆提案手法の合成音声は収録音声と比較して明瞭性が高まった。また，話者埋め込みはより SMA 者の収録音声に近い部分へ埋め込まれるようになり，提案手法は従来手法より話者性に優れた音声を生成できた。

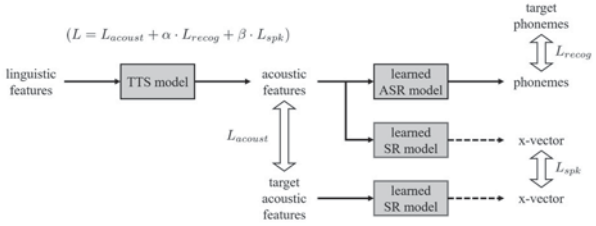


Fig. 1: Adaptation procedure on the proposed TTS system.
All model parameters except the acoustic model in the TTS model are fixed during adaptation.

1-3Q-7

1-3Q-7 キャラクター顔画像を用いた深層学習に基づく音声変換

Voice conversion by deep learning using anime face images

☆前島健哉、菊池英明(早稲田大学)

- ◆ キャラクター顔画像から推定した話者情報を用いた音声変換を提案
- ◆ 提案法はVAE、VQVAEを組み合わせたモデル
- ◆ 実験によって収集した、顔画像と音声のペアデータを用い提案モデルを学習
- ◆ 提案法は変換精度において既存手法を上回ることを確認

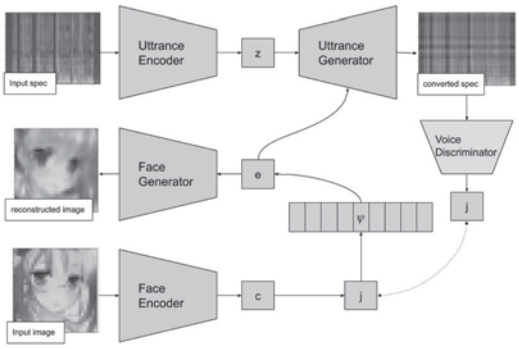


Fig. 1: 提案モデルの概要図

1-3Q-8

1-3Q-8 Emotional Voice Conversion with a Novel Content-Style Fusion Block

☆Xunquan Chen[1], Jinhui Chen[2], Ryoichi Takashima[1], Tetsuya Takiguchi[1]
([1]Kobe University, [2]Prefectural University of Hiroshima)

◆Emotional Voice Conversion (EVC) technology aims to convert a neutral voice into an emotional one while keeping the linguistic information and speaker identity unchanged.

◆In this study, we propose an emotional voice conversion framework with a novel content-style fusion block for rearranging the emotional style distribution. The global-level emotion information can be extracted by encoding the whole utterance into a fixed-length vector. For local-level emotion information, we adopt the proposed content-style fusion block for implementing time varying emotion representation.

◆As show in Fig. 1, the generator is an encoder-decoder module in our work. Our proposed content-style fusion block can appropriately embed an emotion feature which depends on the content information for another phoneme.

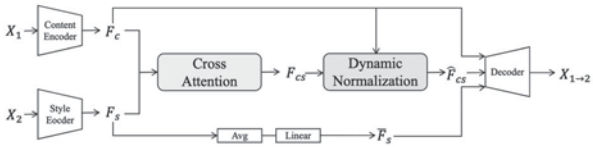


Fig. 1 : The generator architecture of the proposed model. Here, the content-style fusion block is composed of a cross attention module and a dynamic normalization module.

1-3Q-9

1-3Q-9 Tachylone: サイクルー貫学習に基づくニューラルボコーダの話者適応による低遅延リアルタイム声質変換

Tachylone: Low latency real-time voice conversion by speaker adaptation of neural vocoder based on cycle-consistent training

☆今井 柊平, △金垣 葵, 能勢 隆, 府川 省吾, 伊藤 彰則(東北大)

◆本研究では，事前学習したニューラルボコーダを話者適応することで声質変換に応用する手法を検討した．

◆CycleGAN に基づく従来の声質変換手法では，音響モデルとボコーダを別々に学習し，推論時に接続することで変換音声を得る．

◆提案手法では，複数話者コーパスで事前学習したニューラルボコーダをサイクルー貫学習に基づき End-to-End に話者適応する．

◆変換過程が簡略化されることによる低遅延化と，End-to-End 学習による高品質化が可能．

◆実験から，提案手法は音韻性に関して改善の余地があるものの，HiFi-GAN V1 に基づく Tachylone V1 は従来法よりも自然性・話者性の観点で優れることを示した．

Table 1: Results of subjective evaluation and real time factor.

model	MOS ↑		RTF ↓
	Naturalness	Speaker similarity	(CPU 1core)
Natural	4.72 ± 0.14	-	-
CycleGAN-VC2 + HiFi-GAN V1	3.69 ± 0.21	3.58 ± 0.22	2.155
Tachylone V1	3.77 ± 0.20	3.80 ± 0.20	1.717
CycleGAN-VC2 + HiFi-GAN V2	3.16 ± 0.20	3.41 ± 0.24	0.595
Tachylone V2	2.90 ± 0.21	3.51 ± 0.22	0.145

1-3Q-10

1-3Q-10 Dual Diffusion Implicit Bridges を用いた話者間の匿名性を担保した声質変換

Privacy protected voice conversion based on Dual Diffusion Implicit Bridges

☆奥田耕平, 岸田拓也, 中鹿亘(電通大)

◆背景：声質変換において学習データとなる音声の発話内容に秘匿にしたい情報が含まれる時、モデルの学習時や音声の変換時に変換話者が自分の音声データを他者に非公開にしたい。

◆従来手法：学習時に変換話者ペア両方の音声データを必要とする
or 変換時に未学習話者の音声データを必要とする。

◆目的：変換元話者のモデルと目的話者のモデルを拡散確率モデルで別々に学習することで、学習時にも変換時にも話者間の匿名性を担保できる声質変換の手法を提案する。

◆実験：提案手法と従来手法の客観的評価実験の指標としてメルケプストラム歪み(MCD)を用いた。

◆ 結果：女性から男性への声質変換で従来手法と同程度の品質となった。

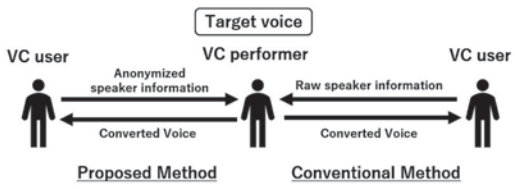


Fig.1 : An example of voice conversion setting on proposed method and conventional method

1-3Q-11

1-3Q-11　合成音声の主観評価結果の統計的解析

Statistical analysis of listening test of synthetic speech

◎安田裕介, 戸田智基 (名古屋大学)

近年の合成音声の品質の向上や品質の自動評価法の発展により、主観評価結果そのものの解析の重要性が増している。本稿では、合成音声の主観評価における統計的解析を行う。標本平均が真の期待値から大きく外れる裾確率を評価し、MOS（Mean Opinion Score）の信頼区間と標本数の関係を調査する。裾確率の評価法では、近似に基づく方法と上界に基づく方法を導入する。実験では、仮想データと実際の聴取試験の結果を解析する。仮想データでは合成音声の品質差はクラウドソーシングで確保可能な標本数の限界に達しつつあることを示し、実際のデータでは標本数の小ささのために信頼区間が被っており有意な差がないシステムが多いことを示す。まとめでは今後の主観評価の方向性を提案する。

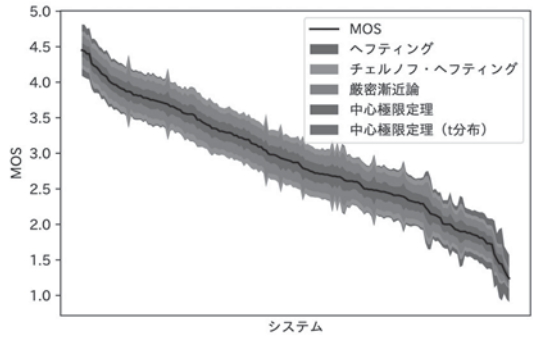


図1: VoiceMOS チャレンジの１７５システムの９５%信頼区間。

1-3Q-12

1-3Q-12　歌唱音声における自然性評価のための基本周波数とスペクトルの微細変動増減処理によるビブラート音声の作成

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

◎鈴木千文, 坂野秀樹, 旭 健作(名城大)

◆本研究ではスペクトルの各ビンの時系列に含まれる微細変動の減衰処理と増幅処理によるビブラート合成音の作成を行った

◆ビブラート合成音作成に用いる指標の検討

- 基準フレームをビブラート終了区間のフレームとした場合のメルケプストラム歪みを算出
- ビブラートの終了区間では、メルケプストラム歪みの値が大きく、スペクトルの形状が大きく変化していることが分かった

◆スペクトルの各ビンの時系列に含まれる微細変動や基本周波数の微細変動に対し減衰処理や増幅処理によってビブラート合成音を作成

◆微細変動に対して減衰処理を行った結果

- 重みを小さくしてもビブラート合成音はほとんど変わらない印象
- 微細変動が聴感上与える影響はほとんどないと考えられる

◆微細変動に対して増幅処理を行った結果

- スペクトルの微細変動に対する結果
  - ✧ 2.1～3.0 まで重みを変化させるとビブラート合成音の自然性が低下する印象
- 基本周波数の微細変動に対する結果
  - ✧ 3.0 まで重みを変化させてもビブラート合成音の自然性への影響はほとんどないと考えられる

1-4P-1

1-4P-1　音環境が持続負荷作業に与える影響～脳波による特徴分析～

Effect of sound environment on sustained workload ~ characteristic analysis by EEG ~

☆大門丈流, 大隅歩, 伊藤洋一（日大・理工),
△栗原幸大, 高田誠（三菱電機）

◆オフィスワークなどの活動は，温度・湿度・照度等の室内環境に加え，適切な音環境を提供することが重要であり，知的生産活動の効率上昇が期待される。

◆無意味音を用いた音環境下で，オフィスワークを想定した持続負荷作業時の High α 波の脳波特性について検証した。

◆その結果，音環境の種類によってその出現傾向が一定の被験者，音環境やそれ以外の要因により出現傾向が大きく変化する者がいることが分かった。

◆また，脳波の出現パターンを５つに分類して評価したところ，作業の集中度とある程度の関連性があることが認められた。

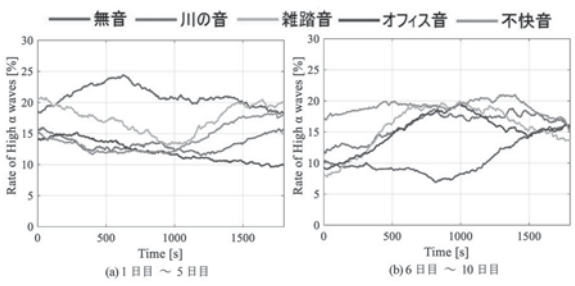


Fig.1 : Rate of High α waves

1-4P-2

1-4P-2　聴覚時間分解能検査の作成：(4)振幅変調検出課題のパラメタの検討

Test of human auditory temporal resolution: (4) Examination of parameters for amplitude modulation detection task

○森本隆司(リオン), 村田悠登(九州大院・情報理工), △岡本康秀(慶応大), 蝦名俊匡(リオン), △神崎晶(慶応大), 森周司(九州大・情報学)

◆短時間で測定可能な時間分解能測定法の確立を目指し，振幅変調検出閾値測定に対して ZEST を適用することを検討している。

◆これまでの検討では，ZEST の初期確率密度関数には正規分布形を仮定してきたが，実際の健聴者と難聴者を合わせた閾値分布は２峰性であることがわかっている。

◆本研究では，初期確率密度関数のパラメタ更新を目的に，健聴者，難聴者の変調周波数８Hz の振幅変調検出閾値分布の検討を行った。

◆健聴者 122 名，難聴者 23 名に対して実際に測定を行い，カーネル密度推定法によって閾値分布を求め，modified hyperbolic secant 及び正規分布でフィッティングを行った。

◆健聴者，難聴者の結果は，どちらのモデルでもよく当てはまり，特に難聴者の場合で左右非対称の分布形となった。

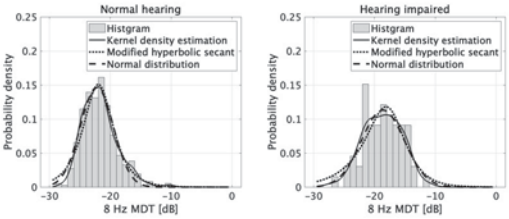


Fig.1 : Threshold distributions of amplitude modulation detection of 8-Hz modulation frequency

1-4P-3

1-4P-3　聴覚時間分解能検査の作成：(5) 無音検出課題のパラメタの検討

Tests of human auditory temporal resolution:
(5) Examination of parameters for gap detection task

☆村田悠登(九州大)，森本隆司(リオン)，△岡本康秀(慶應大)，蝦名俊匡(リオン)，△神崎晶(慶應大)，森周司(九州大)

◆聴覚時間分解能の指標の１つである無音検出閾値の測定時間短縮のため ZEST による測定を検討している。

◆ZEST は閾値測定法の一種であり、実験開始前に閾値の事前推定を確率密度関数としてパラメタ設定することで、短時間での測定を可能としている。

◆健聴者 113 名と難聴者 23 名のデータそれぞれに modified hyperbolic secant と正規分布をフィッティングした結果、どちらに対してもよく当てはまった。

◆今後は本研究で得た健聴者と難聴者の閾値分布を組み合わせた２峰性の閾値分布を作成し、コンピュータシミュレーションによる評価を実施する。

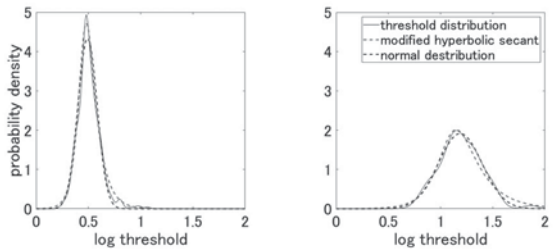


Fig.1：Probability density function
[Normal hearing (right), Hearing impaired (left)]

1-4P-4

1-4P-4　ピッチ不安定性強調現象におけるピッチ知覚メカニズム ー聴取実験による検討ー

* Pitch Perception Mechanism:
Analysis of Pitch Instability Enhancement Illusion
by Psychological Experiment

☆近森正二郎，饗庭絵里子(電通大)

◆ピッチ不安定性強調現象(PIEI)は変化音程が半音より小さい場合などに生じ，音の変り目で音の高さが不安定に知覚される現象である.

◆PIEI への予測や順応といった高次の聴覚情報処理過程の寄与を検討することを目的とし，複雑な音系列を実験刺激として用いる心理評価実験を行った.

◆童謡のチューリップをベースとしたメロディ音列と、１半音を三分割した音系列をランダムに並べた音列の２種類を呈示し，どの音で音の高さが変化したかを回答させた.

◆先行音との音高差 100 Cent 未満の音で不安定に感じられる傾向が見られた．メロディ音列では半音の整数倍差であっても，旋律から逸脱した場合に不安定に知覚された.

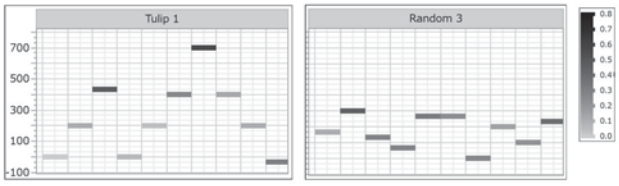


Fig.1 Instability for sound in a sound sequence

1-4P-5

1-4P-5　骨導提示音による気導提示音の抑圧効果

Study on the suppression effect of the air-conducted tone
by the bone-conducted tone

☆井上隼輔(北陸先端大)，鳥谷輝樹(金沢大・理工)，
上江洲安史(北陸先端大)，鵜木祐史(北陸先端大)

◆背景：現在の騒音抑圧法は耳を塞いだ状態で雑音を抑圧するため，抑圧量が大きいが，聞き逃してはいけない音声も同時に抑圧してしまう．
→耳を塞がずに騒音のみを抑圧が可能な騒音抑圧法の実現の期待.

◆目的：骨導デバイスによる騒音抑圧法の実現の基礎検討として，骨導提示音により，気導提示音を知覚的に抑圧できるか明らかにすること.

◆方法：６種類の純音（0.25, 0.5, 1, 2, 4, 8 kHz）について，気導提示音を最も抑圧する骨導提示音の振幅・位相を極限法で決定．得られた振幅・位相条件の骨導提示音による気導提示音の抑圧量を調査.

◆結果：平均で 4.95 dB の抑圧効果を確認（個人間でばらつきあり）．抑圧効果は周波数に依存.

◆結論：各周波数における純音の振幅・位相を制御して骨導提示することで，気導提示音を知覚的に抑圧できることが分かった.

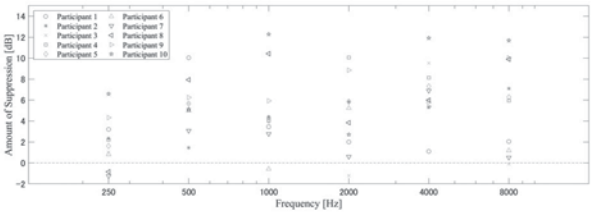


Fig.1 Amount of suppression for AC tone by BC tone

1-4P-6

1-4P-6　周波数と空間的位置を考慮した聴覚オブジェクトの知覚に関する検討

The effect of frequency and sound position
on the perception of auditory objects

○堀田尭八，坂本修一(東北大通研/院情科研)

◆本研究では，音脈分凝における周波数と空間的位置の要因の影響を検討した．実験では，２種類の 1/3 オクターブバンドノイズ A，B を用い，試行ごとに B 音の中心周波数，及び２音源間の角度を変化させて音列を再生した．聴取者は，各刺激を１つの音脈に知覚したかどうかを回答した.

◆その結果，単一のスピーカで音列を提示した時，周波数差に関わらず多くの場合１つの音脈として知覚された．このことから，音脈を判断する要因として，周波数よりも空間的位置の方が影響が大きいと考えられる．また，１つの音脈として知覚される角度の限界（=許容限）は，周波数差が大きくなるにつれて減少した（Fig. 1）．加えて，周波数差が小さいときのみ，前方と比べて側方（30° 左側）で許容限が有意に大きかった．この結果は，前方と側方では，音脈知覚に対する空間情報の影響が異なることを示唆している.

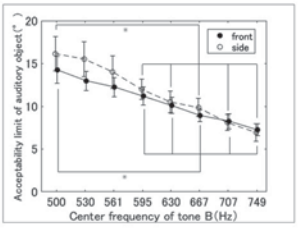


Fig. 1:Acceptability limit of auditory stream at each center frequency.

1-4P-7

1-4P-7　若年者と中高年者の競合音存在下での聴取能力の違いとその要因−ワーキングメモリ容量，および，時間情報処理能力の比較から

Differences in listening performance between young and middle-aged adults in the presence of competing sounds and their factors – a comparison of working memory and temporal information processing ability

☆湯淺　麻生，大塚　翔，中川　誠司（千葉大）

◆中高年期には，競合音存在下での聴取能力が低下することが報告されている．本研究では，若年者と中高年者のエネルギー・情報マスキング下での聴取成績を比較し，各マスキング下の聴取に対する時間情報処理能力，及び，ワーキングメモリ容量の寄与を調べることで，中高年期の競合音存在下での聴取能力の低下の原因の解明を目指す．

◆妨害音の種類に依らず中高年者の方が若年者に比べて聴取成績が低かった (Fig. 1)．中高年者の競合音存在下での聴取能力の低下が確認された．

◆中高年者群を対象とした相関分析から，情報マスキングの効果が大きい同話者条件では，ワーキングメモリ容量の寄与が大きいのに対して，エネルギーマスキングの効果が相対的に大きい条件（妨害音がノイズ，及び，異性・同性話者条件）では，時間情報処理能力の寄与が大きいことが分かった．

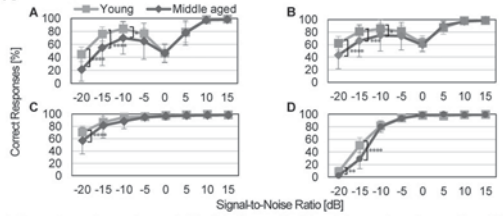


Fig. 1 Percentage of correct speech identifications with different types of maskers: voice of the same talker as the target voice (A), voice of different talker of the same sex as the target voice (B), voice of a talker of the opposite sex of the target voice (C) and noise (D). *p<.05, **p<.01, ***p<.005, ****p<.001 (tested by the simple main effect test).

1-4P-8

1-4P-8　耳介血腫が軟骨伝導の検出閾特性に及ぼす影響

Effects of auricle hematoma on detection threshold of cartilage conduction

☆田村　茜，大塚　翔，中川　誠司（千葉大）

◆骨伝導の装用性に係る問題を解決する方法として ”軟骨伝導” が提案され，補聴器やスマートホン，イヤホンなどに応用されている．

◆耳介血腫は柔道やラグビーの競技者に頻発し，イヤホン等の使用の妨げになる場合が多い．軟骨伝導デバイスが有効であると考えられる一方，耳介の硬度や質量が変化するために，軟骨伝導知覚特性が変化する可能性が大きい．

◆耳介血腫が軟骨伝導知覚に及ぼす影響の調査を目的とし，耳介血腫保有者 (Hematoma) および聴覚健常者 (Normal) の耳介の硬度・形態および軟骨伝導検出閾を計測し，両者の比較検討を行った．

◆多くの耳介部位において，Hematoma の耳介硬度は Normal に対して増加した．サイズの差異は確認されなかった (Table 1).

◆Hematoma の検出閾は Normal に対して 125-1000 Hz で上昇し，2000 および 4000 Hz で低下した (Fig. 1).

◆これらの結果は，血腫により耳介の硬度や質量が増加し，軟骨伝導検出閾に影響を及ぼすことを示している．

	部位	平均±標準偏差 Male	Hematoma
硬度	対輪上部	43.00±6.33	47.05±5.94
	対輪下部	37.88±7.89	51.70±8.68
	耳輪	25.05±6.62	31.75±12.43
	耳珠	42.53±6.88	40.39±7.20
	対珠	31.23±6.69	36.75±5.52
形態	耳長	6.40±0.38	6.20±0.41
	耳幅	3.65±0.27	3.52±0.27
	耳介付着部長	5.59±0.43	5.33±0.30
	耳介軟骨長	5.33±0.31	5.13±0.35
	外耳道までの距離	3.34±0.31	3.24±0.17
	耳垂長	1.89±0.19	1.90±0.22

Table 1: Auricle hardness and sizes (mean±S.D.).

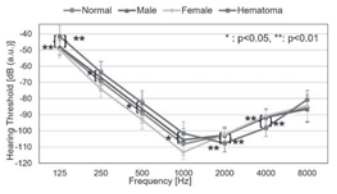


Fig.1：Hearing threshold of cartilage conduction for each subject group.

1-4P-9

1-4P-9　Paired-click 刺激法による蝸牛シナプス障害の評価 —ヒトの蝸牛神経複合活動電位による検討—

Evaluation of cochlear synaptopathy by paired-click stimulus −Assessing the action potential in human−

○藤平晴奈（NTT CS 研・日本学術振興会特別研究員 RPD），山岸慎平，古川茂人，柏野牧夫（NTT CS 研）

◆聴力閾値が正常であるにもかかわらず，ノイズ下での聞き取りが困難になる“隠れ難聴”の一要因として蝸牛シナプス障害が挙げられる．

◆蝸牛シナプス障害とは，強大音の暴露や加齢によって引き起こされる内耳器官の内有毛細胞と蝸牛神経との間のシナプスにおける障害のことであり，その障害は自発放電頻度の低い蝸牛神経 (low-spontaneous discharge rate auditory nerve: LSR-AN) において選択的に生じると報告されている．

◆本研究では，LSR-AN の活動を強調して捉えることができると考えられる paired-click 刺激法による蝸牛神経複合活動電位に着目した．若年者 27 名，中年者 24 名を対象に，この手法による測定値がヒトにおける蝸牛シナプス障害の評価指標となりうるか検証した．

◆Paired-click 刺激の 2 音目に対する反応波形における，LSR-AN の活動を強調すると考えられる解析区間の実効値は，クリック間間隔が短い条件で，中年者が若年者に比べて有意に低い値を示した．さらに，その値はノイズと残響下での聴取成績ともやや強い相関を示した．

◆これらの結果から，paired-click 刺激を用いた 2 音目に対する反応，特にクリック間間隔が短い条件における実効値が，蝸牛シナプス障害の評価指標となりうることが示唆された．

1-4P-10

1-4P-10　Paired-click 刺激法による蝸牛シナプス障害の評価 —計算機シミュレーションによる検討—

Evaluation of cochlear synaptopathy by paired-click stimulus –A study using computer simulation –

○山岸慎平（NTTCS 研），藤平晴奈（NTTCS 研・日本学術振興会特別研究員 RPD），古川茂人，柏野牧夫（NTTCS 研）

◆聴力閾値で正常であってもノイズ下での聞き取りが困難になる“隠れ難聴”の要因として，強大音暴露や加齢による内有毛細胞と蝸牛神経間のシナプス障害（蝸牛シナプス障害）が注目されている．

◆蝸牛シナプス障害は低い自発発火頻度の蝸牛神経（low-spontaneous firing rate auditory nerve: LSR-AN）に選択的に生ずることが報告されている．

◆1-4P-9 において，ヒトを対象に paired-click 刺激に対する蝸牛神経複合活動電位を計測することによって，2 音目のクリックに対する反応が中年者と若年者の間で異なること，および雑音下聴取成績と相関することを報告している．これらの効果はクリック間間隔が短い条件（1.5 ms, 2 ms）でのみ観測された．この結果は LSR-AN が先行音の影響を受けづらい特性を反映したものであると解釈できる．

◆本研究では，計算機シミュレーションによって蝸牛神経活動を再現し，LSR-AN と HSR (high-spontaneous firing rate)-AN の比率や LSR-AN の数を制御することによって，上記の解釈の妥当性を検討した．

◆シミュレーションの結果，ヒトを対象にした実験と同様，クリック間間隔が短い条件では LSR-AN の数や割合が大きいほど先行音の影響が小さくなることが示された．

1-4P-11

1-4P-11　精神疲労がオリーブ蝸牛束反射に及ぼす影響–計算課題の難易度との関係

Effects of Mental Fatigue on Medial Olivocochlear Reflex and its Dependence on Task Difficulty

☆内山寛大，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へ接続した遠心性神経束であるオリーブ蝸牛束は，強大音に反応し外有毛細胞の増幅を抑制することで内耳を保護している．この反応をオリーブ蝸牛束反射 (MOCR) といい，MOCR 強度の計測よって騒音性難聴のリスクを事前評価できる可能性がある．しかし MOCR は計測毎のばらつきが大きく，安定したリスク評価のため MOCR の変動の要因の解明が必要である．

◆ MOCR は認知機能の変動の影響を受けることが報告されている．しかし，認知機能の低下の一因である精神疲労と MOCR 強度の関係の体系的な解明には至っていない．

◆ 本研究では，疲労度合いを計算課題の難易度により調節し，MOCR 強度との関係を調べた．さらに，脳波計測により皮質レベルの活動を推定し疲労に伴う MOCR の変動のメカニズムの解明を試みた．

◆ MOCR は耳音響放射の測定により評価した．精神疲労は，フリッカー検査 (CFF)，自覚症アンケートにより評価した．脳波は聴性誘発電位を測定し，N1-P2 振幅の算出によって評価した．

◆ 全難易度の課題前後で CFF は減少し，自覚症アンケートのスコアは増大した．MOCR，N1-P2 振幅は課題前後で有意な結果は見られなかったものの，MOCR と N1-P2 振幅の疲労課題前後の変動量には，逆の変動パターンを示した(Fig .1)．適度な精神負荷ではメンタルエフォートが増大することが報告されている．Level 2 での N1-P2 振幅（皮質活動）の増加は，このメンタルエフォートの増加を反映している可能性がある．この考察を踏まえると MOCR 強度は，皮質レベルのメンタルエフォートの増加によって減少し，過度な精神負荷に伴う精神疲労によって増加する可能性が推察される．

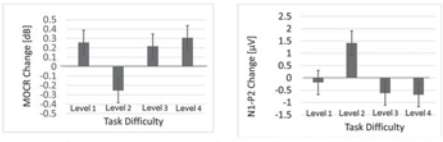


Fig.1 MOCR and N1-P2 amplitude changes after calculation tasks with different difficulty level

1-4P-12

1-4P-12　注意に伴うオリーブ蝸牛束反射の変動の個人差と皮質活動との関係

Relation between individual differences in attention-related changes of medial olivocochlear bundle reflex and cortical activities

☆鈴木麻里子，石坂勇毅，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞に投射するオリーブ蝸牛束は，音響刺激に反応し，外有毛細胞の増幅を抑制するはたらきがある．この抑制反応は，オリーブ蝸牛束反射(MOCR)と呼ばれ，雑音下での聴取を改善する役割を果たすと考えられている．それゆえ，注意状態の変化に伴う MOCR の変化は雑音下での聴取に影響を与えると推察される．

◆ 著者らはこれまでに，妨害課題を課すことで，聴取課題に配分される注意資源を減少させると，聴取成績，及び，MOCR 強度が減少することを示したが，MOCR 強度の減少量の個人差は大きかった．

◆ 本研究では，注意に伴う MOCR の変動の個人差の要因を解明することを目的に，妨害課題負荷時の皮質活動と MOCR との関係を調べた．

◆ 妨害課題負荷に伴う聴取成績の減少量と MOCR 強度の減少量の間には統計的な有意な相関は見られなかったものの，MOCR 強度が減少した人ほど，聴取成績が減少する傾向が見られた．PLV の減少量と MOCR 強度の減少量との間には，θ帯域において有意な負の相関が見られた (Fig. 1)．

◆ 妨害課題負荷によって，θ帯域に関連した皮質レベルの処理が劣化すると，逆に MOCR が増強してその劣化を補う可能性が推察された．

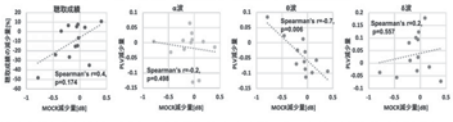


Fig. 1 Relationships of the disturbance-task-related changes in MOCR strength, listening performance and neural oscillation phase-locking values (PLVs)

1-4P-13

1-4P-13　先行音列の規則性に基づく予期によるオリーブ蝸牛束反射の変動が雑音下での聴取に与える影響

The effect of the medial olivocochlear reflex enhancement associated with temporal expectation on listening performance in noise

○石坂勇毅、大塚翔、中川誠司（千葉大）

◆ 脳幹から外有毛細胞に遠心性接続するオリーブ蝸牛束が，音に反応して外有毛細胞の増幅を抑制する働きをオリーブ蝸牛測反射(MOCR) という．

◆ 著者らはこれまでに，先行音列にジッターを加えると，MOCR 強度及びδ波の位相同期が減少し，両者のジッターに対する依存性が類似することを示した．

◆ さらに前報では，ワーキングメモリ (WM) を要する妨害課題を課すと，予期に伴う MOCR の増強が失われることを報告した．

◆ 本研究では，予期に伴う MOCR の増強により雑音下での聴取成績が向上するかを調べるために，妨害課題負荷時の雑音下での聴取成績，MOCR 強度，及び，δ波の位相同期度 (phase locking value: PLV)を比較した．

◆ 妨害課題の負荷が小さい場合には，ジッターが小さいほど，MOCR 強度及びδ波の PLV が増加し，聴取課題成績も上昇した．一方で，妨害課題の負荷量が大きい場合には，MOCR 強度，δ波の PLV 及び聴取課題正答率が減少するとともに，ジッターに対する依存性が失われた (Fig. 1)．

◆ これらの結果は，皮質領域での時間的予期と関連した MOCR の増強が雑音下での聴取を改善する可能性があること，さらに，その増強と聴取改善には聴覚的注意が必要であることを示している．

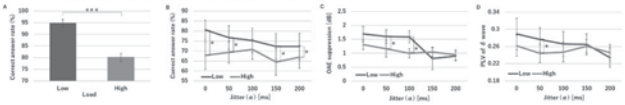


Fig. 1 (A) Mean correct answer rate of working memory task as a function of task load. Mean correct answer rate of auditory task (B), mean OAE suppression (C) and mean PLV of δ wave(D) as a function of jitter (α). *p<0.05, ***p<0.001 (T test without any correction)

1-4P-14

1-4P-14　Distinguishing conspecific bats by their echolocation calls using a convolutional neural network

Andreas Mayer, Soshi Yoshida, Olga Heim, Keisuke Imoto, Shizuko Hiryu
(Doshisha University)

◆As species distinction of bats with **convolutional neural networks** (CNN) has already been implemented, we utilize this method to **distinguish between individuals of the same species.**

◆Spectrograms of calls from bat species *Miniopterus fuliginosus* and *Pipistrellus abramus* were investigated (Fig. 1).

◆Success rates for conspecific *Miniopterus fuliginosus* individuals exceeded 80% consistently.

◆Pair flight of *M. f.*, compared to flying alone, had a decreased performance which may give a hint to understanding jamming avoidance behavior in bats.

Table 1: Neural network classification results for solo flight experiments.

Bat species	F1-score	Accuracy
M. fuliginosus	0.842	0.862
P. abramus	0.652	0.661

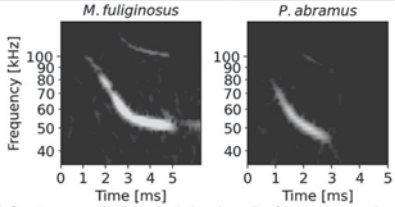


Fig.1 : Spectrograms of individual echolocation calls of the two bat species used.

1-4P-15

1-4P-15　Deep Prior を用いた脳波からの Forward model と電流源の同時推定

Simultaneous estimation of forward model and current source from EEG using Deep Prior

☆北条直樹，矢野肇，高島遼一，滝口哲也（神戸大），中川誠司（千葉大）

◆本研究では，Deep Prior を用いて，**EEG のみから脳内の電流分布を推定する手法**を提案する．

◆従来手法では被験者の負担の大きいMRI の測定を必要としていたが，提案手法では電流源と Forward model を推定する２つの CNN を用いることで，**被験者本人の MRI データを用いずに電流源を推定**する．

◆提案手法を用いることで，被験者本人の MRI データを用いた場合と**同等の電流源推定位置精度**を得ることができた．

◆また，Deep Prior によるノイズ除去の効果により，ノイズありの EEG 観測値から電流源を推定した場合でも，**ノイズなしの EEG 観測値を復元**することができた．

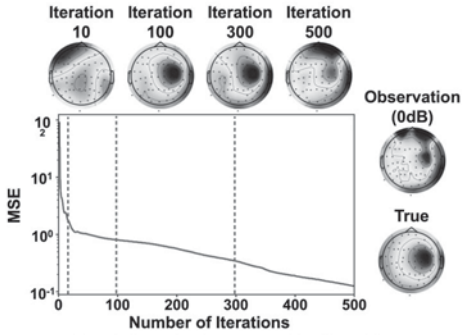


Fig.1 Effect of noise removal by Deep Prior

1-4P-16

1-4P-16　自己正確推定損失を用いた Deep Prior に基づく複数の脳内電流源の推定

Multiple source localization in the brain using Deep Prior with self exact localization loss.

☆矢野　肇，高島遼一，滝口哲也（神戸大），中川誠司（千葉大）

◆以前の発表（2022 年秋）で，提案した自己正確推定損失は，ニューラルネットワークが表現する事前分布（Deep Prior）に基づく脳内の電流源推定を安定化させる効果がある．

◆この手法は脳内に１つの電流源を想定する場合には有効だが，複数の電流源がある条件下ではうまく機能しない．

◆本発表では，脳内に複数の電流源がある場合に，自己正確推定損失を用いて安定して推定を行うための方法を検討する．

◆シミュレーション実験によって，電流源の数と同じ数の異なるネットワークを設け，ネットワークごとに自己正確推定損失を計算することが安定した推定に有用である可能性が示唆された（Fig. 1）．

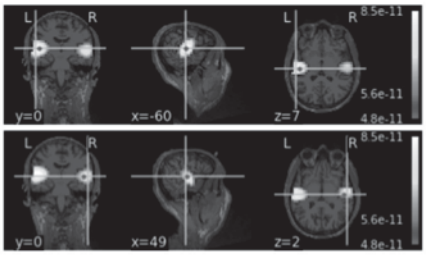


Fig. 1 Estimated current density at the left auditory cortex (top) and the right auditory cortex (bottom) by the proposed method.

1-5P-1

1-5P-1　高槻城公園芸術文化劇場の音響設計

Acoustical design of Takatsuki Arts Theatre

○中川浩一（日建設計）

◆高槻城公園芸術文化劇場は 1508 席の大ホール，205 席の小ホール，大・中・小の計 11 室のスタジオ等から成る複合文化施設である。本稿ではホールを中心に遮音計画，室内音響計画について報告する。

◆各室を独立した平面配置とし，間の通路や諸室が遮音的なバッファーゾーンとなることを意図している。また，小ホールと，スタジオの内，大音量発生が想定される室に防振遮音構造を採用した。

◆地場産の杉材を建物の内外装全体に使用し，大ホールは木キューブ，小ホールおよびスタジオは木リブにより，音の散乱を意図した。

◆大ホールの天井・壁面に設置した木キューブは，見付寸法 120mm 角，奥行寸法３種類（24，48，72mm）をランダム配置した。木キューブの吸音性能は，施工時に試験体を作成し，残響室法吸音率測定を行い確認した。また，現場のモックアップを利用して共振振動の減衰時間を確認し，木キューブの取付方法を検討した。

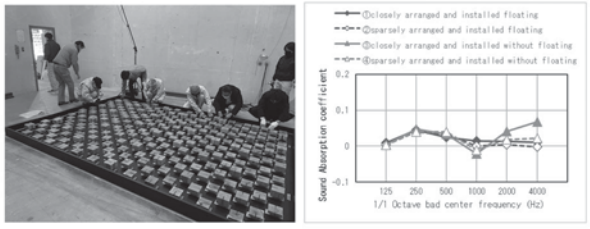


Fig.1：Test specimen of wooden cube and sound absorption coefficient

1-5P-2

1-5P-2　東京文化財研究所「実演記録室（スタジオ）」の改修における音響設計

Acoustical Design of a recording booth for traditional Japanese musical instruments at Tokyo National Research Institute for Cultural Properties.

◎松尾明穂，甲斐智鶴，土倉律子，中原雅考（ソナ），小口恵司（永田音響設計）

◆和楽器の収録に特化した録音ブース「東京文化財研究所 実演記録室（スタジオ）」が，建設以来 22 年振りに改修された。

◆改修は，遮音層や空調及び電気設備は既存利用とし，既存下地を活かした条件で仕上げを全面リニューアルすることとなった（Fig.1）。

◆本報では，改修に際しての音響要望とそれに対する音響設計，試奏による仕上げの微調整や音響可変効果の確認（Fig.2），改修前後における音響測定結果などを紹介する。

Fig.1：Refurbished new studio for recording traditional Japanese instruments.

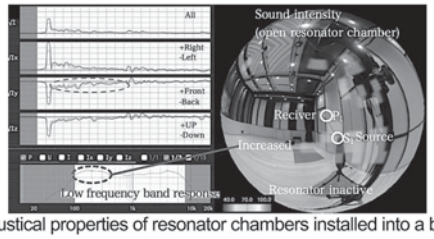


Fig.2：Acoustical properties of resonator chambers installed into a back wall.

1-5P-3

1-5P-3 Study on Blind Estimation of Room Acoustic Parameters from Speech Signals Based on Extended Model of Room Impulse Response

☆Lijun Wang, Suradej Duangpummet, Masashi Unoki (JAIST)

◆**Objective:** To propose a deterministic method for blindly estimating room acoustic parameters (RAPs) from a reverberant speech signal.

◆**Problems:** Insufficient description of a single estimated RAP or objective index for room acoustic characteristics of a sound field.

◆**Solution:** Five RAPs (i.e., T_{60}, *EDT*, C_{80}, D_{50}, and T_s) and STI are blindly estimated to describe the room acoustics comprehensively using an analytical method, as shown in Fig.1.

◆**Dataset:** ATR dataset (ten long Japanese sentences uttered by ten speakers, five males and five females).

◆**Measurements:** RMS errors and Pearson correlation coefficients.

◆**Evaluated Conditions:** Real reverberant environments from the SMILE RIR dataset, which contains 43 realistic RIRs.

◆**Results:**

Table I: Estimation accuracy of the proposed method.

	T_{60}	EDT	C_{80}	D_{50}	T_s	STI
RMS Err. (pro)	0.067	0.256	2.309	14.303	0.052	0.037
Pearson Corr. (pro)	0.993	0.945	0.794	0.680	0.797	0.908

◆**Conclusion:** The proposed method can blindly estimate RAPs and STI effectively without training.

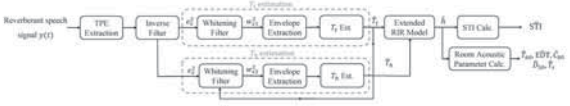


Fig.1: Block diagram of the proposed method.

1-5P-4

1-5P-4 アンサンブル平均を利用した材料の吸音特性の *in-situ* 測定法に関する研究

—微細穿孔板（MPP）を対象とした音響管法との比較—

In-situ sound absorption characteristics measurement method by ensemble averaging technique-Comparison with impedance tube method for microperforated panel absorbers (MPP)-

☆田原有彩（大分大院・工学研），岡本則子，富来礼次，
大鶴徹（大分大・理工），池田夏菜（大分大院・工学研）

◆筆者らは，アンサンブル平均を利用した材料の吸音特性の *in-situ* 測定法（EA 法）を提案している

◆微細穿孔板（MPP）を対象に，背後条件を変化させ，EA 法および音響管法で吸音特性の測定を行った

◆MPP の背後条件は，空気層の場合，ハニカムの場合，多孔質吸音材を設置した場合とし，条件毎に EA 法と音響管法の測定値を比較した

◆MPP の背後空間にハニカムおよび多孔質吸音材を充填した場合，EA 法の吸音率は音響管法や理論値に近似した（Fig. 1，Fig. 2）

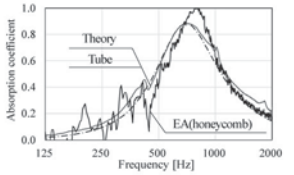


Fig. 1：Comparison of absorption coefficients for MPP with a honeycomb between EA, tube method and theory.

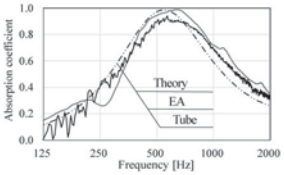


Fig. 2：Comparison of absorption coefficients for MPP with a glass wool between EA, tube method and theory.

1-6P-1

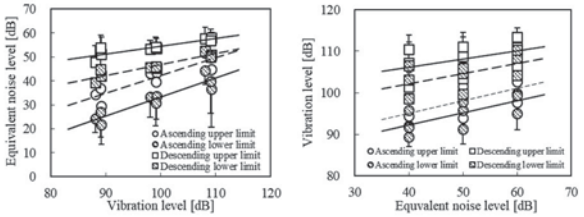
1-6P-1　車両振動を対象とした振動感覚に及ぼす音の影響に関する研究

☆西尾恒紀，笹森咲依，佐々木亮弥（日本大院・理工），
松田礼，町田信夫（日本大・理工）

◆乗物の振動環境において，人間は振動と音が同時に暴露される複合環境下にある。しかし，乗物を含む振動評価規格は振動のみの評価であり，音の影響を考慮した評価基準は定められていない。

◆本研究の目的は，振動と音を同時に暴露したときの振動と音の大きさの優位性を調べることである。本研究では，振動条件を固定して音量を調整する実験と，音量を固定して振動加速度レベルを調整する実験の２種類の実験を行った。

◆振動と音の優位性の判断が不明瞭になる条件は，等価騒音レベルと振動レベルには比例関係がみられ，音量は 20～30dB 程度，振動レベルは 15dB 程度の幅がみられた。



(a) Volume adjustment experiment　(b) Vibration adjustment experiment

Fig. 1: Relationship between vibration level and equivalent sound level (WN, lateral vibration)

1-6P-2

1-6P-2　水平振動の振動感覚に及ぼす音楽の影響に関する研究

Effect of music on vibration sensation of horizontal vibration

☆笹森 咲依，西尾 恒紀，佐々木 亮弥（日本大院・理工），
松田 礼，町田 信夫（日本大・理工）

◆自動車や鉄道車両等の車内環境は，音と振動が同時に存在している。しかし，全身振動に対する国際評価規格では振動のみで評価され，音を含む複合環境の評価方法は定められていない。

◆本研究の目的は，水平振動を全身暴露した時の振動感覚に及ぼす音楽の心理的影響を明らかにすることである。本研究では，振動と音楽を同時暴露した時の振動と音の優位性と，振動の不快感について心理アンケートを用いて測定した。

◆不快に感じる振動環境下では，振動と音の優位性が不明瞭な音量でも音楽よりも振動の方が優位に感じやすくなる傾向が見られた。

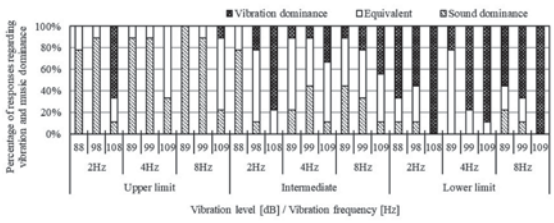


Fig.1：Percentage of responses regarding vibration and music dominance (Lateral vibration)

1-6P-3

1-6P-3　振幅変調低周波音の感覚特性と生体反応に関する研究

Sensory characteristics and biological responses of amplitude-modulated low-frequency sound

☆伊藤優花，吉原山人イスル（日本大院・理工），
松田礼，町田信夫（日本大・理工）

◆周波数 100Hz 以下の低周波音による騒音が引き起こす低周波音問題は，定常性低周波音による騒音を対象とした評価指標がある。しかし，風力発電用風車等から発生する変動性低周波音の評価方法は確立されていない。

◆本研究の目的は，変動性低周波音を構成する物理量と人間の心理・生理反応との関係を実験により調べ，変動性低周波音の人体影響を明らかにすることである。本報では，評定尺度法と ME 法による心理アンケート，及び心電図を用いて生体反応の測定実験を行った。

◆20Hz 以外の変動音は，多くの条件でΔRRI が 1 以下であることから，より交感神経優位，つまりストレス状態であると読み取れる。

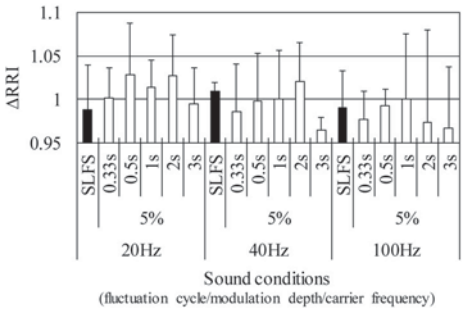


Fig.1 : Changes in RRI ratio by low-frequency sound exposure

1-6P-4

1-6P-4　低周波の吸音特性を重視した音響メタマテリアル吸音材を用いた際の斫り騒音低減効果の評価

Evaluation of concrete chipping noise reduction by using acoustic metamaterial absorbers for low-frequency

○平良優大，五味蔵酒（ピクシーダストテクノロジーズ）
△森圭太郎（JR 東日本研究開発センター）

◆目的：　ブレーカーを用いた斫り解体工事は，多くの解体現場で採用されている。一方で騒音によりその利用を制限される場合があり，特に駅舎などでは駅利用者がいない深夜時帯など限られた時間での工事となっており，工事費用の増加や工事期間の長期化につながる要因となっている。そこで我々は低周波の吸音特性を重視した音響メタマテリアル吸音材で構成する防音壁を構築し，斫り騒音の低減を試みた。

◆方法：　200-2000 Hz で垂直入射吸音率 85%以上を満たす音響メタマテリアルを設計，製造し，音響メタマテリアルで構成する防音壁を作成した(Fig. 1 (a))。

◆結果：　低周波の吸音特性を重視した音響メタマテリアル防音壁を用いることで，従来の仮囲いと比較してより大きな騒音低減効果を示すことが確認できた(Fig. 1 (b))。

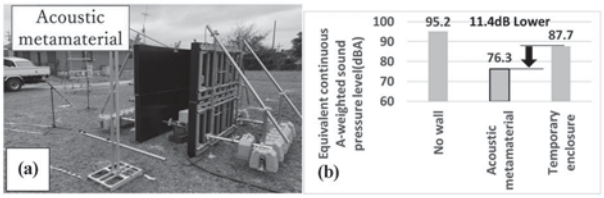


Fig. 1 (a) Photograph of acoustic metamaterial noise barrier. (b) Equivalent continuous A-weighted sound pressure level at 1m from wall.

1-7P-1

1-7P-1　高周波数による液中弾性表面波推進システムの小型化と絶縁膜 SiO_2 の性能検討

Surface acoustic wave propulsion system miniaturization with high frequency and SiO_2 insulating film evaluation

◎孔徳卿，谷村瞭(室蘭工大院)，黒澤実(東工大)，青柳学(室蘭工大院)

◆弾性表面波振動子を用いた液中推進システムは，IDT 電極に高周波電圧を印加することで，スイマーの推進力が得られた。液中音響推進システムの小型化と高推力の両立を目指すため，IDT 電極の表面に絶縁膜 SiO_2 を成膜し、音響推進力を検討する。さらに，高周波数による液中推進システムの小型化を実現し，音響推進力の周波数特性を評価する。

◆Fig. 1 に示すように，SiO_2 で IDT を絶縁した方が推力が高くなることがわかった。印加電圧が 86 V_{pp} の時，絶縁した振動子では 4 mN，絶縁していない振動子では 2.8 mN の推力が得られた。

◆Fig. 2 を示すように，入力電力は 2.2 W の場合，38.5 MHz の時に 0.83 mN/mm^2 の最大単位面積推進力を測定した。

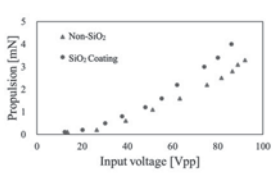


Fig. 1 Propulsion characteristics.

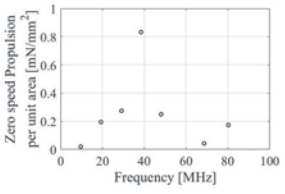


Fig. 2 Propulsion with different frequencies.

1-7P-2

1-7P-2　超音波を用いた下肢浮腫の定量診断に向けた研究 —乳脂肪含有寒天ファントムを用いた基礎検討—

Study for quantitative diagnosis of lower limb edema using ultrasound -Fundamental study using the agar phantoms containing the milk fat-

☆神戸大希，坂本眞一，神戸春輝，越山雅文，生田宴里，渡邊友美子
（滋賀県立大）

◆心不全や腎不全などの病気を早期発見するために初期状態の下肢浮腫を診断することが望まれる.

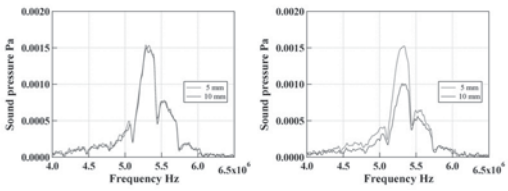
◆先行研究では浮腫と肥満の超音波による識別に課題があった.

◆乳脂肪含有寒天ファントムを製作して，脂肪と水の混在する媒質の音響特性を検討した.

◆乳脂肪の混入による超音波の減衰を確認した(Fig. 1).

◆乳脂肪含有寒天ファントムの厚みの増加に伴い減衰量も増加した.

◆減衰量は浮腫と肥満の識別に有効であることを示した.



(a) Water agar phantoms.　(b) Agar phantoms containing the milk fat.

Fig. 1　Frequency spectra of the reflected waves.

1-7P-3

1-7P-3　診断用超音波計測のための ハイドロホン感度位相特性の比較校正 —周波数が異なる複数の正弦バースト波を 並べた信号を用いる方法の有効性—

Calibration of phase frequency response of hydrophone sensitivity by comparison for diagnostic ultra-sound measurement: Effectiveness of the method using the signal aligning multiple sine-burst wave with different frequencies on a time axis

○千葉裕介，吉岡正裕（産総研）

◆診断用超音波の性能・安全性評価のためには，ハイドロホン感度の振幅だけでなく位相特性も用いた瞬時音圧の精密計測が必要である．

◆本稿では，必要な校正周波数を厳密に設定可能でかつ計測中の水温変化の影響を受けにくい比較校正法を提案し，有効性を検討した．

◆提案法で得られた感度位相特性を英国の校正機関 NPL の校正値と比較した(Fig. 1)．その結果，周波数毎の変化の傾向が概ね一致した．

◆一方，提案法による計測値は多くの周波数で NPL による校正値の拡張不確かさの範囲を超えた．今後更なる検討が必要である．

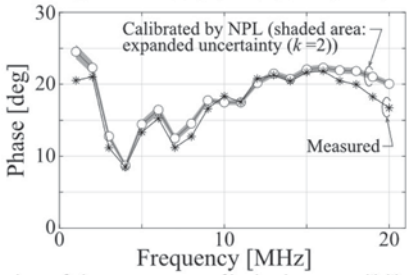


Fig. 1 Comparison of phase responses of hydrophone sensitivities between proposed method and calibration by NPL.

1-7P-4

1-7P-4　集束送信ビームを用いたベクトルドプラ法における同時送信数の検討

Investigation on multi-line transmit in vector Doppler method with focused transmit beam

☆鈴木陸斗，大村眞朗，長岡亮，長谷川英之(富山大)

◆超音波血流計測には,ドプラ効果を利用した手法が用いられている．

◆超音波ビームの送信時に MLT(Multi-Line Transmit)と呼ばれる一回当たり複数の超音波ビームを送信することで，フレームレートを向上できる．

◆本研究では,先行研究をもとに新たに作成したベクトルドプラ法に対して同時送信数を 4 まで増やした場合の計測精度への影響をシミュレーションによって示した．

◆シミュレーションでは，超音波周波数，血管の傾き，ビーム角度，最大流速についてそれぞれ条件を変えて検討を行った．

◆Table.1 に同時送信数 2 と 4 のそれぞれの誤差の最大値とその時の条件を示す．Table.1 より，同時送信を 4 まで増やしても，精度の悪化は見られなかった．

◆結果より，精度を保ったままフレームレートを上げられることが確認できた．

Table1 Maximum RMSE of 2MLT and 4MLT

MLT	frequency	Flow angle	Beam steering angle	Maximum velocity	RMSE
2	4.0 MHz	0 deg	10 deg	1000 mm/s	20.73 %
4	7.5 MHz	0 deg	30 deg	1000 mm/s	22.90 %

1-7P-5

1-7P-5　脈波速度の周波数特性に関する検討

Investigation on frequency characteristics of pulse wave velocity

☆坂口裕基，茂澄倫也，大村眞朗，長岡亮，長谷川英之(富山大・工)

◆脈波速度推定法を用いる場合，推定値は周波数に依存するため，再現性の良い計測のために周波数特性を検討する必要がある．本報告では in vivo での脈波速度の推定値と等方性動脈における脈波速度の理論値を比較する．

◆脈波速度推定を行う関心点を頸動脈の前壁に設定した．

◆理論値の算出に関する血液粘性とヤング率を変数とし，各注目周波数において推定値との誤差の和が最小となる値を決定した．

◆Fig. 1 に推定値と理論値を示す．Fig. 1 より推定値と理論値は同様の傾向であり，5 - 15 Hz 付近で安定した値を取ることがわかった．

◆結果から，局所脈波速度推定法による脈波速度推定値は 10 Hz 付近の周波数において安定した値を取る可能性が確認できた．

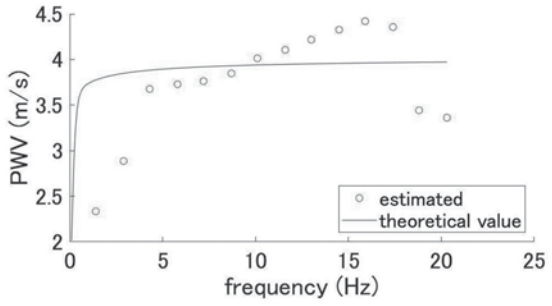


Fig. 1 Comparison of estimated and theoretical values.

1-7P-6

1-7p-6 エコーロケーション中のコウモリの頭部方向と超音波放射方向に関する検討

A study of head direction and ultrasonic emitted direction in bats during echolocation

○河村拓，△長谷川雄大，△源田祥子，
△小原大知，手嶋優風，飛龍志津子（同志社大院・生命医科学研）

◆コウモリは種によって超音波の時間周波数構造や帯域が異なることから，エコーロケーションの空間戦術にも種毎の適応が予想される．

◆本研究では，放射する超音波の方向とその際のコウモリの頭部方向をマイクロホンアレイとモーションキャプチャシステムを用いて計測し，２種類のコウモリについて結果を比較した．

◆頭部方向と放射方向の角度差を調べた結果，キクガシラコウモリ（Fig.1B）の方がユビナガコウモリ（Fig.1C）に比べ，飛行中の頭部方向と放射方向の角度差が大きいことが分かった．

◆キクガシラコウモリは超音波を放射する鼻孔の周囲に鼻葉(noseleaf)を持っており，頭部方向とは独立に超音波の放射方向の調整が可能であること考えられる．

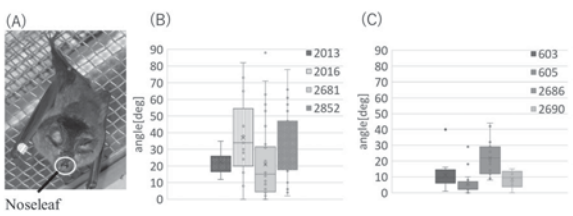


Fig.1 : (A)Noseleaf of *Rhinolophus ferrumequinum nippon*. Angular differences between pulse direction and head direction in flight of *R. ferrumequinum* (B) and *M. fuliginosus* (C).

1-7P-7

コウモリの生物ソナーデザインを模倣した自律飛行ドローンによる障害物回避飛行 —サーモホンを用いた非線形 FM 音による空中超音波ナビゲーションの実践—*

Obstacle avoidance navigation by an autonomous flying drone with simply designed active sonar system inspired by bats

◎山田恭史（広島大），△渡部佑真，△佐々木晋一，
浅田隆昭（村田製作所），飛龍志津子（同志社大）

◆自律飛行ドローンの社会普及が注目を浴びる中，非 GPS 環境下においてシンプルなデザインでの超音波センサ運用が期待されている．

◆広帯域超音波が放射可能な熱音響変換素子サーモホンをドローンへ搭載し，LPM パルスを用いたパルス圧縮センシングを実践（図 1A, B）．

◆飛行ノイズやドップラーシフトの耐性を確保できた結果，1 送信器・2 受信器の超音波センシングで障害物回避飛行を実現（図 1C）．

◆本検証により，シンプルな生物ソナーを模倣した超音波運用法が，空中ナビゲーションシステムとして機能することが示せた．

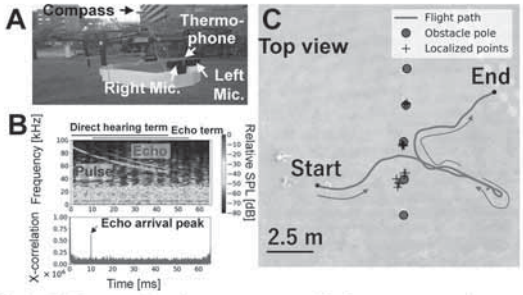


Fig.1: (A).Drone with active sonar system (C) Spectrogram and cross-correlation signal while drone detect the pole's echo during flight. (D) Example of obstacle avoidance navigation.

1-7P-8

1-7P-8　二層型カロリメトリ用水槽内に設置した断熱シートの超音波パワー計測への影響

Thermal insulating sheet films installed in a two-layer calorimetry tank influence on ultrasonic power measurement

◎上原長佑（沖縄医療工学院），内田武吉（産総研）

◆本研究では，カロリメトリ法に用いられる水槽内に断熱シートを設置し，断熱シートの厚さが超音波パワー計測に与える影響について有限要素法を用いて検討した。

◆本研究は，二層型のカロリメトリ用水槽の簡易モデルを用いて有限要素法によるシミュレーション解析を行った。超音波振動子から 10 mm の点に断熱シートを設置した。厚さは 0.01 mm，0.1 mm，1mm の 3 種類である。超音波振動子から 20 mm の点において音圧を測定した。得られた音圧を FFT 処理し，パワースペクトル密度の換算後，時間積分した値について相対比較を行った。

◆結果として，断熱シートなしと比較して，断熱シートの厚さ 0.01 mm および 0.1 mm では，それぞれ -0.011 dB，-0.064 dB であり，減衰率はそれぞれ 0.27 %，1.48 % 程度と 1.5 %以下であった。1 mm では -0.683 dB，14.8 % 程度の影響があった。断熱シート 0.1 mm より薄い厚さであれば減衰率は低く，有効である可能性が高い。

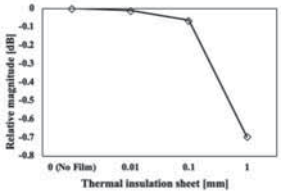


Fig.1: Relationships between the ultrasonic power and thermal insulation sheet films thickness

1-8P-1

1-8P-1　日本人大学生における英語習熟度の指標とリスニング力との関係

A study on the relationships between English acquisition parameters and listening abilities of Japanese college students

○須藤路子（順天堂大），伊藤憲治（国立感覚器セ）

本研究の目的は、英語の生成（発音）と知覚（リスニング）に関し、発音とリスニング力の習得到達度はどのような関係なのか検討することである。今回は、発音訓練の後、発音訓練がリスニング力と他の英語習熟度の指標に与える影響を測定した。

本研究の被験者は、海外在住経験の無い日本人大学生 10 名(男性 3 名,女性 7 名)であった。被験者グループのリスニング力とリーディング力を TOEFL IPT®により測定した。スピーキング力は、TOEIC® Speaking Test で測定した。読解速度と語彙数も標準テストにより、測定した。また、リスニングを困難にしているストレスを担わない弱形のリスニングテストを 2 種類独自に作成し、発音訓練の前後で、異なる version の 2 種のリスニングテストを実施した。

被験者の TOEFL ITP®テストの平均点は、発音訓練前 452 点で、訓練後は 477 点で、グループの平均点も被験者の個人スコアも上昇した。リスニングセクションのスコアも上昇した。TOEIC Speaking は、10 名中 2 名にスコアの上昇が示された。読解速度に関し、著しい習熟度の上昇が示された。語彙サイズも上昇した。2 種の弱形リスニングテストの結果では、訓練の効果は観察されなかった。

発音訓練の効果は、TOEFL ITP®テストスコア、読解速度と語彙サイズの向上に観察された。弱形の聞き取りの習得は困難であることが示された。

1-8P-2

1-8P-2　英語絵本読み聞かせ訓練をタスクとした学習者音声の韻律分析とその改善

Analysis and improvement of English learners' prosodic control in the task of storytelling training

☆正田千宙，高英翔，峯松信明，齋藤大輔（東大），中西のりこ(神戸学院大)

◆**英語絵本の読み聞かせ音声**を用いて、日本人が苦手とする英語の**韻律表現**を調査した。

◆適切な韻律表現で英語を話すことで聞き手側に伝わりやすくなる。

◆絵本を目の前の子どもに読み聞かせているような韻律表現豊かな音声や、Happy, Angry, Sad といった感情を込めた音声を日本人は模倣できるかを調べた。

◆どのような文のどのような箇所で母語話者英語と日本人英語のずれが生じているか詳しく分析した。

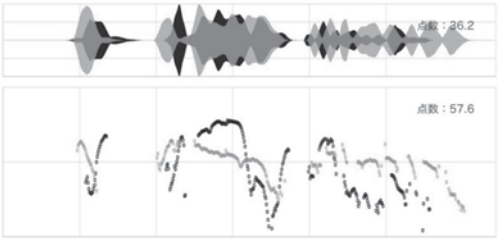


Fig.1: Appearance and Scoring of gap of intensity and pitch control between native speaker's sound and English learner's

1-8P-3

1-8P-3　SNS を活用した音声障害リハビリテーション支援システムの有用性評価

Assessment of SNS-based voice rehabilitation support system

○川村 直子(姫路獨協大), 北村 達也(甲南大)

- ◆ 音声障害のリハビリテーションでは，患者は言語聴覚士に指導された発声法を自宅でも適切かつ継続的に自主練習する必要がある.
- ◆ 音声障害のリハビリテーションで広く用いられているチューブ発声法を用いた自主練習の支援技術を開発した.
- ◆ このシステムは IoT 機器，クラウドシステム，患者のスマートフォンから成る．練習後のフィードバックと練習を促すリマインダ通知が患者のスマートフォンに LINE で届く「声かけ機能」を実現した.
- ◆ 評価実験を実施し，声かけ機能が自主練習の継続性に寄与することを確認した.

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

1-8P-4

1-8P-4　喉頭室の大きさを考慮した仮声帯発声実験

Influence of laryngeal ventricle on physical model of the vocal and ventricular folds

☆韓聰（立命館大），松島大輔（立命館大），鎌田将史（立命館大），徳田功（立命館大）

- ◆声帯および仮声帯モデルを用いて、喉頭室の大きさを考慮した仮声帯発声実験を行った。
- ◆声帯モデル単体の振動周波数は 95 Hz、仮声帯モデル単体の振動周波数は 146 Hz であった。
- ◆いずれの大きさの喉頭室においても、声帯と仮声帯は同期的に振動した。
- ◆喉頭室を最も狭めたときには、仮声帯が 1 回振動する間に、声帯は 2 回振動していることを確認した。

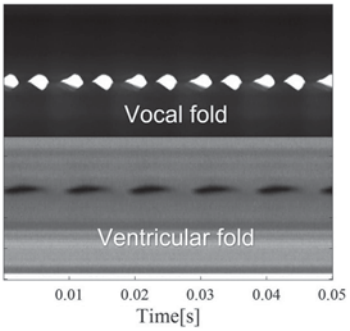


Fig.1 Kymograms observed when the ventricle size was 14mm.

1-8P-5

1-8P-5　声道を取り付けた声帯膜物理モデルの振動解析

Influence of vocal tract on physical model of the vocal membrane

☆槌田悠希,吉谷友紀, △小畑大樹, 徳田功(立命館大)

- ◆声帯膜物理モデルに声道モデルを取り付け、声帯膜の振動にどのような影響を与えるかを調べた。
- ◆声帯膜の長さ、硬さをそれぞれ二種類作成し、物理モデル単体の場合と声道有りの場合で吹鳴実験を行い、比較した。
- ◆声道によって物理モデルが音響フィードバックを受けていることが確認できた。また声帯膜が長く柔らかい場合に、声道と声帯膜の相互作用によって、不規則なカオス振動が起こることが確認できた。

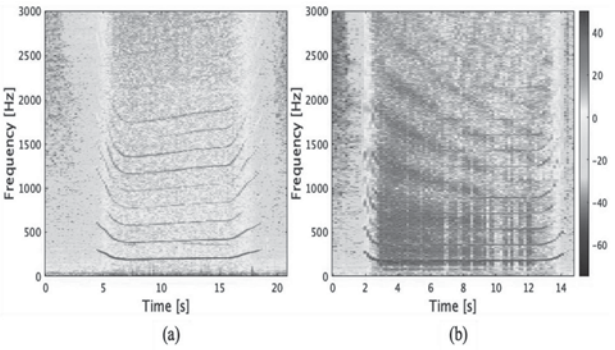


Fig. 1 (a) Spectrogram without VT. (b) Spectrogram with VT.

1-8P-6

1-8P-6　演技未経験者の感情音声の演技における台本の影響: 音響解析と主観評価による検討

The influence of the script on the acting speech of the emotional voice: evidence from　acoustic analysis

☆坂下尚史(豊橋技科大院)，河原英紀(和歌山大)，松井淑恵(豊橋技科大)

- ◆日本語感情音声データセット収録のため、演技未経験者による演技発話の台本の影響を音響解析と主観評価実験で調査した。
- ◆台本は２種類用意し、台本１は表現する感情と感情が起きる状況、台本２では話し手および聞き手の情報、発話背景を詳細に設定した。
- ◆20 代男性 6 名、女性 6 名の合計 12 名を対象に音声収録を行い、WORLD を用いて F0 解析を行った。
- ◆収録音声のうち男性5名と先行研究で用いられたプロの男女１名の音声に対してクラウドワーカー106 名を対象として主観評価実験を行った。
- ◆２要因分散分析を行ったところ、男性話者・女性話者ともに F0 平均では台本の主効果、感情指示の主効果、および台本と感情指示の交互作用が有意であった。
- ◆男性話者・女性話者ともに F0 偏差では台本の主効果、および台本と感情指示の交互作用が有意であったが、感情指示の主効果は有意でなかった。
- ◆主観評価の結果、台本１と台本２どちらにおいても「怒り(冷たい)」と「嫌悪」が混同される傾向にある。
- ◆台本１に比べ、台本２の方が、恐怖の音声を正しく「恐怖」と回答した割合が 20 ポイント高かった。

1-8P-7

1-8P-7　日本語音声のモーフィングに対する性別判断: 心理測定関数による知覚特性の調査

Gender judgments for morphing of Japanese speech sounds: Investigation of perceptual characteristics using psychometric functions

☆菅野聖真(豊橋技科大), 河原英紀(和歌山大), 松井淑恵(豊橋技科大)

◆男性話者と女性話者の発話音声をモーフィングし，その音声に対して性別を判断させることで，日本語母語話者の知覚特性を調査した．日本語母語話者の健聴者男性８人，女性８人が参加した。

◆FW03の男性２名, 女性２名の４モーラ単語を原音声として使用した．モーフィング割合は0%から100%まで10%刻みとし，全パラメータをこの割合とした．全ての男女ペアについてモーフィングしたものを刺激音声として使用した．

◆モーフィング割合ごとの「話者は女性」として判断した割合を，心理測定関数として累積ガウス分布にフィッティングした（Fig. 1）．

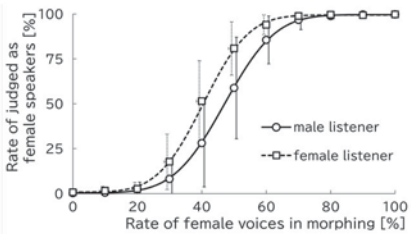


Fig.1：Comparison of psychometric functions by gender of participants.
(○：male listener, □：female listener. Error bars: SD.)

◆女性聴取者の PSE より男性聴取者の PSE が高かった（有意傾向）．このことは，男性聴取者が女性話者だと判断する際に，女性聴取者が求めるより強い女性の声としての特徴を求めることを示している．

1-8P-8

1-8P-8　マルチモーダル情報に基づく対話エージェントの応答時韻律・表情制御の評価

Evaluation of Prosodic and Facial Expression Control during Response of Dialogue Agents Based on Multimodal Information

☆渡辺稜哉, 伊藤彰則, 能勢隆(東北大), 千葉祐弥(NTT)

◆既存の擬人化対話エージェントの多くは韻律・表情を個別に制御
- ➢ 韻律と表情の間の関連を考慮できていない

◆提案モデルはシステム応答の単語単位韻律・表情制御信号系列を生成，学習時は２種の信号を同時生成するように学習（マルチタスク学習）
- ➢ 上記の関連を考慮した**学習の効率化・高精度なモデリング**を期待

◆これまでにマルチタスク学習の優位性は客観評価では示されていたものの，主観評価については未検討

◆本稿では主観評価より，マルチタスク学習は対話としての自然性の観点で**シングルタスク学習を有意に上回る**ことを確認

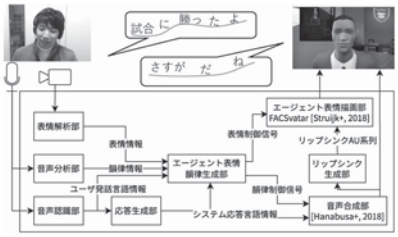


Fig.1：Overview of proposed

Table.1：Subjective Results
(Votes, MT：Multitask learning, ST：Singletask learning)

Presentation condition	MT	ST	P-value
Speech + Facial Animation	**178**	**102**	**< .001**
Face Animation	147	133	.43
Speech Animation	152	128	.16

1-8Q-1

1-8Q-1　音読支援アプリの開発に向けたインタフェースの予備的検討

Preliminary examination of an interface for the development of oral reading support application

○白勢彩子(東京学芸大), 宮澤幸希, △佐藤可直(フェアリーデバイセズ)

◆小学生の宿題の「定番」に音読がある。音読は国語の授業で重視されるが，教員にとっては自宅での実施内容が不明瞭，家庭では児童の動機づけなどの課題がある。

◆我々は，小学生の音読の宿題に特化したアプリによる課題解決を考え，児童向けの音読アプリの開発を目指して研究を進めている。本稿では，小学生児童の保護者と教員にアンケート調査を実施し，結果を踏まえてインタフェースについて検討する。

◆保護者調査：低・中学年では毎日課されるために積極的に取り組みにくいことや，課題の性質上，聞き手が必要になること，評価の基準や方法に問題があると捉えられた。「音声をアプリが自動で評価する」と「ゲーム形式で楽しく音読できる」機能への期待があった。

◆小学校教員調査：提案したアプリの機能については全般に好意的な意見であった。録音履歴が確認できる機能，教員が確認・評価する機能は，「頑張った履歴が本人に伝わるようなもの」，「日々の成長を確認できるといい」など，児童の実施歴が確認できる点が評価された。

◆調査結果を踏まえ，Fig.1 のようなインタフェースを提案する。

Fig.1 Sample of interface. Left: opening session, right: confirming of recordings.

1-8Q-2

1-8Q-2　英語リスニング学習支援のための低話速音声の聞き取りに関する検討

A Study on Listening to Natural and Simulated Low Speaking-Rate Speech for English Listening Learning Support System

☆千葉理揮, 能勢 隆, 伊藤 彰則 (東北大院・工学研)

◆本研究では，話速を制御可能な英語リスニング学習支援システムの実現に向け，人間による遅い話速の音声と，速い話速の音声を信号処理により一様伸縮して低話速化した音声を比較し，その知覚への影響について調査する．

◆一様伸縮による影響を分析するために，英文の書き起こしテストを用いた実験を行う．この際，話速や英語学習者の聴解に大きく関係する要素として，リンキング・リダクションに着目した．

◆リンキング・リダクションの有無で聴解に差が生じることを確認した．また，収録時の話速が聴解に影響を与えることが確認できた．

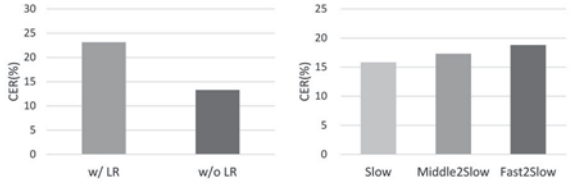


Fig.1： Character Error Rate
(w/ LR：with Linking or Reduction, w/o LR：without Linking and Reduction)

◆英語のリスニングにはリンキング・リダクションが大きく聴解に影響を及ぼすことから，一様伸縮による話速の制御よりも，これらの発音の変化を考慮できる話速制御手法が望ましい．

1-8Q-3

1-8Q-3　モノローグ発話における音声のパワーと口形の関係の調査

An investigation of relationship between speech power and mouth shape in monologue speech

☆川西翔貴（東北大），伊藤彰則（東北大），千葉祐弥（NTT），能勢隆（東北大）

◆口形は音素の種類だけではなく感情や個人性などのさまざまな影響を受ける可能性がある．また呼気吸気の関係で，音声のない区間においても口唇の動きがある可能性がある．

◆本稿ではより人間らしい対話エージェントを構築することを目的として，これらの要因と口唇の動きの関係を調査した．

◆調査には ITA コーパス[小口他, 2021]の文を読み上げたマルチモーダルデータベース[1]を用いた．

◆調査の結果，発話前，発話後には無音区間であっても口を大きく開けていることが示唆されたほか，その度合いには個人差があることがわかった．また，感情も口の開きの大きさに影響を与えることが示唆された．

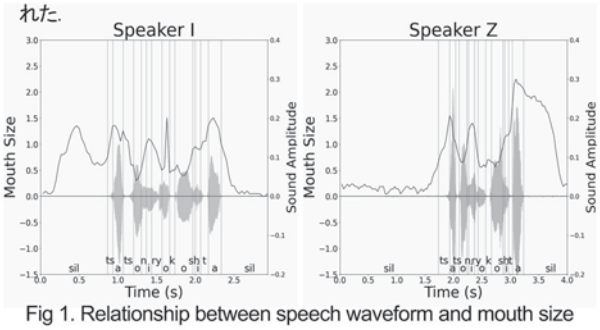


Fig 1. Relationship between speech waveform and mouth size

[1] https://zunko.jp/multimodal_dev/login.php

1-8Q-4

1-8Q-4　モーフィング音声を用いた話者類似性判断と話者同定判断の比較:発話単語の同異と鼻音の有無による影響

A comparison of speaker similarity judgement and speaker identification judgement using morphed speech sounds: effects of nasal and word difference

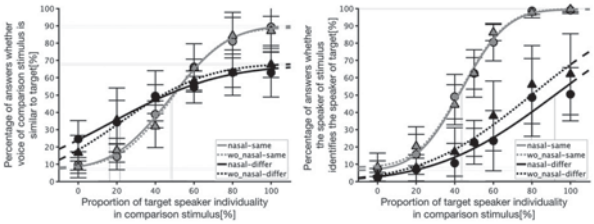
☆内堀颯太（豊橋技科大院），河原英紀（和歌山大），松井淑恵（豊橋技科大）

◆音声を異なる話者間でモーフィングし，モーフィング割合が段階的に変わるとき，話者の類似性・同定判断がどのように変わるかを調べた．

◆6 種の日本語 4 モーラ単語 (3 種は 3 モーラ目の子音が鼻音) を発話した音声を用いて，聴き比べる音声の発話単語の同異と，単語中の鼻音の有無が話者識別判断に及ぼす影響を定量的に調べた．

◆結果，単語に鼻音を含むことは話者識別判断を容易にしなかった．一方で，発話単語が異なる音声を聞き比べた場合，発話単語が同じ音声を聞き比べた場合より，話者の類似性・同定判断が難化した．

◆「同一話者である」と同定することと，「話者が異なる」と弁別することとで，判断基準が異なる可能性が示唆された．



(a) speaker similarity judgement　(b) speaker identification judgement

Fig.1: The results of two speaker recognition judgements (Nasality (●:nasal, ▲:without-nasal), Words (gray:same, black:differ)).

1-8Q-5

1-8Q-5　超音波診断装置による調音運動観測におけるプローブ位置の補正

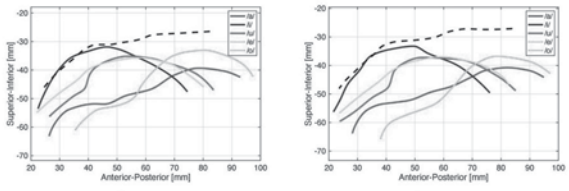
Calibration of relative probe position on measurement of articulatory movements by ultrasound

○大山陣，北村達也（甲南大），孫静，林良子（神戸大）

◆超音波診断装置は舌の運動を簡便に可視化できる比較的安価な装置として広く利用されているが，得られる画像には下顎の動きが反映されない．

◆観測対象の頭部とプローブに貼り付けた 2D マーカーを用いる簡便なプローブ位置補正法を提案する．

◆超音波撮像と同時に Web カメラにより話者の側方の映像を撮影し，後者を用いて求めた頭部とプローブの位置関係から超音波画像を補正する．



(a) Original　(b) Calibrated

Fig.1: The palate and tongue measured during production of the five Japanese vowels. Tracings of (a) original and (b) calibrated images.

1-8Q-6

1-8Q-6　喉頭室の影響を考慮した声帯物理モデルの吸気発声実験

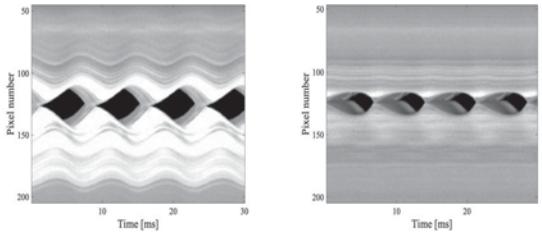
Effect of laryngeal ventricle on inspiratory phonation of the vocal fold physical model

☆長谷川寛人（立命館大），△足立恭平（立命館大），徳田功（立命館大）

◆大きさの異なる四種類の喉頭室モデルと MRI 声帯物理モデルを用いて、呼気・吸気発声の比較実験を行った。

◆同一の声帯物理モデルを用いて、吸気発声時および呼気発声時における声門下圧、音声、高速度動画を計測した。

◆吸気発声では、喉頭室の大きさに依らず、基本周波数、呼気発声ともに、ほぼ一定であった。これらの傾向から、喉頭室の大きさは、あまり吸気発声に影響しないことが分かった。



◆

Fig.1 Kymograms of expiratory (left) and inspiratory (right) phonations.

1-8Q-7

1-8Q-7　アカゲザルの声帯物理モデルで観測された声帯振動から声帯−仮声帯振動への遷移現象

Transition from vocal fold oscillations to vocal-ventricular fold co-oscillations in a physical model of rhesus macaque (*Macaca mulatta*)

☆吉谷友紀，上田可奈子，小畑大樹（立命館大），
△西村剛（京都大），徳田功（立命館大）

◆アカゲザル摘出喉頭のマイクロCT画像をもとに作成された物理モデルを使用し，声帯振動から声帯−仮声帯振動への遷移現象を調査した。
◆流量を60〜160 L/minまで，20 L/min間隔で変更した。
◆流量が増加することで，声帯振動から声帯−仮声帯振動への遷移がみられた。
◆遷移区間において，振動が複雑化し，発声効率が低下した。

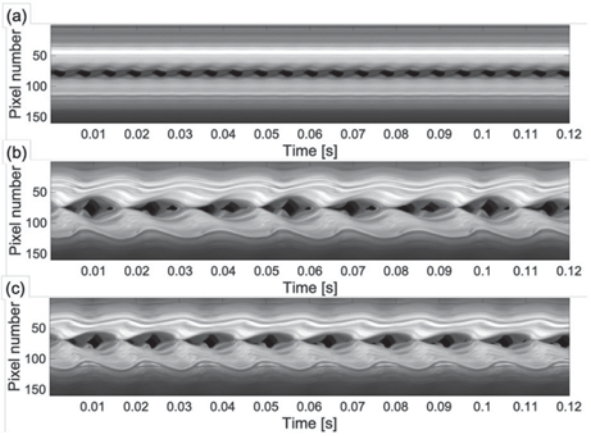


Fig.1：Kymograms with flow-rate of (a) 100, (b) 120 and (c) 140 L/min.