3-1-1

3-1-1　位相の混合微分に基づく最適化アルゴリズムを用いた調波打撃音分離

Harmonic and percussive source separation with optimization algorithm based on mixed partial derivative of phase spectrogram

◎赤石夏輝（早大理工），矢田部浩平（農工大），及川靖広（早大理工）

◆背景：調波打撃音分離（HPSS）

・・・音響信号の調波成分と打撃音成分を分離するタスク

我々はこれまでに**位相の混合微分**に基づく手法を提案

◆従来法：分離過程で位相が入力信号から変わらない

→ 自然な信号による分離ができていない

◆提案法：ADMM アルゴリズムの形を適用

→ **分離過程で位相を修復**し安定した分離を可能に

◆結果：反復によって分離性能が**大きく向上**した

打撃音側の SDR（100 反復目）：**+1.46 dB**

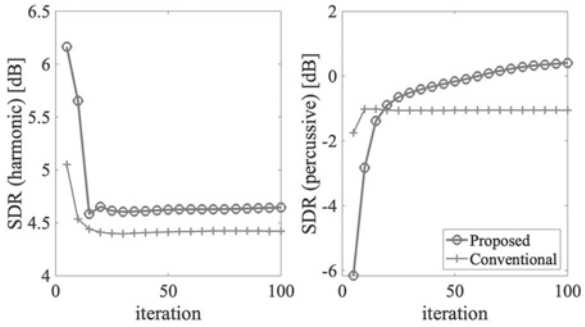


Fig.1: Comparison of separation performance for each iteration between conventional and proposed method.

3-1-2

3-1-2　Synchrosqueezing を用いた高精度クロマベクトルのオンライン推定

Highly-accurate chroma vector online estimation with Synchrosqueeezing

☆ 伊藤陸人（早大理工），矢田部浩平（農工大），及川靖広（早大理工）

- 目的：
  高精度クロマベクトルのオンライン推定
- 再割り当て法を用いた従来手法：
  スパースな表現を得るにはスペクトログラム全体の計算が必要
  → 精度は高いがオンライン推定は不可能
- 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
  → クロマベクトルの**オンライン推定可能**
  → 従来法に遜色ない結果がオンラインでも得られた

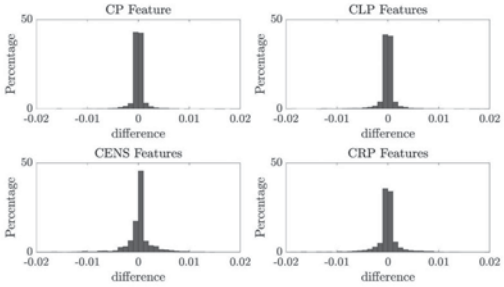


Fig1: Difference of correlation coefficient between true and predicted chroma with conventional and proposed method

3-1-3

3-1-3　Fine-tuning ImageNet Pre-trained Models To Be Transferred for Predominant Instrument Recognition in Polyphonic Music

☆ Lifan Zhong, Daisuke Saito, Nobuaki Minematsu (UTokyo)

◆Recently, the convolutional neural networks (CNNs) were proved efficient in detecting the predominant instruments in polyphonic musical content, with the Mel-spectrogram as the input. Such methods are, to some extent, similar to image classification.

◆In this work, we proposed to pre-train the CNN models with the ImageNet dataset, a large image corpus, and then fine-tune the pre-trained models with the Mel-spectrograms of the musical data for predominant instrument recognition.

◆We evaluated our systems on the IRMAS dataset, an instrument recognition dataset containing professionally produced real-world music recordings with different genres and audio quality.

◆The evaluation results showed that the ImageNet pre-trained models outperformed the ones trained from scratch, even though there is no apparent relationship between the images of real-world objects and the Mel-spectrograms.

◆In our experiments, the best model was the ResNet50 model with ImageNet pre-trained weights, achieving a micro-weighted F1 score of 0.649, a 7.8% relative improvement compared to the baseline.

3-1-4

3-1-4　時間周波数表現を用いた畳み込みニューラルネットワークに基づく演奏楽器推定

Convolutional Neural Network-Based Musical Instrument Identification Using Time-Frequency Representations

☆城間 佑樹（都立大），水野 賀文，高橋 祐（ヤマハ），
塩田 さやか（都立大），近藤 多伸（ヤマハ），貴家 仁志（都立大）

◆提案法

- 入力音の時間周波数表現に画像特徴抽出法を適用
- ConvMixer(画像分類で高い性能の分類器)を用いて演奏楽器推定

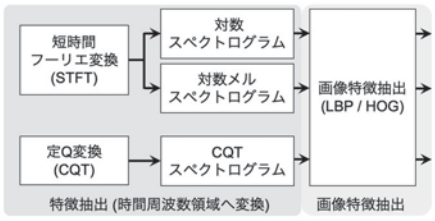


Fig.1：A flow of feature extraction in the proposed method

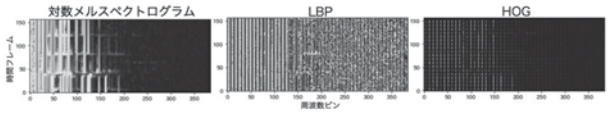


Fig.2：Time-Frequency representations with image feature extraction

◆実験結果

- 画像特徴抽出法を含む手法のアンサンブルを実施

分類正解率(従来法):調波楽器: 86.27%，非調波楽器: 97.39%

(提案法):調波楽器: 98.53%，非調波楽器: 99.55%

3-1-5

3-1-5　各楽器音源に着目した楽曲間類似度学習の評価

Evaluation of music similarity learning focusing on each instrumental sound

☆橋爪優果，李莉，戸田智基(名大)

◆ **背景**：インターネット上の楽曲の量は膨大であるため，ユーザが効率的に好みの楽曲を見つけるための楽曲推薦，検索技術が必要となり，その基準として楽曲間類似度が重要となる．

◆ **目的**：楽曲の部分的な特徴に基づいた類似度を計算することで，より自由度の高い楽曲推薦，検索を実現できる．本研究では，我々が以前提案した，各楽器音に着目した楽曲間類似度を距離学習により計算する手法のさらなる評価を行う．

◆ **実験**：楽曲を構成する各楽器音を入力に用いて各楽器音類似度を計算する．学習に用いる楽曲数を変化させた場合，学習データと秒数の異なるデータでテストした場合について評価を行う．

◆ **結果**：学習データを増やすと精度が向上すること，モデルの学習に用いたデータより長い秒数のデータを入力すると精度の高い結果を得られることがわかった．

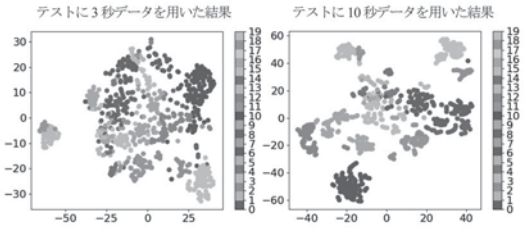


Fig.1：2次元で可視化したピアノ音(楽器音分離による)の埋め込み空間

3-1-6

3-1-6　複数音高推定における音響特徴量およびモデルによる推定性能の違いに関する調査

Investigation of the performance difference by the acoustic features and models in multiple pitch estimation based on deep learning

☆川凌司，坂野秀樹，旭健作(名城大院)

◆複数音高推定における深層学習の入力音響特徴量およびモデルを変更するとどのような推定性能の違いとなるのかを調査し、音響特徴量の策定とモデルの策定の参考とする

◆実験条件

- 音響特徴量は時間波形、スペクトル、対数スペクトルの３種類
- モデルはCREPE、DA-Net、Transformer の３種

Table 1 Each metric in the test results for different models and acoustic features

Model	Acoustic feature	Precision	Recall	F1
CREPE	waveform	0.660	0.513	0.577
	spectrum	0.645	0.619	0.645
	log spectrum	**0.687**	0.542	0.606
DA-Net	waveform	0.651	0.565	0.605
	spectrum	0.665	**0.665**	**0.665**
	log spectrum	0.667	0.643	0.655
Transformer	waveform	0.647	0.481	0.552
	spectrogram	0.672	0.590	0.629
	log spectrogram	0.630	0.492	0.552

3-1-7

Note-level Automatic Guitar Transcription Using Attention Mechanism and Multi-task Learning

注意機構とマルチタスク学習を用いた音符単位ギター自動採譜システム

〇金世訓，林知樹，戸田智基(名大)

◆In this research, we propose a deep neural network (DNN) based **note-level automatic guitar transcription system** that can generate **both frame-level and note-level** guitar transcription.

◆We use **attention mechanism**, along with **convolutional neural network (CNN)** to effectively **model both short and long-term characteristics**.

◆In our experiment, we confirmed that 1) the **proposed method significantly outperforms** the conventional method based on a CNN in frame-level estimation performance and that 2) the proposed method can **also generate note-level guitar transcription.**

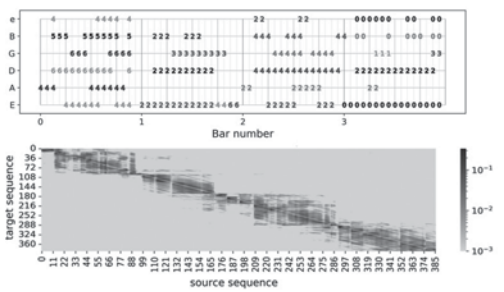


Fig.1：An example pair of note-level estimation result (top) and corresponding attention map from the self-attention mechanism of our model (bottom).

3-1-8

3-1-8　複数 BGM を手掛かりとしたロボットの自己位置推定

Mobile robot localization based on sound source localization under mixed background music

☆津田龍星(阪産大院),赤塚俊洋(阪産大),高橋徹(阪産大院),
江川琢真(立命大院),中山雅人(阪産大院)

目的: 複数 BGM を手掛かりにロボットまでの自己位置推定

・BGM を再生するそれぞれのスピーカーまでの距離から自己位置を推定

手段: (1) 複数 BGM の混合音からその中の１つの BGM 信号との相関を最大化し伝達遅延を推定

(2) 混合音から直接特徴量を求める

（音源分離など重い処理を回避）

・バイナリクロマスペクトル(BCS)を特徴量として採用

→背景雑音・非目的音に頑健、時間周波数領域に薄く広く情報を有す

評価: 時間波形の相互相関・提案法（BCS の相互相関）を推定値の平均絶対値誤差で比較

結論: 提案手法は 2m 以上で従来法より低誤差

Table.1：Error (Average of absolute error)

真値(m)	従来法(m)	提案法(m)
1	3.5	6.4
2	5.7	3.2
3	8.3	0.0
4	15.5	3.6
5	19.2	4.0
6	21.6	7.3
7	25.0	11.9

3-1-9

3-1-9　仮想残響エンベロープを用いた音楽試聴時のリラックス感測定の可能性

Possibility of relaxation level measurement during music listening using hypothetical reverberation envelope

☆笠島菜那(大阪工大)，脇田由実(大阪工大)

我々は，残響の聴こえに関する楽曲の音響信号特徴量の指標として従来提案されているエンベロープ指数（E 値）に注目し，E 値がリラックス感を与える音楽の指標となる可能性を探るために，同一音源，同一楽器を用いて，異なる奏法で演奏された 4 つの音源を聴取させ，E 値の違いがリラックス感に与える影響をアンケートによって調査した．その結果，E 値が低いほどリラックス感を感じる傾向にあることがわかった。下図は奏法が異なる各音源に対して，好みであるか，リラックスできるかを 5 段階評価で質問したアンケート結果を示し，（ ）内の数値は各音源の E 値を示している．

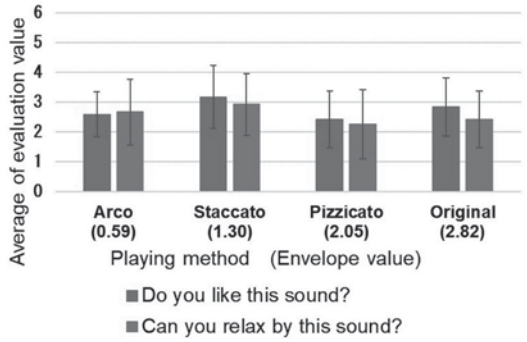


Fig1 Average of evaluation value for each playing method

3-1-10

3-1-10　リズム情報共有のための心拍音可視化手法

Visualization of heart beats for sharing rhythm

○高橋 徹（阪産大院），塩安 佳樹（はごろも内科・小児科）

目的: 聴診器で観測する心拍音のデジタルデータ化
特にリズム情報を重点化
１サイクルに２パルスあり従来の基本周波数推定は不適当

手段: 聴診器にマイクロホンを接続し録音
雑音に強いリズム情報（＝基本周期）を示す特徴量の探索
リズム情報　⇒　自己相関関数による基本周期の明瞭化

評価: 胸部に直接聴診器を当てる場合とＴシャツの上からの場合を比較

結論: 自己相関関数表現による可視化は基本周期の提示に有効

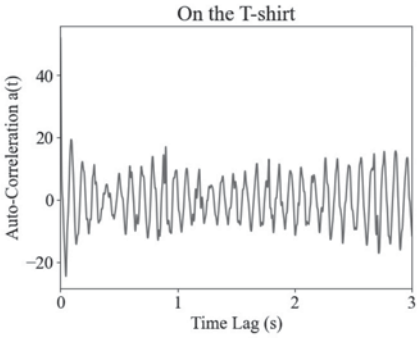


Fig.1：Autocorrelation waveform of heartbeat sound recorded from the top of the T-shirt.

3-2-1

3-2-1　カーネル補間に基づく空間能動騒音制御における二次音源からの外部放射抑圧

Suppression of exterior radiation from secondary sources for kernel-interpolation-based spatial active noise control

☆有川和志，児島孝明，小山翔一，猿渡洋（東大院・情報理工）

◆カーネル補間に基づく空間 ANC の従来法では，二次音源からの外部放射が考慮されていなかった．

◆そこで本稿では，<u>二次音源からの外部放射エネルギーを抑圧する</u>カーネル補間 ANC を提案した．

◆二次音源からの外部放射エネルギーをペナルティ項として加える場合と，不等式制約として加える場合の二通りを考え，それぞれで適応フィルタアルゴリズムを導出した．

◆二次元自由空間シミュレーションにより，提案法は対象領域で音圧を抑圧しつつ，外部放射エネルギーを小さくできることを確認した．

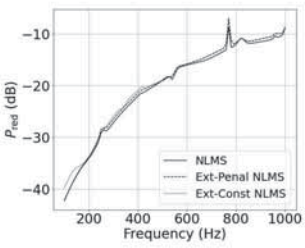


Fig.1：
Final regional noise power reduction for each frequency

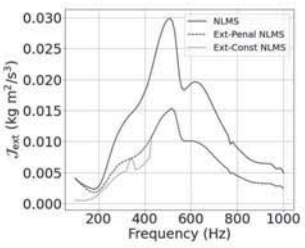


Fig.2：
Final exterior radiation power for each frequency

3-2-2

3-2-2　方向重み付き外部放射抑圧を用いた多点振幅制御による局所再生

Local-field sound reproduction based on amplitude matching using directionally weighted exterior cancellation

☆冨田佳秀，小山翔一，猿渡洋(東大院・情報理工)

◆マルチゾーン音場合成技術には，多点音圧制御法，多点振幅制御法，モードマッチング法などが知られるが，いずれもスピーカ外部への放射が問題となる．

◆外部放射の方向ごとの許容度を重み関数とする，<u>方向重み付き外部放射エネルギー</u>を導出．

◆<u>多点振幅制御法</u>に<u>方向重み付き外部放射エネルギー</u>を罰則項として加えた<u>局所再生手法</u>を提案．

◆数値シミュレーション実験，実環境実験により，提案手法が内部の合成精度を保ちつつ，外部への放射を抑圧できることを実証．

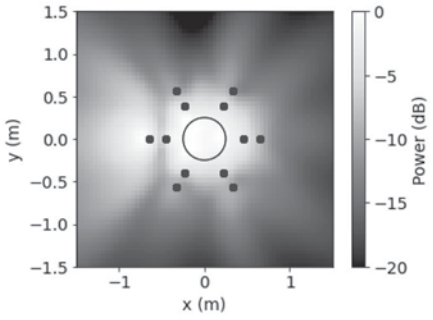


Fig. 1: Power distribution of the sound field synthesized by the proposed method.

3-2-3

3-2-3　聴覚特性を利用した音圧・振幅マッチングの併用による音場合成

Sound field synthesis combining pressure and amplitude matching using auditory property

◎木村 圭佑，小山 翔一，猿渡 洋（東大院・情報理工）

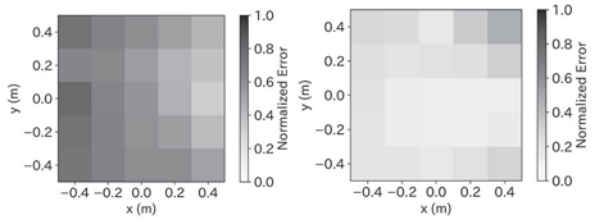
◆音場合成において空間エイリアシングの聴感上の影響を軽減する，音圧・振幅マッチングの併用による新たな手法を提案した．

◆低周波数帯域では所望の振幅・位相分布を合成するが，水平方向の知覚において両耳間レベル差（ILD）が主要な手がかりとなる高周波数帯域では所望の振幅分布のみを合成し，位相分布を任意とする．

◆交互方向乗数法に基づくアルゴリズムを定式化した．

◆数値シミュレーション実験により，対象領域内において高精度にILDを合成できること，平坦な振幅特性を実現できることが確認された．

◆実システムを用いた主観評価実験により，提案法が聴感上も高い品質で音場を合成できることが確認された．



(a)　音圧マッチング法　　(b)　提案法

Fig.1：対象領域内を5×5の計25区画に分割し，各区画の中心に頭部をおいた際の，ILDの正規化誤差の分布．

3-2-4

3-2-4　スプライン補間に基づく音場表現を用いた Physics-Informed Neural Networks による音場推定 ー散乱体を含む領域に関する検証ー

Sound Field Estimation Based on Physics-Informed Neural Networks Using Spline-Interpolation-Based Sound Field Representation - Case of Region Including Scatterers -

☆重見和秀，小山翔一，中村友彦，猿渡洋（東大院・情報理工）

◆音場推定問題に対し，従来の畳み込みニューラルネットワーク（CNN）に基づく手法は，少ない観測点からの音場推定を可能にする一方，推定結果が物理的に実現可能なものとなる保証はない．

◆本稿では，従来の CNN に基づく手法に対し，Helmholtz 方程式に基づく損失関数を導入し，推定結果をより物理的に実現可能な解へと誘導する手法を提案する．

◆空間方向のラプラス演算子を含むHelmholtz方程式からの逸脱を評価する上で，従来法における CNN は対象領域内の離散点上の推定音圧値を出力するため，CNN の出力をスプライン補間により連続関数として表現することで，この評価を可能にした．

◆数値実験により，特に推定対象領域の内部に散乱体がある条件の下，推定精度を保ちつつ，より物理的に実現可能な推定結果が得られる傾向を確認した．

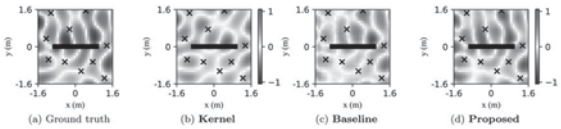


Fig.2：Estimated sound pressure distribution. Crosses indicate the observation points (M=10). NMSEs of Kernel, Baseline, and Proposed were -0.69, -7.71, and -10.7 dB, respectively. Relative Helmholtz Equation Error (RHE) was -6.27 dB.

3-2-5

3-2-5　16チャネル小型円形スピーカアレイを用いたマルチスポット再生システムの実装

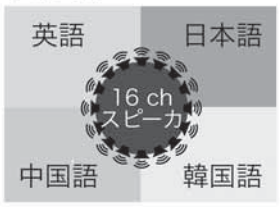
Implementation of multiple sound spot generation system using a compact circular array of 16 loudspeakers

○岡本拓磨 (NICT),
△大山慎二，△上野克司，△岡部司，△谷健太郎 (北日本音響)

Q：「持ち運び可能なマルチスポット再生システムを実現したいのですが?」

オカモト&北日本音響：「それならこれで!!」

16チャネル
小型円形スピーカアレイ



ニューラル音声合成を用いた
4言語マルチスポット再生

＋

タブレットを用いて
エリアを自由に回転可能
(岡本ら，音講論，2012秋)

スーツケースにスタンド・ケーブル・アンプ・PC含めて持ち運び可能

発表会場にて実機デモあり!!

3-2-6

3-2-6　音響シーン認識のための サブアレイ間相関特徴量の検討

An investigation of correlation-based features calculated between subarrays for acoustic scene analysis

☆河村 隆生，木下 裕磨，小野 順貴（都立大），シャイブラー ロビン（LINE）

◆目的

- 分散マイクロホンアレイでサブアレイが同期している場合に，サブアレイ間相関特徴量の音響シーン分類への寄与を調査する。

◆概要

- サブアレイ間相関特徴量として２つのサブアレイ間で計算された白色化相互相関関数 (GCC-PHAT) を検討した。
- 「GCC-PHAT を用いたシーン分類性能の評価」と「ラベルごとに平均した GCC-PHAT の概形を比較」した。

◆まとめ

- GCC-PHAT を用いた分類結果（macro F1-score）は 72.7%
- ３つ以上のサブアレイ間で計算された分類により性能向上が期待

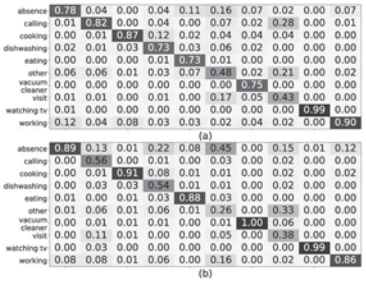


Fig.1：Normalized confusion matrices (nCM). nCM is normalized by the marginals of either the column or the row. (a) precision nCM. (b) recall nCM

3-2-7

3-2-7　ヘッドホン型デバイスを用いたオンライン音響イベント定位

On-line SELD utilizing headphone-type device

◎安田昌弘，中山彰，齊藤翔一郎，日和﨑祐介（NTT）

【音響イベント定位】

- ■ 音響イベント定位 (SELD) は音響信号から，いつ・どこで・何が起こったのかを推定するタスク

→ 本研究では特にヘッドホン型デバイスを用いた SELD に着目

【課題】

- ■ 音源定位のための入力特徴量として用いられる音響強度ベクトル (IV) を，ヘッドホン型デバイスから導出することは困難
- ■ IV の代替としてマイク間の位相差 (IPD)が考えられるが，メモリ制約のあるデバイス上のオンラインシステムには不適

【提案法】

- ■ 頭部を剛体球であると仮定して，マイクロフォンから得られた信号を球面調和関数展開することにより擬似的なアンビソニックス信号 (擬似 FOA) を導出
- ■ 擬似 FOA から得られた擬似 IV に対して各成分の正規化を行うことで，剛体球近似による影響を補正する

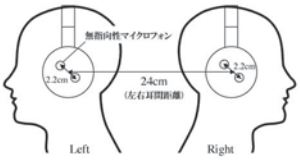


Fig.1 Headphone-type device used in this study and microphone placement on it.

3-2-8

3-2-8　音響イベントトリアージ：クラスの優先度を考慮したイベント検出

Sound Event Triage: Sound Event Detection with Priority of Classes

◎砺波紀之(NEC)，井本桂右(同志社大学)

◆音響イベント検出 (SED)の新たなタスク音響イベントトリアージ (SET)を提案する。

◆従来の SED は全てのイベントクラスを平等に検出することが目的であった。

◆しかしながら，ユーザやアプリケーションによって検出対象となるイベントは異なる。

◆そこで，各イベントクラスに検出優先度を設定する新たなタスク SET と単一ターゲットのための深層学習ネットワークを提案する。

◆評価実験により，提案手法が従来の SED 手法と比較して特定のイベントで 7.53%ポイントの F-score が向上することがわかった。

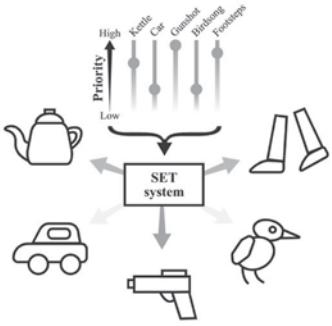


Figure: Overall of SET

3-2-9

3-2-9　MIMII DG：稼働音診断におけるドメイン汎化手法向けデータセット

MIMII DG : Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task

◎土肥宏太，西田智哉，Purohit Harsh，田邊亮，遠藤隆，山本正明，二階堂悠貴，川口洋平(日立製作所)

◆熟練点検者の不足に対処するため、機械の稼働音から自動で点検を行う異常音検知システムを開発している。異常音検知タスクにおいて、ドメイン汎化性能を評価するための稼働音データセットを作成した。

◆異常音検知システムの運用時にドメインシフトが発生すると、正常音データの分布が変化し、検知性能が低下する。ドメインシフトに対処する方法としてドメイン適応があるが、ドメインシフトが高頻度で発生する場合やドメインシフトの発生が検出困難な場合には、ドメイン適応が困難である。これらの場合においては、ドメインシフトに依存しない表現を抽出するドメイン汎化が有用と考えられる。しかしながら、従来の異常音検知タスク向けデータセットはドメイン適応の使用を想定して作られたため、ドメイン汎化手法の性能評価には適さない。

◆そこで、ドメイン汎化性能を評価するためのデータセット MIMII DG を作成した。MIMII DG は 5 種類の異なる機械種で構成され、各機械種に 3 種類のドメインシフトを含む。ドメイン汎化を行うため、各ドメインシフトが最低でも 3 種類の異なる録音条件を含むようにした。また、対象機械の周囲で稼働する機械の状態が変化するドメインシフト等、検出が困難なドメインシフトを作成した。

◆ベースラインシステムを用いた異常音検知性能評価の結果、ドメインシフトによる性能劣化が再現できていることが分かり、MIMII DG がドメイン汎化性能の評価に有用であることが示唆された。

3-2-10

3-2-10　最小全域木を用いた複数時間差の同時推定

Robust estimation of multiple time delays utilizing minimum spanning tree

◎山岡洸瑛，中嶋大志，小野順貴（都立大）

◆時間差はアレイ信号処理において重要な空間情報であり，これまでに 2 チャネル間の時間差を推定する手法が盛んに研究されてきた．

- ➢ 3 チャネル以上のアレイでは，マイクペア毎に時間差を推定することですべてのペア間の推定値が求まる（ペアワイズ法）．

◆Mチャネルのアレイではペアワイズ法により$M(M-1)/2$個の時間差が求まるが，実際の時間差の自由度は$M-1$であり，冗長である．

◆本発表では，冗長な時間差から最適な$M-1$個の時間差を選択するという問題を，最小全域木を求める問題へと帰着させ，ペアワイズ法による推定誤差に頑健な複数時間差の同時推定法を提案する．

- ➢ 更に，以前提案した補助関数型時間差推定を後段の処理として用いることで，初期値に頑健かつ高精度な時間差推定を達成した．

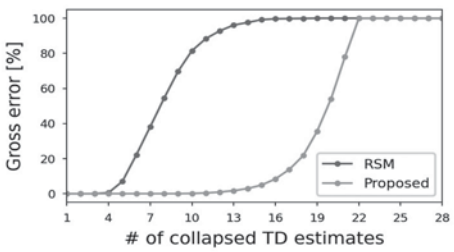


Fig. 1: Simulation results of the time delay estimation using an $M = 8$ channel microphone array (1000 trials). The horizontal axis shows the number of incorrect TDs out of 28 redundant TDs. RSM: naïve method.

3-2-11

3-2-11　光ファイバマイクロホンを用いた分散型光ファイバセンシングによる音源位置推定の基礎的検討

Basic Study of Sound Source Localization by Distributed Optical Fiber Sensing with Optical Fiber Microphone

◎美島咲子，△河野航，△松下崇，△樋野智之（NEC）

◆光ファイバ内を伝搬する光の位相変化を利用して音や振動を計測する分散型音響センシング (Distributed Acoustic Sensing; DAS)は，光ファイバに沿った複数地点の信号を単一の装置で計測できるため，広域空間の音響計測手段として有望である．

◆光ファイバマイクロホンとは円筒の外周に光ファイバを巻き付けた部材を指し，共振を利用した受音感度の向上と同一受音地点での多チャネル計測が実現できる．

◆光ファイバマイクロホンで得られる多チャネル信号を用いた音源位置推定法を提案する．屋外環境下での評価実験の結果，光ファイバマイクロホンで囲まれた領域内で音源を放射した場合の誤差距離において，従来法と比較して 4.2m の改善を確認した．

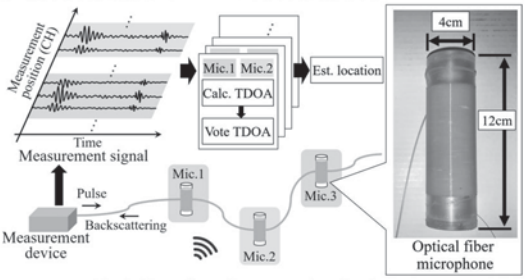


Fig.1：Overview of proposed method.

3-2-12

3-2-12　超低周波音を用いた音波インバージョンによる津波早期検知法

Early tsunami detection method using infrasound data inversion

○大久保 寛，△齊藤 義騎(東京都立大学)
△上原 謙太郎，△舘畑 秀衛，△林 健次(日本気象協会)
△橋詰 正広(中部電力)

◆本研究では，波源と大気モデルを考慮した大規模伝搬数値シミュレーションを実装し，津波による超低周波音の伝播現象について数値的に再現する方法を開発した．

◆さらに，開発した数値解析モデルを用いて，単位波源からの音波波形を再現し，それらを用いて，インバージョンによって津波波源を推定する方法を提案する．

◆超低周波音の波源は海面変動であり，津波の波源と極めて関係性が深い．また，津波波源を早期に求めることができれば，津波伝搬計算によって陸地に到達する津波現象を即座に求めることもできる．

◆低周波音波を用いた津波波源推定は新たな早期津波検知法としての可能性を大きく広げることができるだろう．

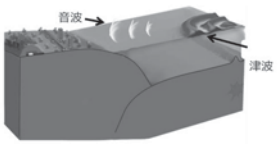


Fig. 1 Schematic diagram of secondary phenomenon of wave motion caused by tsunami.

Fig. 2 Tsunami source estimation results from infrasound data inversion.

3-2-13

3-2-13　delayed-X harmonics synthesizer アルゴリズムによるハウリングキャンセラ

Howling canceller with delayed-X harmonics synthesizer algorithm

☆信川裕介，下倉良太，△飯國洋二（大阪大）

◆近年急速に普及している遠隔会議システムにはハウリングキャンセラが搭載されているが，いまだ発展途上であることが知られている．

◆本研究では，現在消音することのできない「同室で２台の PC が干渉することで発生するハウリング」を消音できるようなアルゴリズムの提案を行った．

◆倍音構造を持つというハウリングの特性から，周期的なノイズの制御に特化した delayed-X harmonics synthesizer アルゴリズムを改良した．これは，ノイズの基本周波数と倍音数を用いることで倍音成分の消音が可能であるため，振幅スペクトルを求めることで推定を行った．

◆提案アルゴリズムを用いた結果，音声の音質劣化を防ぎつつハウリングの消音が可能であることを確認した．（Fig.1,2）

◆ただし，録音した音声データに対して消音を行っているため，今後はシステムの遅延を考慮し，アルゴリズムのリアルタイム化を行う必要がある．

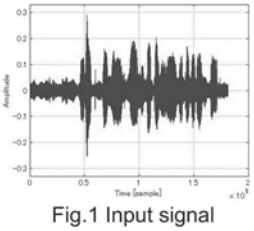


Fig.1 Input signal

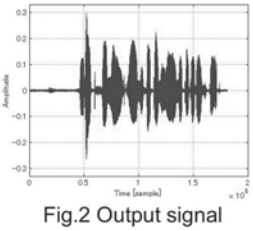


Fig.2 Output signal

3-2-14

3-2-14　周波数変換法のマルチチャネル化に向けた検討

Study on multi-channelization of frequency conversion method

○藤井健作(コダウェイ研)，△棟安実治(関西大・エ)，苣木禎史(千葉工大)

能動正弦波騒音制御システムのマルチチャネル化は，参照信号間に独立成分が存在しないために困難とされる．本報告では，スピーカの入力側に逆フィルタシステムを挿入することによってスピーカから誤差マイクロホンに至る交差経路が相殺され，各チャネルが分離されることによって，この問題が解決されることを示す．その構造を Fig.1 に，分離されたチャネルの構造を Fig.2 に示す．本報告では，周波数が 200, 300, 400, 500, 600 Hz の正弦波の和を制御対象騒音として，Filtered-x LMS 法と周波数変換法を適用して性能を比較した．その結果，周波数変換法で計算量が少なく，その低減速度も高いことが確認された．

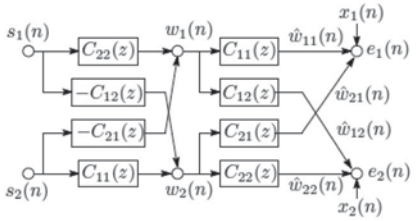


Fig.1：Configuration of proposed system

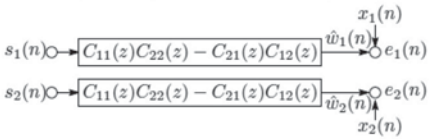


Fig.2：Equivalent configuration

3-4-1

3-4-1　日本の音楽における機能的な役割 —仕事歌と下座音楽の紹介を中心に—

Functional roles in Japanese music -Focusing on the introduction of work songs and Geza music (sound effects in Kabuki)-

〇安藤珠希（箏曲）

◆現在の生活において、身の回りには音や音楽があふれている。
◆式典や行事でも音楽が演奏され、祭礼の場合は、音楽が宗教的な役割や奉納の役割を担う。
◆一方で、第二次世界大戦中においては、国民の心を一致団結して士気を高める曲も作曲されている。
◆音や音楽には様々な場面で様々な役割があったり、求められたりするが、本発表では、数ある日本の音楽の中から仕事歌と下座音楽を取り上げ、紹介しながら、音楽の機能的な役割を考える一助としたい。
◆仕事歌はもともと、生産に伴う労働・仕事を行うとき、作業能率を高めるために拍子を取ったり、共同作業をするときに動作を合わせたり、先導の役割をしたり、単調な作業に飽きる気持ちを紛らわせたり、労働時の厳しくつらい気持ちを和らげたり、愚痴ったり、または作業の合間に疲れを癒したり、もうひと踏ん張りの元気をだすためなどに歌われてきた。
◆下座音楽は、歌舞伎の演出効果を高めるための音楽で、幕開き、場面や情景を表すもの、人物の登退場、役者の所作に対応するものなどがあり、歌舞伎の演出には欠かせない。
◆あらためて音楽の持つ役割が幅広いことを確認したい。

3-4-2

音楽でがん患者さんをハッピーに！？

〇三浦智史（国がん東）

がんは日本の国民の２人に１人がかかる病気となっており、うまく付き合っていくことが必要な時代になってきています。本講演では、がん患者がどのような時にストレスを感じているのか、がんという病気の経過、がん患者の価値観や思考の変化などについてお話します。

音楽療法はセラピストが関わる能動的音楽療法と、関わらない受動的音楽療法があります。今回は受動的音楽療法について、２つお話をさせて下さい。１つ目は音楽による不安や痛みの軽減効果についてです。音楽や音の利用は、侵襲が少なく手軽であり、かつ、不安や痛みの軽減効果が示されていることから、取り組むことのハードルは高くないと考えます。2 つ目は音楽の利用により期待したいと考えている、がん患者の思考の幅を広げる体験を手軽に行いうるのではないかということです。個人の対処能力の改善は調査票で測定することが難しい部分もありますが、音楽や音を通じて、がん患者の気持ちが和らぎ、がんを克服していくための一助となればと願っています。

3-4-3

3-4-3　公共空間における 立体音響による音楽再生の試み

An attempt of music reproduction using 3D audio in public space

〇亀川徹（東京藝大）

◆店内の BGM や駅の発車メロディーなど，公共空間で音楽を使用する場合，通常の２チャンネルステレオで作成したものを天井などに設置されたスピーカから再生する場合が多い。
◆本講演では，立体音響システムを用いた公共空間における音楽再生の例として，京成上野駅の地下通路などの取り組みを紹介する。
◆事前に再生空間のインパルス応答や暗騒音を収録し，それらを用いてスタジオでシミュレーションした上で，最終的には現地でバランスを調整した。
◆再生空間の音環境を考慮して再生チャンネル数や配置を検討することで，従来の２チャンネル再生では実現できない聴取空間を実現できる。

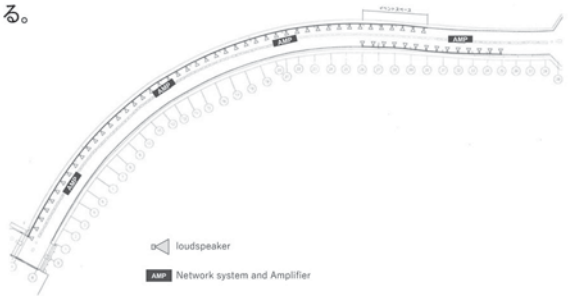


Fig.1：3D sound reproduction system installed in the underground passageway connecting Keisei Electric Railway's Ueno Station and the Tokyo Metro Ginza Line.

3-4-4

3-4-4　我が国における BGM の歴史と現状について

The history and status of BGM in Japan

〇小原良夫（株式会社リプル・JBA）

◆本稿は商業施設や銀行、レストランなどで利用されている BGM について取り上げたものである。
◆我が国で、商業 BGM が始まったのは１９５０年代後半のことで、ホテルの一室から各客室、ロビー、レストランなどに音楽を配信することから始まった。
◆その後、記録メディアになり、当初はオープンテープであったが、フィデリパックというエンドレステープが登場し、爆発的に普及した。記録メディアは、テープから CD へ移り、現在はインターネットを活用した通信型 BGM となっている。
◆商業 BGM は幾つかの転換期を経て、ユーザーニーズに応えてきたが、１９７３年の消防法施工令の改正による非常放送設備の義務化には大きく影響を受けた。それ以降、放送設備と BGM は違う部署が管理するのが一般的になり、トータルな音環境づくりが難しくなった。
◆現在でも BGM に対する理解度は高いとは言えず、商業施設など設計において BGM に配慮したケースは殆ど見られない。
◆BGM を含む音環境は人間の五感に関わる問題であり、これを機に少しでも BGM に興味を持っていただければ幸いである。

3-4-5

3-4-5　自動運転中の車内環境におけるBGMのテンポがドライバの反応時間と主観的印象に及ぼす影響

Effects on driver reaction time and subjective impressions by the tempo of the background music in autonomous vehicle environment

☆尾崎真弘，星野博之（愛知工業大学）

◆目的　本研究では、自動運転中の車室内を模した環境下で「ハイテンポ」および「ローテンポ」のクラシック音楽を聴取しながら視覚反応タスクを行い、ドライバを快適かつ覚醒度の高い状態に保つ音楽を明らかにする。

◆方法　防音室内において高速道路走行中の映像と走行音付きの楽曲（第一実験：6曲、第二実験：4曲）を10分間流し、反応時間のタスク評価を行う。その後、SD法による印象評価を形容詞対「不快-快適」「眠い-目覚めている」「時間遅い-時間早い」「速度遅い-速度速い」「つまらない-楽しい」「集中できない-集中できる」に対して行う。

◆結果　重回帰分析から音楽のシャープネス値と「目覚めている」「快適」の主観値が反応時間に寄与する可能性が示された。（図1）

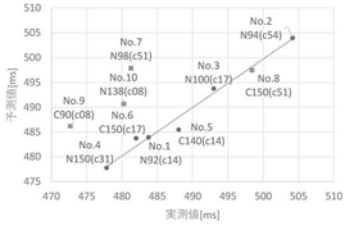


図1　反応時間の予測値と実測値（○：第一実験の重回帰結果
□：第二実験に第一実験の重回帰式を適用した結果）

3-4-6

3-4-6　サウンドロゴによる企業ブランディングに関する研究

Corporate branding with sound logos

☆池田雅子（九州大・芸工 / パナソニック），△杉本美貴（九州大・芸工）

◆企業ブランディングにおいて，音を利用している例は古くからあり，テーマ曲やサウンドロゴでブランドストーリーを伝えている企業も多い。本研究では，企業ブランディングの中でのサウンドロゴの役割と効果，ブランドイメージとサウンドロゴの関係などを探ることで，新たなデジタルメディアで企業のブランドイメージを効果的に伝えるサウンドロゴのあり方を導出することを目的とする。

◆電機メーカーに対して実施した企業サウンドロゴに関するアンケート調査から，サウンドロゴの実態とサウンドロゴに対する企業の想いを把握することができた。また，企業の想いが，どのようにユーザーに伝わっているのかを探るべく実施したユーザー調査からは，企業自体の認知とサウンドロゴの想起に関連性が見られた。

◆サウンドロゴを音響特徴量から考えるべく，ロゴデザインを参考に，サウンドロゴの構成要素，スタイル，デザイン要素について検討した。サウンドロゴのデザイン要素としては『リズム』『メロディ』『和声』『音色』『音声』などがあると考えられ，それぞれの要素が持つ意味やイメージが組み合わさってサウンドロゴとしてのイメージができていると考えられる。

◆今後は，ユーザー調査で使用した６社のサウンドロゴについて，デザイン要素ごとに分析し，企業のブランドイメージを効果的に伝えるサウンドロゴのあり方を導出していく。

3-4-7

3-4-7　教育学部生の生活騒音に対する意識・知識と小学校における音の学習に対する考え

Consciousness of daily life noises, knowledge, and thoughts of sound education in elementary schools by undergraduate students of education

○豊増美喜（大分大），△橋本紗弥（元福岡教育大），△鈴木佐代（福岡教育大）

◆小学校家庭科で「音と生活とのかかわり」の内容が扱われている。本研究では、教育学部生（主に初等課程生）自身の、生活騒音に関する実態や知識、小学校における音の学習に対する考えを明らかにし、今後の授業提案等の基礎資料とする。

◆生活騒音による被害意識・加害感ともに高いのは「話し声や騒ぎ声、歌声」である。生活騒音によるトラブルを減らすためには「時間や場所など各自が考えて行動する」の回答が最も多く、遮音等の対策も必要だが、まずは周囲の人々に配慮することが大切と考える学生が多い。

◆小学校における音についての学習に対する考えについて、「とても必要」の割合が高いのは、「音の感じ方は個人差があることを学ぶ」と「生活騒音の発生に配慮する必要性について知る」で、それぞれ84.6%である。また、「音の伝わり方について学ぶ」の項目は、音の性質を、生活騒音の対策に関連づけできる内容であるが、他の項目と比較して、「とても必要」と考える割合が低い（Fig.1）。

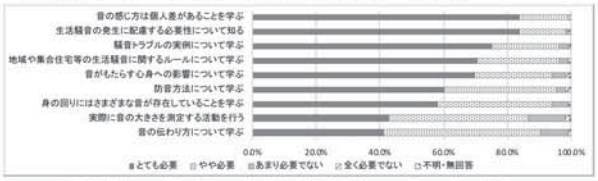


Fig.1：Thoughts of sound education in elementary schools by undergraduate students of education

3-4-8

3-4-8　完全五度の調整課題に関する訓練法の効果について：音大生と音楽経験のある一般の大学生との比較

The effect of the training method for the adjustment task of perfect fifth: Comparison of music college students and university students who have music experience

☆常　慶旻，荒井 隆行，△平山 結佳理（上智大）

◆ 音楽演奏者にとって音程を正確に調整できることは重要である。しかし、今まで音程を調整する能力について重視されてこなかったのは大きな問題である。

◆ 我々は以前から音程調整の正確性を改善することに注目し、完全五度の倍音とうなりの特性を利用した訓練法を提案した。本稿では、音楽経験のある一般大学生と音大生の 2 グループに対象をして、訓練法の有効性について検証した。

◆実験 1 と 2 はそれぞれに、一般の大学生・音大生を対象にして、Training 前と Training 後の改善率について調べた。実験は Pre-test（feedback なし）→ Training（feedback あり）→ Post-test（feedback なし）の順で行った。

◆両実験とも、訓練後の結果は訓練前より改善することができた。特に一般の大学生では、60%以上に改善することができた（以前の実験結果）。一方、音大生の pre-test の音程の正確性は一般の大学生より高かった。音大生の Training 前の音程の正確性が元々高かった原因もあり、Training 後の改善率は一般大学の学生より低かった。しかし、音階にない微分音の課題の結果では、音大生の Training 後の改善率（80%以上）は非常に高かった。この訓練法は音大生と音楽経験のある一般の大学生の両グループとも有効であると考えられる。

3-4-9

3-4-9　プログラミング学習における音読の有効性に関する検討-その２

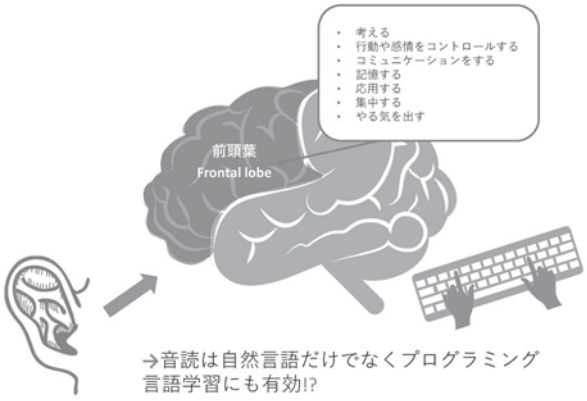
A Study on the Effectiveness of Reading Aloud in Programming Learning - Part 2

○上田麻理，小田桐空大，春日秀雄，田中哲雄(神奈川工科大)

＜あらまし＞

自然言語学習における音読の有効性に関する知見を参考に，プログラミング言語学習にも音読が有効だと考えた．

前報では，大学生のプログラミング言語学習における音読学習の有効性を明らかにするために，音読プログラミング学習システムを作成した．本稿では，引き続きプログラミング言語学習における音読の有効性を確認するために，システムの改良及び，講義やゼミナール等で実践したのでその結果を報告する．



3-4-10

3-4-10　「ST 養成校の音響学教師のつながり」を作る

Making "connections among acoustic teachers of the speech therapist courses."

○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)

◆ことばのリハビリを行う言語聴覚士（S T）の養成校では，音響学が必修科目であるが，学生にとって最も苦手で嫌いである科目であるとされる。

◆臨床に結び付いた，分かりやすい授業にするためには，音響学教師はＳＴの臨床について知る機会や，他の音響学教師との交流や，授業見学などの機会が必要である。

◆本発表では，養成校の音響学教師間のつながり，さらには現役 ST とのつながりを作る活動を紹介する。

◆音響学・聴覚心理学教師のための Facebook の非公開グループを作成した。

◆現役ＳＴのために開催していた講習会「ＳＴのための音響学」に音響学教師間や言語聴覚士との交流や，教授法を学ぶ会などの機能を持たせた。

◆言語聴覚士側は，音響学教師に対して具体的な要望を持ち，今後の共同作業を歓迎していることが分かった。

◆音響学教師側は、悩みはあるが，まだ参加している者が少ないこともあり，積極的につながろうと考える者が少ない結果となった。

◆今後の課題は，Facebook 等と講習会を継続するとともに，全国の言語聴覚士養成校の音響学・聴覚心理学の教員に対する調査をし，より多くの音響学教師の意見をうかがっていきたい。

3-4-11

3-4-11　巨大メガホンを用いた音声の長距離伝達実験

Experiments of long-distance speech transmission using huge megaphone.

○高橋 義典 (工学院大)，△金 孝義 (NEXTEP(株))

* 博物館等のポピュラーな屋外展示としてパラボラ集音器が挙げられる。パラボラはアンテナとしては一般的であるものの，音響学の分野では，ホーンの方が日常生活でよく目にするデバイスである．
* 本報告では，2022年4月29日のNHK総合および2022年7月8日BSプレミアムで放送された「世界！オモシロ学者のスゴ動画祭３」における巨大メガホンを使った実験「人間の声はどこまで届くのか？」の実験について述べる．
* 遮断周波数100Hzのハイパボリックホーン(Fig. 1)を使用したところ，2kmまで音声が届くことが確認できた．
* シミュレーションではRASTIが60%以上となる距離は，暗騒音レベルが40dBの静かな環境であれば3.5km程度確保できるものの，50dBになると2km程度となると予想される(Fig. 2)．
* 本実験で得られた知見から，博物館等の屋外展示として巨大メガホンの設置する際には，展示付近での来場者の話し声なども考慮して，巨大メガホンと受聴者の距離は500m～1km程度とすることが望ましい．また，音声が発せられるタイミングを受聴者が把握できないと，受聴者は長距離伝送されてきた音声を聞き漏らしてしまう可能性があることから，話者と受聴者との間でお互いの様子を確認できるモニタの設置が望ましい．

Fig. 1　huge megaphone

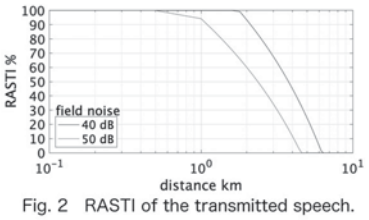


Fig. 2　RASTI of the transmitted speech.

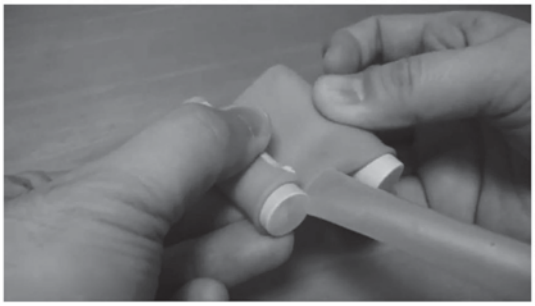
3-4-12

3-4-12　声帯模型の試作に関する一考察

A Study on Making a Functional Vocal Cords Model

○青木 直史（北大）

模型教材の活用は五感に対する訴求力に優れており，初学者の興味を引きつけるアプローチとして教育的効果が高い．本研究では，これまで検討事例が少なかった声帯をターゲットとして，機能に焦点をあてた実体モデルの作成を試みている．



シリンダ方式の声帯模型

デモ動画：https://youtu.be/afKG1YsnhNQ

3-4-13

3-4-13　管理栄養学科における音響教育

Acoustic Education in the Department of Nutrition

○上田麻理，饗場直美，田中哲雄（神奈川工科大学）

＜あらまし＞

プロの料理人らは，調理状態の表現に「シュワシュワという音がしたら揚げ時である」等の擬音語を使用しており，調理状態を耳で判断するといったことが日常的に行われているようである．それらの音の表現は感覚的であることが多く，食品の調理状態や物性値の対応が明確でない場合がある． 調理時に発生・利用する音の感覚的な表現を音響学的に理解できれば，より的確に調理状態を再現することができると考えた．

そこで，今回はプロ或いは料理に係る専門家をめざす学生向けの音と調理支援アプリケーションを開発したので報告する．

3-4-14

3-4-14　聴能形成システムのオーディオネットワークを用いた構成事例

Design of technical listening training system using audio network

○河原一彦，山内勝也（九州大・芸工）

九州大学大橋キャンパスには，前身である九州芸術工科大学設立当時から，聴能形成を実施する専用教室（「音響心理実験室」と称された）が設置されていた。しかし築後約 50 年を経て，2021 年に全面的な改修工事が行われた。

改修工事にともない，聴能形成システムの音声回線についても構成を刷新した。九州大学の聴能形成システムでは，音声再生用の PC からの音源送出だけでなく，準備室と教室との間の音声コミュニケーションの回線が必要である。本事例では，それらの音声回線をオーディオネットワーク規格 “Dante” を用いて配線を敷設した。

本報告では，オーディオネットワークを用いた聴能形成システムの構成計画，および設置調整，運用に関して報告する。

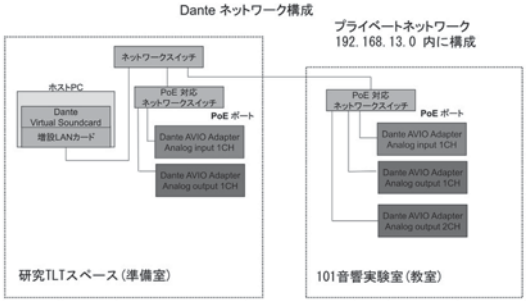


Fig.1：本報告で構成した Dante ネットワーク

3-6-1

3-6-1　光音響を用いた油中の微粒子観測*

Observation of fine particles in oil with photoacoustic effect

☆古谷彬人、小池義和（芝浦工大）△福田啓輔△ラジャゴパランウママヘスワリ

◆現在、油などの液体中の微粒子濃度を計測可能となる技術が求められている。本研究では光音響効果を使用した油中の微粒子濃度測定技術について検討を行う。光音響効果とは物体に断続的な光エネルギーを与えた際に物体が光の熱による振動から音波を発生する現象であり、本研究では発生した音波から油中の微粒子観測を行う。

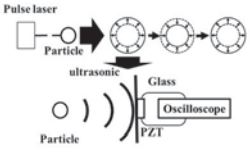
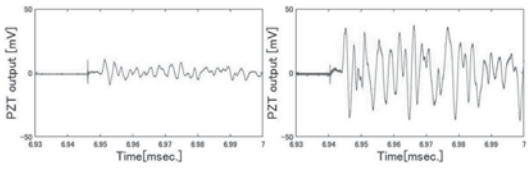


Fig.1　Photoacoustic effect

◆ 油中の微粒子から発生した音波を外部に設置した PZT 素子により測定した電圧値の波形を Fig.2 に示す。粒子濃度の増加によって波形の振幅値に差が出ていることが分かった。粒子や PZT、油の容器等の共振周波数から音波の周波数成分について考察していく。



(a) 0g/L　　(b) 120mg/L

Fig.2　Waveform with Photoacoustic effect

3-6-2

3-6-2　超音波センサを用いた撹拌擂潰状態の進行状況推定の検討

Study of Estimating the Progress of Stirring and Grinding Process Using ultrasonic sensors

☆林僚哉（芝浦工大），△WangkanklangEkkawit（NRRU），小池義和（芝浦工大）

◆ 石川式撹拌擂潰機（図１参照）は、乳鉢と二本の杵、撹拌棒、モーターから構成されている。また、擂り潰し、撹拌、分散、混錬の複数作業の同時処理を行うことができ、工業材料、医薬品、食品加工などに幅広く利用されている。近年、発展が著しい IoT 技術を用いて遠隔に計測、推定するシステムを導入する必要がある。例えば、石川式撹拌擂潰機では、被擂潰物の撹拌擂潰状態の進行度合いを確認する際、一度工程を停止し、被擂潰物のサンプルを抽出し、サンプルの状態を顕微鏡などで確認する必要があり、使用者にとって負担と手間がかかる。そこで、遠隔で撹拌擂潰推定状態を実現する必要がある。

◆ 本研究では、石川式擂潰機の撹拌擂潰時に発生する音からパワースペクトル密度（PSD）を測定し、撹拌擂潰状態の進行状況推定方法を提案し、精度の確認、システムの実現を目指している。

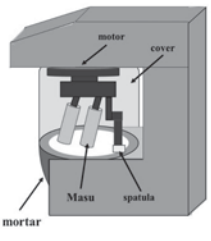


図 1. 石川式撹拌擂潰機

3-6-3

3-6-3　パラメトリック差音を用いた音速画像の取得

Sound speed image using parametric sound at difference frequency

☆井ノ口楓, 野村英之（電通大）

◆本研究では同サイズの音源から直接放射した低周波数超音波と比較して指向性が鋭いパラメトリック差音を用いた CT (Computed tomography)法により，対象物の音速画像の取得と音速推定を行った。

◆Fig.1 にアルミニウム丸棒の音速推定の結果を示す。高周波超音波とパラメトリック差音で得られた結果に大きな差異はなく，400 kHz のパラメトリック差音を用いた場合において，空間分解能が良い 2.2 MHz の高周波超音波と同精度で音速画像の取得と音速推定が可能であることがわかった。

◆今後内部構造を持つ対象物での実験を行うことにより，パラメトリック差音を用いた低周波超音波画像法実現の検討を行う。

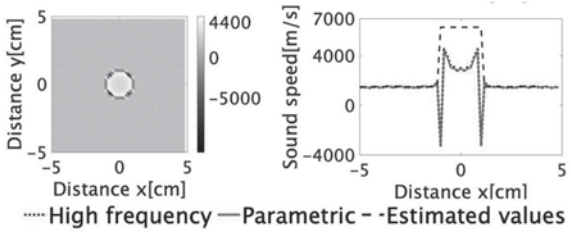


Fig. 1 Obtained sound speed image(left) and estimated sound speed (right) by parametric sound.

3-6-4

3-6-4　2MHz焦点型圧電高分子トランスデューサを用いた空中超音波によるIC内部の位相差画像形成

Phase difference imaging of Integrated Circuit with through-transmission method at 2 MHz in air by using focal type P(VDF/TrFE) transducers

○高橋貞幸(山形大学)

超音波による観測対象物の画像形成には、主に 2 つの方式がある。一般的に用いられる方式は、受信した超音波の振幅強度差による画像形成である(Amplitude-imaging)。この方法は、一定電圧の送信超音波からの反射(または透過)強度を数値化するため、送受信回路も比較的容易であり、頻繁に用いられる。一方、送信超音波と受信波間の時間(位相)差を用いて、画像形成を行う方法が、位相差画像(Phase difference imaging)である。位相差法が強度法よりも画質向上を期待できる。本研究では、集積回路(Logic-IC)[11229-12(ROCKWELL)]を、Fig.1に示すシステムにより位相差画像形成を行ったことを報告する。

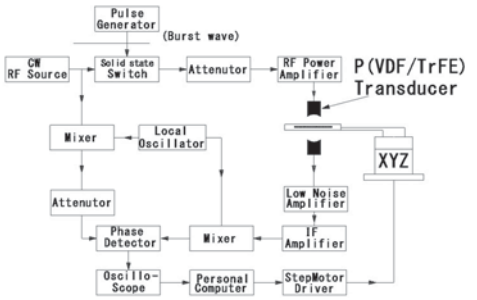


Fig.1 : Block diagram of the experimental ultrasonic imaging system with using P(VDF/TrFE) transducers.

3-6-5

3-6-5　音響放射力励起による過渡応答を用いた欠損検出法の検討

Defect detection using transient response excited by acoustic radiation force

☆北村香子, 野村英之(電通大)

我々はこれまでに，音響放射力によって生じる過渡振動応答から対象物の振動特性を取得する方法を検討してきた．本報告では，本手法を物体の欠損検出へ応用することを目的とする．特に，円形状に周端固定したアルミニウム箔を一次元走査し，欠損の有無による違いを評価した．

Fig. 1 に取得した振動特性分布を示す．Fig. 1 (A)より無欠損の場合は対象物の振動モードに対応するような分布が取得された．一方で欠損を加えた場合は，Fig. 1 (B)より，切り目左右で振動モードに差が生じることから取得分布に不連続性が表れた．

これより，欠損の有無によって異なる振動特性分布が取得されることがわかった．

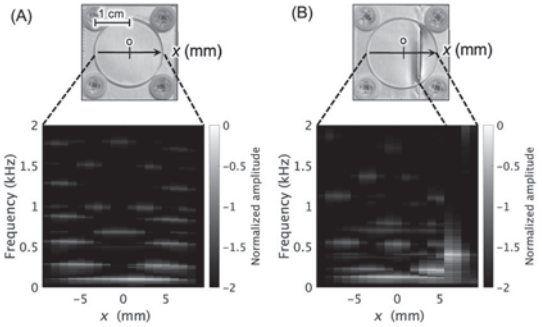


Fig.1 : Distribution of vibration characteristics of aluminum foil. (A) without defect (B) with defect at x = 5 mm.

3-6-6

3-6-6　基板実装型超音波液晶レンズの検討

Board-mounted ultrasonic variable-focus liquid crystal lens

☆黒田悠真[1]　ジェシカオナカ[1]　△原田裕生[1]　△江本顕雄[2]
松川真美[1]　小山大介[1]（1. 同志社大, 2. 徳島大）

◆<u>透明電極と機械的機構を必要としない</u>，<u>超音波振動</u>を用いた<u>可変焦点液晶レンズ</u>について，基板への実装方法を検討した．

◆<u>基板にレンズの振動子内径と同径の穴</u>を設け，接着することにより，レンズ中心にたわみ振動モードを励起することができた．(Fig. 1(a))

◆レンズの入力電流が大きいほど焦点距離が小さくなる，<u>可変焦点型凸レンズ</u>を作製することができた．(Fig. 1(b))

◆基板への実装はレンズの<u>温度上昇</u>および<u>液晶の相転移</u>を防ぐことに繋がり，従来型と比較し<u>可変焦点範囲を拡大</u>できる可能性が示唆された．

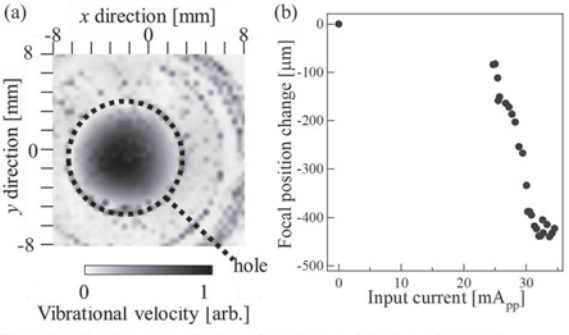


Fig.1 : (a) Vibrational distribution of the lens and (b) the relationships between the input current and the focal length

3-6-7

3-6-7　たわみ振動板と反射板による多数液滴の同時浮揚と同時滴下の検討

Simultaneous levitation and dispensing of multiple droplets using 28-kHz flexural vibrating plate and reflector

☆Wang Yilin，和田 有司，中村健太郎（東工大）

◆A horizontal setup of a parallel pair of vibrating plate and reflector was studied for levitating multiple droplets of liquid.

◆The length of a 28-kHz flexural vibration plate was experimentally trimmed to resonate with the excitation system composed of a Langevin transducer and a horn.

◆Vertical and horizontal configurations for reflector-vibration-plate setup were investigated.

◆In horizontal configuration (Fig. 1), a vibrating plate and a reflector are parallelly located in side-by-side.

◆Vibration amplitude on the plate required for levitation of droplet was numerically studied through finite element analysis of the ultrasonic field and the resultant acoustic radiation force.

◆We succeeded in levitating five water droplets with an equal interval as demonstrated in Fig. 2.

◆Maximum volume for each droplet was up to 1 μL.

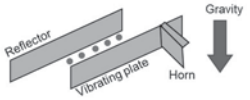


Fig. 1 Horizontal Setup for multi-drop levitation.

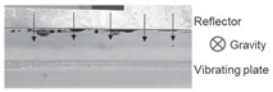


Fig. 2 Top-view photo of multi-drop levitation.

3-6-8

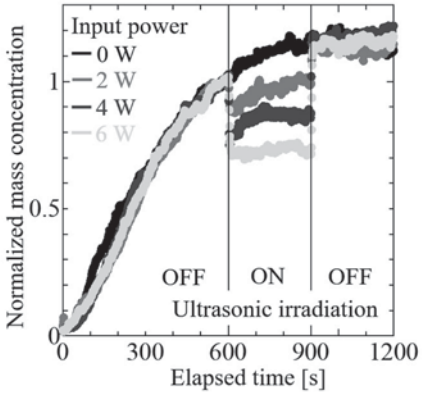
3-6-8　小型空中超音波源を用いた容積の異なる装置による煙霧質の凝集

Agglomeration of aerosol by devices of different volumes using small aerial ultrasonic source

☆小野湧喜，淺見拓哉，三浦 光（日大・理工）

◆本研究では，煙霧質が通る流路の長さが異なった装置を製作し，単位容積あたりの入力電力を一定にした凝集実験を行い，容積が凝集効果に与える影響を検討した。

◆下図は凝集実験の結果である。図より，煙霧質の濃度の低下は容積が大きくなるほど低下することが分かった。凝集率はそれぞれの装置において，凝集装置１（入力電力２W）で 11 %，凝集装置２（同４W）で 21 %，凝集装置３（同 6W）で 32 %であった。



Agglomeration experiment results

3-6-9

3-6-9　定在波音場中で捕捉物体の大きさにより変化する力の評価

Evaluation of forces varying with the size of trapped objects in standing wave fields

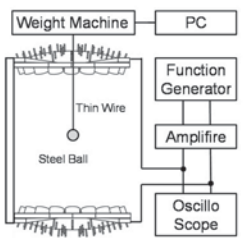
○小塚晃透，桶田大晟（愛工大），豊田昌弘（本多電子），畑中信一（電通大）

◆空中で上下２方向からの超音波の干渉により形成される定在波音場中で、物体に作用する力の評価について研究を行った。

◆糸で吊した鉄球を音場中に投入し、電子天秤で荷重の変化を測定した（Fig. 1）。音響放射圧による力は、音圧の節との相対位置により異なるため、周波数に差を与えて、音圧の節を下から上に移動させた。

◆音圧の節の移動に合わせて、最大荷重（音圧の節が下方にあり、鉄球を引き下げる）から徐々に最小荷重（音圧の節が上方にあり、鉄球を持ち上げる）まで変化する様子が確認された（Fig. 2）。その後、力の作用しない状態を経て最大荷重に戻り、繰り返すことが確認された。

◆鉄球の大きさによる影響は、超音波の半波長より小さい範囲で、鉄球が大きいほど作用する力が強いことが分かった。



Fi g.1 : Experimental apparatus

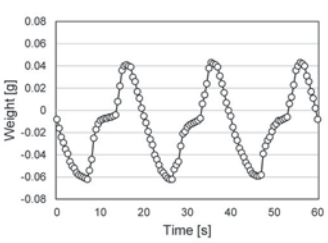


Fig. 2 : Experimental result (ϕ 3 mm)

3-6-10

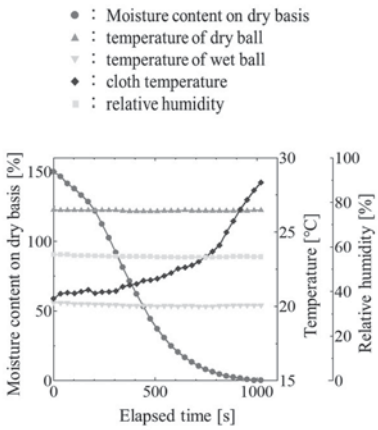
3-6-10　空中強力超音波による濡れた布の乾燥中の温度

Temperature during drying of wet cloth by aerial intense ultrasonic

☆伊藤美桜，淺見拓哉，三浦 光（日大理工）

◆筆者らは，空中強力超音波を用いた濡れた布の乾燥について検討を行っている。本稿では，これまで検討されていなかった乾燥中の試料の温度について検討を行った。

◆2 台の空中超音波源の振動板面が平行かつ向かい合うように設置し，板間に強力な定在波音場を形成した。定在波音場内の音圧の節の位置に試料を設置し，赤外線サーモグラフィカメラを用いて試料の温度の測定を行った。図より，試料の温度は恒率乾燥期間までは湿球温度付近であるが，減率乾燥期間に入ると時間の経過に伴って上昇し，測定終了時には乾球温度を上回ることが分かった。



Moisture content and temperature change.
The total input power is 10 W.

3-6-11

3-6-11　水中定在波音場における水出し珈琲の抽出促進

Promotion of extraction of cold brew coffee in the underwater standing wave field

☆渡辺啓介，浅見拓哉，三浦　光（日大・理工）

◆周波数 20 kHz の水中定在波音場を用いて，水出し珈琲抽出中に超音波を照射した場合のクロロゲン酸の抽出促進について検討した。

◆超音波照射による水温上昇を鑑み，超音波照射しない場合も水温を追従させて検討した。

◆クロロゲン酸抽出量をペーパークロマトグラフィーの結果から画像処理を施し，簡易的に可視化した。

◆Fig. 1 は超音波照射有り，Fig. 2 は超音波無しの散布図を示す。これらより，超音波照射有りの方が広い範囲に分布しており，超音波を照射した方がクロロゲン酸を多く抽出できることが分かった。

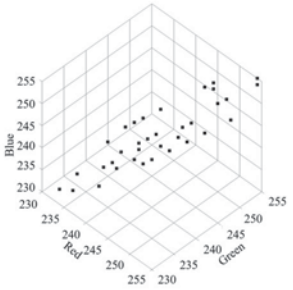


Fig.1：Scatter plots of RGB light in ultrasonic irradiation.

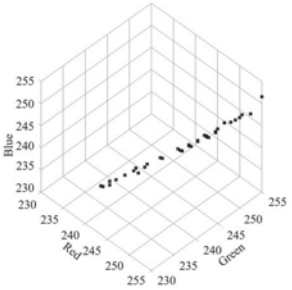


Fig.2：Scatter plots of RGB light without ultrasound.

3-6-12

3-6-12　小型熱音響システムの実現に向けた管内音波の波長とスタック長さの関係が熱音響効果に与える影響

Influence of the relationship between the wavelength of sound in the tube and the length of the stack on the thermoacoustic effect for the realization of a compact thermoacoustic system

☆小野　悟，坂本眞一（滋賀県立大）

◆熱音響システムの小型化に向けて，管内音波の波長とスタックの長さに着目した検討を行った.

◆メッシュ数が 50 の金属メッシュを用意した.

◆無次元量$\omega\tau\sim\pi$となる音波の周波数を算出し，音波の 1 波長と同等の長さとなるようシステム全長を設定した.

◆システム全長 0.67 m，音波の周波数 520 Hz の条件のもと，スタックの長さを変化させ，スタック両端温度差を測定した.

◆スタックの長さ範囲が 1~12 mm のとき，スタック両端温度差が形成され，10 mm のとき最大となることを確認した.（Figure）

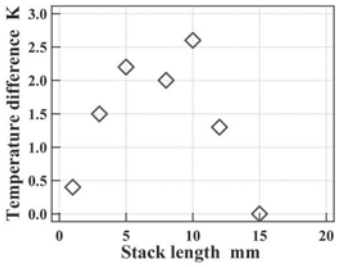


Figure　Experimental result for stack length.

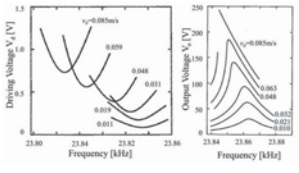
3-6-13

3-6-13　圧電素子の「跳躍・降下現象」の機構解明

Elucidation of the mechanism of "jumping and dropping phenomena" of piezoelectric elements

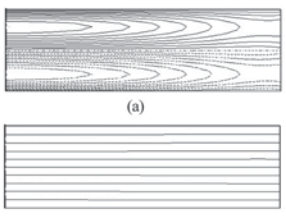
○海老原拓矢，足立和成（山形大院・理工学研），山吉康弘（山形大・工　技術部）

定電圧駆動される圧電素子の機械共振周波数付近で起きる「跳躍・降下現象」は、振動速度の増大に伴って顕在化する圧電材料の弾性的な非線形性に起因するものと従来されてきた。しかし、定動電流駆動時には同現象が起きないことから(Fig. 1)、その原因を弾性的な非線形性に帰することには無理がある。圧電性を考慮した有限要素法による振動解析を行なったところ、圧電素子の内部電界にはその機械共振周波数付近において著しい乱れが生じており(Fig. 2)、同現象はこの乱れに伴う電界集中がドメインにもたらす局所的分極反転などに起因している可能性が高い、と考えた。そこで、数値解析による同現象の大まかな定性的再現を試みるとともに、この仮説の実験的検証を行ったので報告する。



(a)　(b)

Fig. 1 Resonance curves of the driving voltage and the output voltage proportional to the in-flow electric current of a piezoelectric element under the conditions of (a) constant motional current and (b) constant applied voltage respectively, with the vibration velocity v_0 as a parameter [S. Takahashi et al., Jpn. J. Appl. Phys., 31, 1992, p. 3055].



(a)

(b)

Fig. 2 Calculated axisymmetric electric fields inside a circular piezoelectric ceramic disk at its (a) resonance and (b) anti-resonance frequencies. The solid and dashed thin lines represent positive and negative equipotential contours, respectively.

3-6-14

3-6-14　市販 FEM 最適化ソフトを用いた最適化試行 —超音波振動工具のトポロジー最適化(3)—

On the optimization trial using commercial finite element software – Topology optimization of large ultrasonic tools (3) –

○和田有司，中村健太郎（東工大・未来研）

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

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

◆市販の有限要素法解析ソフト(COMSOL 6.0)の最適化モジュールを用いて振動工具の振動分布を平坦にする手法について検討する。

◆目的関数は複数点の入出力内積のうち、最小のものを最大化する MINMAX 法を p-norm 法により近似したものを使用する。

◆Fig. 1(a)が最適化の設計領域、(b)が最適化後の振動分布であり平坦な振動分布を得ている。 (c)が最適な材料配置である。

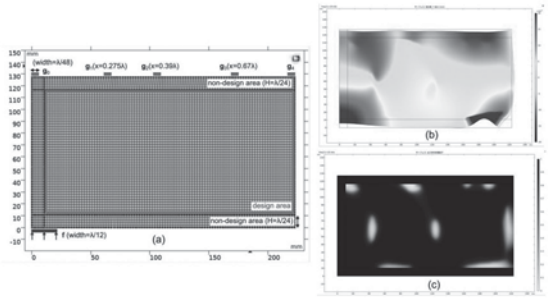


Fig.1：(a) Design area and optimal material distribution, (b) vertical displacement after optimization, and (c) optimal shape

3-6-15

3-6-15　強力超音波用ボルト締めランジュバン型振動子の最適設計に関する研究

Study on Optimal Design of Bolt-Clamped Langevin Type Transducers for Strong Ultrasonic Waves

☆山崎凌汰，足立和成（山形大院・理工）

大振幅を必要とする強力超音波の振動源として、一般的にはボルト締めランジュバン型振動子(以下 BLT)が用いられている。BLT は圧電セラミックスと電極板を交互に重ね、金属ブロックで挟み込みボルトで締めつけて一体化した構造となっている。BLT を高周波化するべく単に金属ブロックの全長のみを短くすると、部品間の界面外周部において静的接触応力が不足し、大振幅駆動時に動的振動応力が静的接触応力を上回り部品間界面での接触が十分に保てなくなってしまい、機械的損失の増大やそれに伴う圧電セラミックスの特性劣化を引き起こす可能性がある。本研究ではそうした問題を克服し、有限要素解析を用いて大振幅駆動時でも部品間界面での接触を十分に保ちながら高い周波数をもつ BLT の構造を検討した。

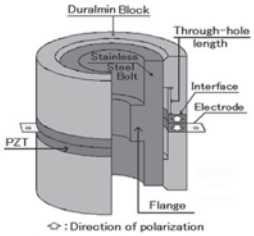


Fig.1: A Hollow cylindrical bolt-clamped Langevin type transducer

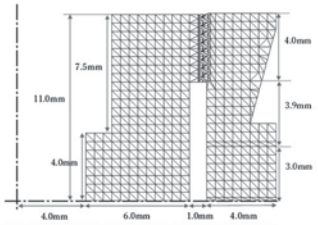


Fig.2: Finite-element representation of an optimized BLT structure

3-6-16

3-6-16　BLT と一体構造とした小型空中超音波源

Compact aerial ultrasound source integrating with BLT

☆大渕　稜太（日大・理工），△笠島 崇，△伊藤伸介（日本特殊陶業），
浅見　拓哉，三浦　光（日大・理工）

◆空中超音波センサなどでは小型であることが求められており，強力な音波が発生しにくいことが問題になっている。

◆本稿では振動面を BLT と一体構造とした小型空中超音波源の開発を目的としている。

◆シミュレーション解析により Fig. 1 に示したジオメトリ図の寸法を求め，実際に製作を行い特性の測定を行った。

◆Fig. 2 は入力電力と音圧の関係を示したグラフである。図より，音圧は入力電力の増加に伴って上昇し，入力電力 8.5 W で最大音圧 513 Pa（音圧レベル 148 dB）の強力な音波を放射できることが分かった。

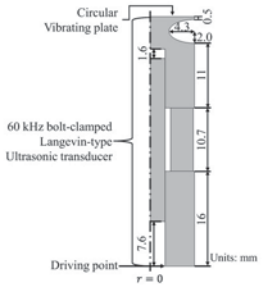


Fig. 1. Outline of the sound source.

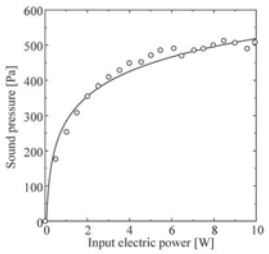


Fig. 2 Relationship between sound pressure and input electric power.

3-6-17

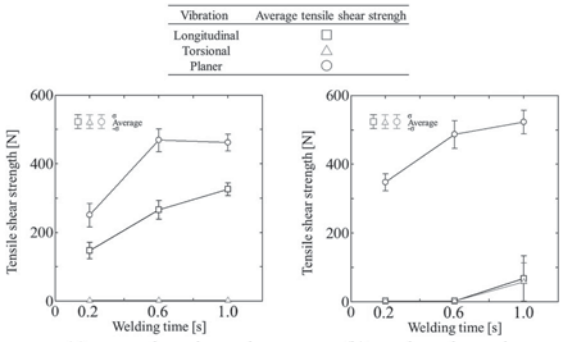
3-6-17　超音波複合振動源による接合試料の表面粗さと接合強度の関係

Relationship between surface roughness and weld strength using ultrasonic complex vibration

☆山崎梨菜，浅見拓哉，三浦　光（日大・理工）

◆本検討では，試料の表面粗さが超音波接合に与える影響を検討するため，元の状態のアルミニウム板の場合とアルミニウム板にエメリー紙で表面を粗くした場合について，銅板との接合実験を行い，引張せん断試験の結果から接合強度の比較検討を行った。

◆図(a)は元の状態のアルミニウム板を用いた接合実験の結果，図(b)はエメリー紙 #220 で表面を粗くしたアルミニウム板を用いた接合実験の結果である。両図より，面状振動の接合は表面を粗くしても接合強度に大きな影響はないが，縦及びねじり振動の接合においては影響が大きいことが分かった。



(a) no roughened samples.　(b) roughened samples.

Relationship between tensile shear strength and welding time.

3-6-18

3-6-18　超音波溶着加工時に発生する非線形振動の分析

Analysis of nonlinear vibration generated under ultrasonic welding.

☆矢部光平（室蘭工大），笠井昭俊，宮田勝，河野太郎（精電舎電子工業），
青柳学（室蘭工大）

◆Fig.1 の測定システムを用いて，溶着中に土台中を伝搬する振動を計測し，FFT 解析による周波数分析を行った。

◆Fig.2 は，溶着過程および試験片の作製方法を示した図である。

◆Fig.3 は，引張試験で得られた溶着時間に対する平均溶着強度（AWS）である。0.38 s 以降から溶着強度が変化しないことが確認できた。

◆Fig.4 は，時間による沈み込み量，電力および THD を示したグラフである。THD および電力は，0.38～0.40 s に極大値および極小値を持っており，Fig.3 の溶着強度が最も高い時刻とほぼ一致した。

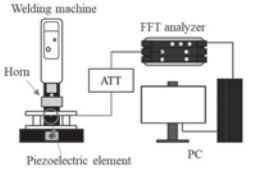
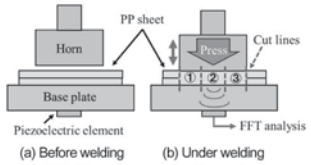


Fig.1 Measurement system.



(a) Before welding　(b) Under welding

Fig. 2 Welding process and cutting method.

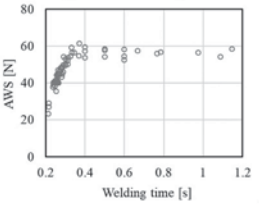


Fig.3 AWS vs welding time.

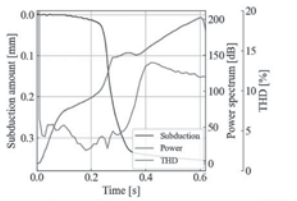


Fig.4 Time variation of subduction amount, power and THD.

3-6-19

3-6-19　空中超音波励起によるガイド波パルス圧縮のための音源駆動条件の検証

Verification of sound source drive conditions for guided wave pulse compression by airborne ultrasound excitation

☆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(Airborne Ultrasound Phased Array:AUPA)を用いた弾性波源走査法による非破壊検査について研究を行っている.

◆本報告では，空中超音波励起によるガイド波の分散性を考慮したパルス圧縮の基礎検討として，検査に用いる空中超音波音源の放射音波が最大音圧に達するまでの最短時間について検証した.

◆その結果，いずれの周波数においても，入力波数 20 波以内で放射音波が最大音圧にあることを確認した.

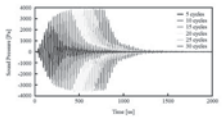


Fig. 1. Time waveform (40 kHz).

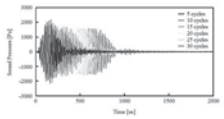


Fig. 2. Time waveform (45 kHz).

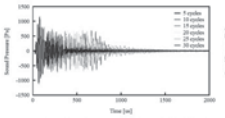


Fig. 3. Time waveform (50 kHz).

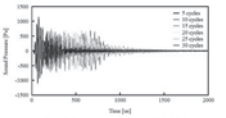


Fig. 4. Time waveform (55 kHz).

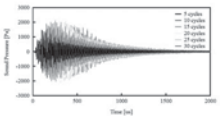


Fig. 5. Time waveform (60 kHz).

3-6-20

3-6-20　非線形空中超音波を利用したピッチキャッチ法による金属薄板を対象とした非接触非破壊評価

Non-contact and non-destructive evaluation of thin metal plates by pitch-catch method using nonlinear airborne ultrasounds

☆濱田郁哉，清水鏡介(日大院・理工)，大隅歩，伊藤洋一(日大・理工)

◆非線形空中超音波と受信器アレイを組み合わせた非接触非破壊検査方法について研究を行っている.

◆提案手法を実際の非破壊検査へ適用するには，音源と受信器アレイを検査対象に対して同一面に設置するピッチキャッチ法での計測が必要となる.

◆非線形空中超音波励起によるピッチキャッチ法を用いた金属薄板内欠陥の非接触非破壊検出について実験的に検討を行った.

◆結果より，提案手法による欠陥の可視化を確認した.

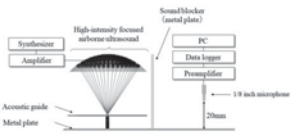


Fig. 1. Schematic of experimental equipment.

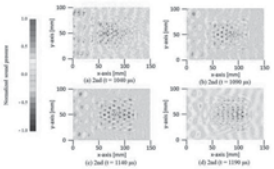


Fig. 2. Experiment results (with sound blocker).

3-8-1

3-8-1　明示的な音声特徴量に基づくDNN音声合成

DNN speech synthesis based on explicit speech features

☆後藤仁，小澤凜夏（千葉工大），竹本浩典（千葉工大），平井啓之，前川喜久雄（国語研）

◆本研究では，MRI 動画からフォルマント・F0・非周期性指標を抽出できるか検討し，これらを話者埋め込みベクトルとして音声を合成して話者性を評価した。

◆フォルマントは声道断面積関数を抽出して伝達関数を計算することで求めた。F0 は喉頭周辺の画像と実音声の F0 を学習させた CNN により抽出した。しかし，声門の開度などに由来すると思われる非周期性指標は画像から抽出することができなかった。

◆そして，MRI 動画から抽出したフォルマント・F0 と実音声から抽出した非周期性指標を埋め込みベクトルとして音声を合成した。その結果，従来の x-vector を用いて合成した音声と同程度に話者類似度を再現できたが，いずれも値は低かった。その要因として，録音した音声の暗騒音や MRI 撮像時の姿勢の影響などの影響が考えられる。

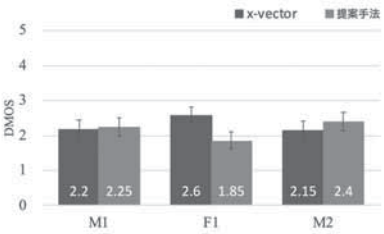


Fig.1 : Results of speaker similarity evaluation between x-vector and the proposed method

3-8-2

3-8-2　音声認識コーパスを用いた高品質複数話者テキスト音声合成に向けたニューラルボコーダによる帯域拡張

Bandwidth extension with a neural vocoder towards high-fidelity multi-speaker text-to-speech trained 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)

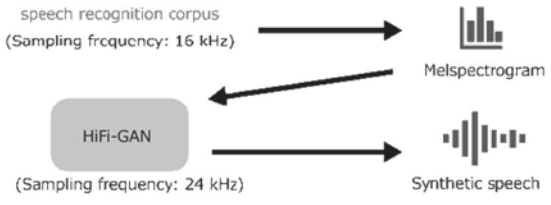
◆通常，TTS やニューラルボコーダの学習には，広帯域かつクリーンな音声合成コーパスが用いられる。

◆全ての言語において，十分な複数話者コーパスがあるとは限らない。

◆狭帯域かつ雑音を含むことのある音声認識コーパスを用いたサンプリング周波数(fs)を 24 kHz とする複数話者 TTS の実現を検討する。

◆本稿では，上記の初期検討として音声合成コーパス(fs:24kHz)で学習した複数話者 HiFi-GAN を用いて，音声認識コーパス(fs:16kHz)の帯域拡張を行った。

◆MOS 評価において，音声合成コーパスでは，合成音声(fs:24kHz)がそれをダウンサンプリングした合成音声(fs:16kHz)を上回った。音声認識コーパスでは，その２つに大きさ差は見られなかった。音声認識コーパスの音声合成コーパスと比べて低い音質が原因であると考える。



3-8-3

3-8-3 Interpretable Emotional Control for Text-to-Speech System toward Development of Sympathetic Educational-Support Robots

共感型教育支援ロボットの実現に向けた
テキスト音声合成システムにおける解釈可能な感情制御

☆Jingyi Feng (Nagoya Univ.), △Tomohiro Yoshikawa (Suzuka Univ. of Medical Science), Tomoki Toda (Nagoya Univ.)

- A speech synthesis system is constructed for sympathetic educational-support robots for emotional expression and sympathetic expression (changes in an emotional expression).
- The emotion categories and the intensity of the emotional expression in the synthesized speech can be controlled by implementing a GSTs module and an A-V tokens module into an original Fastspeech2 model.
- In the synthesis, the interpretability of arousal and valence values in the emotion space is exploited to achieve interpretable control of the intensity of the synthesized speech emotional expression by changing the arousal and valence values.

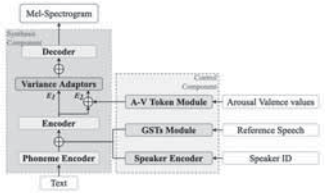


Fig.1 : Overview of Proposal structure
(Participants □:Spring, ■:Fall, Papers ○:Spring, ●:Fall)

3-8-4

3-8-4 テキストを入力とした音声・顔ランドマーク系列の同期生成の検討

An investigation of synchronized speech and facial landmark sequence generation from text

◎三井健太郎，沢田慶(rinna)

- ◆同期された音声・顔動画を生成する３つの枠組み
  - Sequential: テキスト→音声，音声→顔動画を独立に生成
  - Separate: テキスト→音声，テキスト→顔動画を独立に生成
  - Parallel: テキスト→音声・顔動画を１つのモデルで同時生成
- ◆FastSpeech に基づく３つの構造を提案・比較
  - 音響特徴量・顔ランドマーク系列のフレームレートの差異を２種類の length regulator を用いて表現
  - いずれの枠組みも高品質な顔ランドマーク系列を生成可能
  - 単一話者実験では Sequential が最も高品質な lip-sync を達成
- ◆複数話者音声・単一話者顔動画を用いて話者共通の顔ランドマークを生成する応用を検討
  - 音声と顔動画で独立に学習および生成が可能な Separate が最も高品質な顔動画，lip-sync を達成

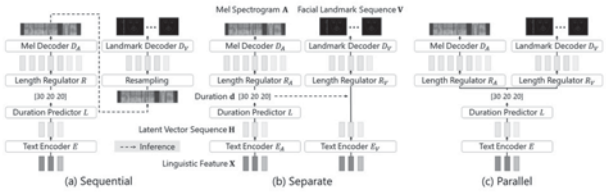


Fig.1: Model structures for Sequential, Separate, and Parallel generation of synchronized mel spectrogram and facial landmark sequence from text

3-8-5

3-8-5 フィラーを含む自発音声合成モデルの品質低下原因の調査と一貫性保証による改善

Investigation of causes of quality degradation in FP-included spontaneous speech synthesis model and quality improvement with consistency guarantee

☆松永裕太，佐伯高明，高道慎之介，猿渡洋（東大）

- ◆課題：
  - 自発音声合成モデルにおけるフィラー挿入による品質低下
- ◆本稿：
  - モデルのどこでフィラー挿入による影響が生じているかを調査
  - フィラーを含む場合と含まない場合の非フィラー部の一貫性を保証する学習手法 (一貫性保証学習) を提案し，有効性を検証
- ◆主観評価結果：
  - 提案手法によりフィラーあり合成音声の自然性が改善

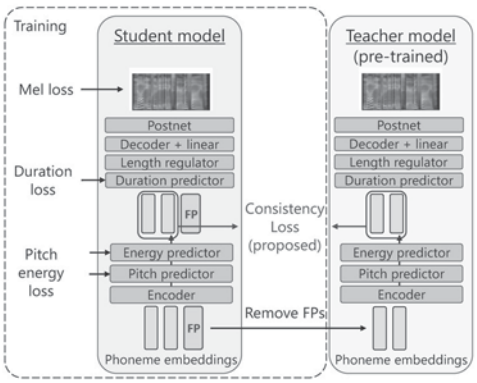


Fig. 1: Overview of consistency guarantee

3-8-6

3-8-6 隠れセミマルコフモデルに基づく構造化アテンションを用いた音声合成におけるパラメータ共有構造の検討

Parameter sharing structures in speech synthesis using structured attention based on a hidden semiMarkov model

☆石田龍成，藤本崇人，橋本佳，南角吉彦，徳田恵一 (名工大)

本研究では，隠れセミマルコフモデルに基づく構造化アテンション (HSMMアテンション) にパラメータ共有構造を導入し，エンコーダとデコーダに統計的生成モデルとしての一貫性を持たせることにより，学習の安定化を期待する．HSMMアテンションのパラメータ共有構造は共有するパラメータの種類や共有の度合いにより，いくつかのバリエーションがあるため，主観評価実験を行うことで，最適なパラメータ共有構造について検討する．実験結果より，学習の安定化を確認し，共有構造の有効性を示した．

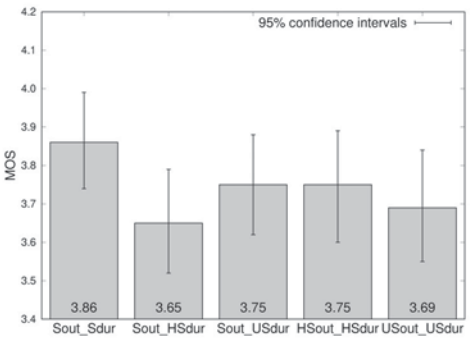


Fig. 1: Results of MOS test

3-8-7

3-8-7　日本語人対人会話データにおける発話タイミングのモデリング

Response Timing Modeling of Japanese Human-Human Dialog

☆佐久間仁（早大），藤江真也（千葉工大/早大），小林哲則（早大）

◆ **研究目的**

人間の発話タイミングの分布をモデル化し，音声対話システムに適用する

◆ **提案概要**

構文完全性を推定して得られた特徴量（SCP; Syntactic Completeness Projection）をタイミング推定に利用

- ストリーミング音声認識の結果から言語モデルにより後続の発話内容を生成
- M 文字先までに発話が終了する確率を SCP として算出
- 音響，言語，時間情報に加え，SCP をタイミング推定に利用

◆ **日本語対話データを用いたタイミング推定実験**：

- 許容する誤差を変えて算出した適合率，再現率, F1 を評価

Proposed　：従来の音響，言語，時間情報と SCP を用いた提案手法
w/o SCP　：提案モデルから SCP の利用を除いた手法
Fixed　：ターン交代のみ推定し，固定タイミングで発話する手法

◆ **結果(Fig.1)**：

- 提案法によりタイミング推定性能が向上することを確認

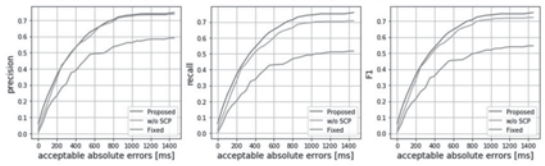


Fig.1 precision/recall/F1 curve

3-8-8

3-8-8　高効率対話方策学習のための規則知識を統合した深層 DYNA-Q

DNN-rule hybrid Dyna-Q for sample-efficient dialog policy learning

ZHANG Mingxin，〇篠崎 隆宏（東工大）

◆タスク達成型の対話エージェントを構成する上で，強化学習は柔軟な挙動を可能とする強力な手段である。しかし学習にエージェントによる多くの試行が必要となる。

◆エージェント内部で対話相手であるユーザーの行動をシミュレートし経験を拡張することで学習効率の改善を図る手法として， Deep Dyna-Q が提案されている。

◆本研究では対象タスクにおけるドメイン知識を規則としてユーザーモデルに統合することによる，Deep Dyna-Q の改良法を提案する。

◆映画チケット予約タスクにおいて，提案法が効率的な学習を実現することを示す。

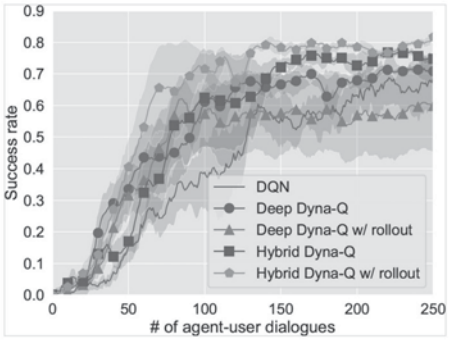


Fig.1 : Learning curves of agents.

3-8-9

3-8-9　傾聴対話におけるユーザ発話に同調する相槌の韻律制御

Prosody prediction of backchannels synchronizing with user's utterance in attentive listening dialogue

〇越智景子，井上昂治，△ディベッシュ・ラーラー，河原達也（京大）

◆**背景**　音声対話システムにおいて、効果的な相槌の制御は不可欠であり、ユーザの発話に肯定的に耳を傾ける傾聴対話においてとくに重要な役割を果たすと考えられる。これまで、聞き手の相槌とその直前の話し手の発話との基本周波数（F0）と F0、パワーとパワー同士などの間に相関がみられ、韻律の同調が報告されている。

◆**目的**　ユーザへの共感の表現、信頼感を感じるような相槌の生成の実現を目指し、ユーザの発話に韻律を同調させた相槌の韻律的特徴の制御を行うことを目的とする。
とくに、異なる種類の韻律同士(F0 とパワー)の関係も調査する。

◆**方法**　ユーザと、オペレータが音声出力・操作をしたアンドロイドの Wizard of Oz (WoZ)会話音声を用い、オペレータの相槌「うん」「はい」とそれに先行するユーザの発話の基本周波数(F0)とパワーの相関を調べた。また、先行するユーザ発話の韻律からオペレータの相槌を一般化線形モデル(Generalized linear model: GLM)とサポートベクター回帰(Support vector regression: SVR)により予測した。

◆**結果**　相槌の F0 の中央値の予測では、
先行発話の F0 の特徴量のみからの相槌の F0 の中央値予測よりも、
先行発話の F0・パワー両方からの予測のほうが良好な結果
⇒異なる種類の韻律的特徴が制御に関与していることが示唆。

3-8-10

3-8-10　クラウドソーシングを利用した感情演技発話マルチモーダルデータの収録と分析

Collection and evaluation of multimodal acted emotion data using crowdsourcing

☆ 堀井大輔，伊藤彰則，能勢隆（東北大）

◆本研究ではバランスよく自然な感情表現を含む発話映像を収集するために，話者に文脈付きの演技発話を読み上げてもらい，それを収録することを検討した.

◆現在，クラウドソーシングを活用し，映像の収集を行っており，本稿ではその概要や，収集した映像・音声に関する分析について報告する.

◆音声に関し自然性評価を行なったところ，既存の演技音声感情コーパスよりも本研究収録のものの方が自然なものであることが判明した.

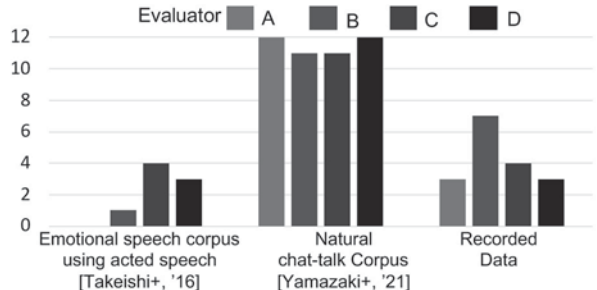


Fig. 1 : Number of utterances evaluated as “Part of a natural dialogue”

◆用意した発話文を収録者それぞれが自分なり改変して発話することを許可したのが自然性向上につながったと考えられる.

3-9-1

3-9-1　音声情報伝達の確実性向上のためのポップアウトボイスの特徴解明

Pop-out voice for reliability improvement of speech communication

○天野成昭(愛知淑徳大), 河原英紀(和歌山大), 牧勝弘(愛知淑徳大),
北村達也(甲南大), 山川仁子(尚絅大), 能田由希子(ATR)

ポップアウトボイスとは，背景雑音などの妨害音が存在する状況であっても非常に目立って知覚される音声である。この音声の特徴を解明し，音声情報伝達の確実性の向上に繋げることを大きな目的として，ポップアウトボイスの基礎データを得る実験を行った。バブルノイズを重畳した 94 話者の音声を 12 名の実験参加者にヘッドホンで呈示し，各話者の音声が背景雑音からポップアウトする程度を 5 段階で評定させた。その結果，評定値が幅広く分布すること(Fig. 1)，および男女の話者間の平均評定値に有意差が無いことが分かった。

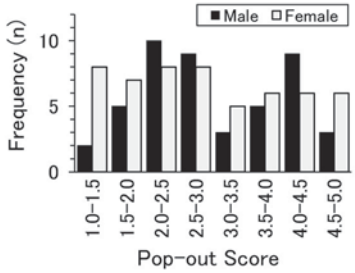


Fig. 1. Distribution of pop-out score for 100 speech items pronounced by 94 speakers (44 male and 52 female speakers).

3-9-2

3-9-2　母語がポップアウトボイスの検出に及ぼす影響

Effects of native language on detection of pop-out voice

○天野成昭(愛知淑徳大), 山川仁子(尚絅大), 河原英紀(和歌山大)

母語がポップアウトボイスの検出に及ぼす影響を明らかにすることを目的として，日本語を知らない英語母語話者 15 名に，ポップアウトの評定値ランクが高，中，低の日本語のポップアウトボイスを様々な SN 比において検出させる課題を行わせた。その結果，日本語母語話者に日本語のポップアウトボイスを検出させた場合(北原ら，2022)とほぼ同様の検出曲線(Fig. 1)および検出閾(Table 1)が得られた。これはポップアウトボイスが，母語の言語知識にはあまり依存せず，音響的特徴に大きく依存していることを意味している。

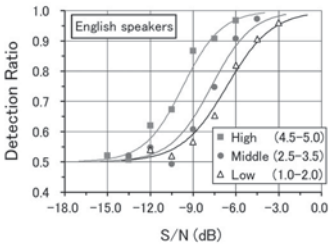


Fig. 1. Detection ratio of Japanese voice in babble noise by English native speakers (n = 15) as a function of S/N and pop-out score rank. Fitted logistic curves were also plotted.

Table 1. Detection threshold (dB) of pop-out voice by English and Japanese speakers.

Native language	Pop-out score rank		
	High	Middle	Low
English	−9.7	−7.7	−6.6
Japanese	−10.9	−8.8	−7.0
Difference	1.2	1.1	0.4

Note. Japanese data were obtaind from Kitahara et al. (2022).

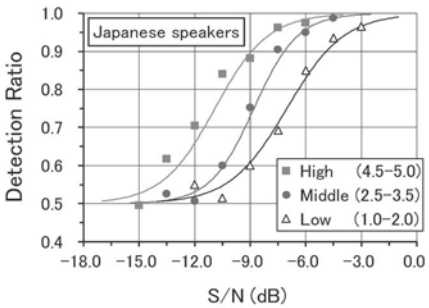
3-9-3

3-9-3　様々な SN 比におけるポップアウトボイスの検出

Detection of pop-out voice at various SN ratios

○北原真冬(上智大), 田嶋圭一(法政大), 米山聖子(大東文化大),
北村達也(甲南大), 河原英紀(和歌山大), 天野成昭(愛知淑徳大)

背景雑音などの妨害音が存在する状況であっても非常に目立って知覚される音声をポップアウトボイスと呼ぶ。天野ら(2022)の実験で得たポップアウト評定値のランクに基づき音声とバブルノイズを様々な SN 比で重畳して刺激を作成した。日本語母語話者 40 名にこれらの刺激と，ノイズのみの刺激の一対比較課題を行わせ，音声の検出率を測定した。その結果，ポップアウトのランク間で SN 比にして約 2dB ずつ異なる検出閾が得られた。これはポップアウトの評定値が高いほど音声が検出されやすいことを示している。



Detection ratio of Japanese voice in babble noise by Japanese native speakers (n = 40) as a function of S/N and pop-out score rank. Fitted logistic curves were also plotted.

3-9-4

3-9-4　バブルノイズ環境下におけるポップアウトボイスの周波数スペクトルパタン

Frequency spectrum pattern of pop-out voice in bubble noise

○牧 勝弘（愛知淑徳大), 坂野 秀樹（名城大), 河原 英紀（和歌山大),
天野 成昭（愛知淑徳大）

◆背景雑音などの妨害音下において目立って知覚される音声をポップアウトボイスと呼ぶ。

◆バブルノイズ環境下においてポップアウトする音声の音響的特徴、特に時間的に平均された周波数スペクトルに焦点をあて分析を行った。

◆その結果，ポップアウトする音声は、ポップアウトしない音声に比べて、男女共に 1 kHz 以上の振幅が大きいことが分かった（図 1）。

◆周波数スペクトルの 1－8 kHz の平均振幅値とポップアウトの評定値との間には、男女共に正の高い相関が見られた（男性 : r = 0.86）。

◆音声を時間的に 4 分割し、どの時間区間がポップアウトに寄与しているのかを調べた結果、どの時間区間も同程度の寄与率であった。

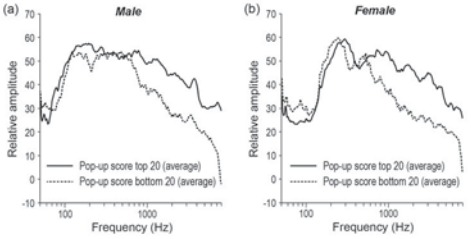


Fig.1 : Frequency spectrum of voices with low and high pop-out scores. (a) Male voice. (b) Female voice.

3-9-5

3-9-5　拡張音声モーフィングによるポップアウト属性の検証可能性

Possible verification of contributing attributes of ``pop-out'' voices using extended morphing tools

○河原英紀(和歌山大)，森勢将雅(明大)，榊原健一(北海道医療大)
北村達也(甲南大)，牧勝弘(愛知淑徳大)

◆駅や地下街などの喧騒の中でも目立って注意を引きつける声がある。そのような声をポップアウトボイスと呼び，多数の音声資料を用いて，知覚的属性，音響的属性・生成機構との関連などを調べてきた。

◆これまで STRAIGHT に基づいて開発してきた拡張モーフィングを，高品質 VOCODER である WORLD をベースとして新たに設計しなおし，支援ツールを MATLAB の新しい開発環境 Appdesigner を用いて実装し，オープンソースとして公開した。

◆この拡張されたモーフィングを用いた予備検討を行った。その結果，獲得された知見に基づいて導かれた音響的属性操作による知覚的な効果を，拡張されたモーフィングを用いて検証する可能性が示された。

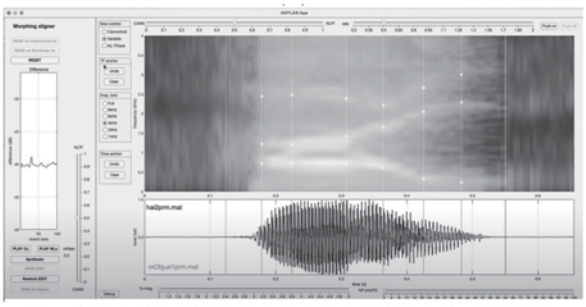


Fig.1：GUI interface of morphing assisting tool for time-frequency anchor assignment

3-9-6

3-9-6　声道断面積関数の抽出手法の改良とポップアウトボイスへの適用

Improvement of the extraction method of vocal tract area function and its application to pop-out voice

☆多湖崇宏，奥田康弘(名城大院)，旭健作，坂野秀樹(名城大)

◆声道断面積関数の抽出手法の改良

- 分析パラメータ（低域・高域の線形予測次数，帯域分割周波数）
- 低域・高域スペクトルの平坦化

◆改良した手法をポップアウトボイスへ適用し，**ポップアウトの程度** と **推定した声道形状の動的・静的特徴** との関連性を分析（Fig.1 にその時の散布図を示す）

- 動的特徴
  - ✧ 相関係数：男性 0.171，女性 0.152（相関なし）
  - ✧ 今回の指標では両者に相関を確認できなかった
- 静的特徴
  - ✧ 相関係数：男性 0.414，女性 0.524（相関あり）
  - ✧ 喉頭の面積に比べて喉頭より上の面積が大きくなる傾向あり

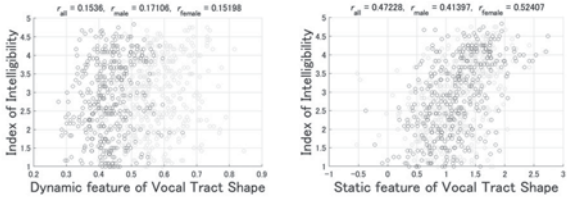


Fig.1：Scatter plots between degree of pop-out and dynamic and static features (Cyan：Male speech, Magenta：Female speech)

3-10-1

3-10-1　多言語同時拡声による案内放送の了解度（1）言語数の影響

Intelligibility of speech guidance simultaneously presented in multiple languages: (1) The effect of the number of languages.

○森本政之(神戸大)，佐藤逸人，穴水智也(神戸大院・工学研)，佐藤洋(産総研)

◆多言語同時拡声による案内放送の実現可能性を検討するために，同時に再生する言語数と目的音声の時間的位置をパラメータとした単語了解度試験を行った。

◆聴取者の正面に位置した 1 つのスピーカから，5 つの連続した単語を複数の言語（日本語，中国語，英語，韓国語）で同時に提示し，指定した時間的位置（1，3，5 番目）の日本語単語を書き取らせた。

◆2 言語および 3 言語を同時に提示した場合は，時間的位置の影響は有意ではなく，単語了解度の平均値はそれぞれ 91.3%と 82.9%であった。

◆4 言語を同時に提示した場合は，時間的位置の影響は有意であり，位置が後になるほど単語了解度は上昇した。時間的位置が 5 番目の単語了解度は，妨害音が定常雑音で同じ SN 比の条件(Amano *et al.*, 2009)に近い値まで上昇した。

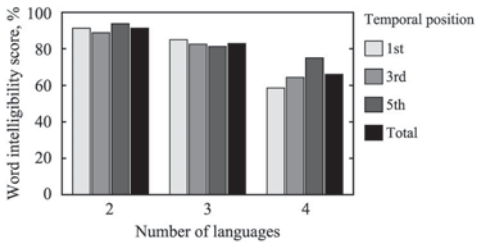


Fig.1：Word intelligibility scores for the single loudspeaker system as a function of the number of languages.

3-10-2

3-10-2　多言語同時拡声による案内放送の了解度（2）目的単語の時間的位置と空間的位置の影響

Intelligibility of speech guidance simultaneously presented in multiple languages: (2) The effect of the temporal and spatial position of the target speech.

○佐藤逸人，穴水智也(神戸大院・工学研)，森本政之(神戸大)，佐藤洋(産総研)

◆3-10-1 と同様の実験を，空間的に離れた複数のスピーカからそれぞれの言語を提示するシステムで行った。

◆3 言語同時提示の場合，スピーカ数が 1（3-1）と 3（3-3）の両者で時間的位置の影響は有意ではなかった。3-1 の単語了解度（83%）との差は有意ではないが，3-3 の単語了解度は 90%程度まで上昇した。

◆4 言語同時提示の場合，スピーカ数が 1（4-1），2（4-2），4（4-4）のいずれにおいても時間的位置の影響は有意であった。時間的位置が 1 番目と 3 番目の単語了解度は 4-1 よりも 4-2 あるいは 4-4 の方が有意に大きく，時間的位置が 5 番目の単語了解度は 4-2 と 4-4 の両者で 85%程度まで上昇した。

◆以上より，時間的位置と空間的位置を最適化すれば，4 言語同時提示の場合でも実用可能性のある単語了解度が得られると考えられる。

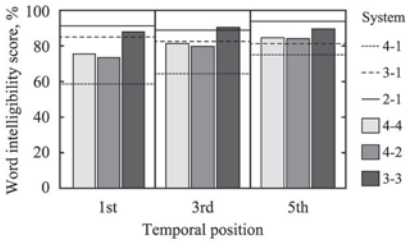


Fig.1：Word intelligibility scores averaged within the same system for each temporal position of the target word.

3-10-3

3-10-3　健聴者における単音節明瞭度および文了解度閾値の比較

Comparison of monosyllabic intelligibility and intelligibility thresholds of the sentence with normal hearing

◎岡龍也(リオン)，中市健志(リオン)，△山岡香央(ファンケル)，△坂井友美(ファンケル)，△岡本康秀(済生会中央病院)

◆本測定では健聴被験者 70 名を対象に静寂下および雑音負荷条件における 67-S 語音聴力検査の結果と雑音下における文了解度閾値検査である J-HINT，J-Matrix Test を測定しその結果を比較した．

◆静寂下および雑音下における 67-S 語音聴力検査の結果を Fig.1，J-HINT，J-Matrix Test の結果を Fig.2 に示す．

◆67-S 語音聴力検査では，雑音信号を負荷することで天井効果の影響を受けずに被験者間の個人差を抽出することができた．

◆J-HINT と J-Matrix Test の両検査は同様の傾向を示した．J-HINT と比較し J-Matrix Test の値の方が低い結果となった．

◆雑音負荷による単音節明瞭度と文了解度閾値間には相関がみられず，各検査より得られる結果は異なるものであることが示唆された．

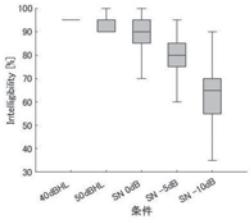


Fig.1：Result of 67-S speech audiometry

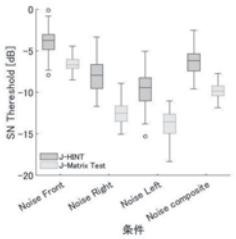


Fig.2：Result of J-HINT and J-Matrix Test

3-10-4

3-10-4　臨床現場で用いられる語音検査用音源の単音節レベルに関する検討

A study for monosyllabic levels of testing used in clinical practice

○中市健志(リオン)，△和佐野浩一郎(東海大学医学部耳鼻咽喉科)

◆本報告では，耳鼻科臨床現場や聴覚心理実験にて用いられる９つの単音節音源について校正音に対する単音節のレベルを調査した。

◆音源は 57-S 語表 (1982)，補聴器適合評価用 CD:TY-89 (1990)，単音節語明瞭度検査音 TS-1 VU45100 (1992)，補聴器適合検査用音源 KR2000A (2000)，親密度別単語了解度試験用音声データセット FW03 単音節音源 (2003)，語音聴取評価検査 CI-2004(試案) 単音節検査 (2004)，以上女性話者５名，男性話者４名を用いた。

◆各音源を PC に取り込み，各単音節の VU 値，実効値，等価騒音レベル，ピーク値を求めてそれぞれ校正音のレベルとの比を算出した。子音のレベルは子音部と母音部を目視にて切り出し。母音部の実効値を基準として算出した。

◆各音源における相対レベルの基準はそれぞれである事，母音に対する子音部のレベルは，有声音，無声音や構音様式別の傾向はみられない事がわかった。

Table 1 Relative levels for each test signals as referenced calibration sound

	dBvu		rms dB		$L_{Aeq,T}$ dB		Peak dB	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
FW03:fhi	3.7	3.4	4.0	2.7	**0.2**	**0.3**	16.7	1.7
FW03:flo	3.2	2.5	3.3	2.1	**0.1**	**0.3**	15.1	1.4
FW03:mis	1.0	2.5	2.8	1.8	**0.2**	**0.3**	17.5	1.5
FW03:mya	0.8	2.5	2.8	2.0	**0.2**	**0.2**	17.6	1.4
57-S	0.8	1.9	1.6	1.8	-0.8	2.8	**11.2**	**1.0**
CI-2004	**-1.6**	**1.0**	-0.9	1.3	-4.1	2.7	13.6	2.1
TY-89	**-10.3**	**1.7**	-9.6	2.0	-13.0	3.4	5.6	2.6
KR-2000A	**-1.0**	**0.4**	-1.5	0.6	-5.5	2.7	9.8	1.4
TS-1 VU	**0.0**	**0.0**	2.0	1.2	-0.9	3.0	14.6	2.3

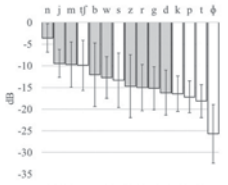


Figure 1 Relative levels of consonant sounds as referenced vowel sounds

3-10-5

3-10-5　話者年齢の人と機械学習による推定の傾向の分析

Trends analysis in speaker age estimation between human and deep neural network model

○北岸佑樹，俵直弘，小川厚徳，井島勇祐，増村亮(NTT)

◆DNN による自動話者年齢推定はコールセンタでのオペレータによる顧客の話者年齢推定の代替といった利用法が期待される

◆実利用化を想定し，**模擬コールセンタ通話音声に対してアノテータ**（オペレータ経験者 21 名）**と DNN による年齢推定実験で両者の違いを調査**

- アノテータ：10 代前半以下～90 歳以上の 5 歳単位の年齢クラスを推定
- DNN：1 歳単位で年齢を自動推定し，5 歳単位の年齢クラスに変換
- 年齢クラスに若い順に番号を振り，真の年齢クラスとアノテータ・DNN による推定クラスの絶対誤差の平均値 c-MAE と相関係数 ρ を Fig. 1 に示す（c-MAE 1 ポイントは平均±約 5 歳の推定誤差を意味する）

◆**DNN による自動推定はオペレータ 21 名と同等以上の推定精度**

◆**以下の傾向がアノテータによる推定結果の平均値だけに見られた**

- 中年の，特に女性話者の年齢を実際より若く推定する
- 男性より女性話者の音声を大きく誤って推定する

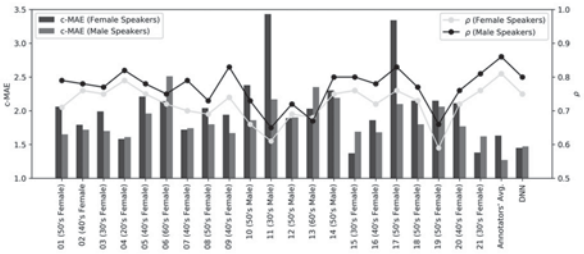


Fig.1：Results of Speake Age Estimation by Annotators and DNN

3-10-6

3-10-6　宣伝音声から生起される感情及び音声印象を考慮した購買意欲モデルの提案とその評価

Proposal and evaluation of the consumer behavior model via perceived emotion and voice quality from advertising speech

◎長野瑞生，井島勇祐，廣谷定男(NTT)

◆感情を媒介とする消費者行動モデル

- 音声が購買意欲に及ぼす影響を音声→感情→購買意欲のプロセスで説明することが出来る [長野他，音講論(春), 2022]
- しかし感情のみを媒介とするだけでは不十分
- 音声印象を媒介に加えることで説明力が向上する可能性がある

◆本研究では，感情を媒介とする行動モデルを用いて，音声特徴が購買意欲に及ぼす影響における音声印象の媒介効果（説明力）を検証

- 平均 F0，話速，F0 分散を変更した宣伝音声を聴取し、自身の感情、音声印象、購買意欲を回答する主観評価実験を実施
- 各仮説モデルにおける感情及び音声印象の媒介効果を比較

◆感情と音声印象を並列に媒介させるモデル(Parallel モデル)が最も媒介効果が高い

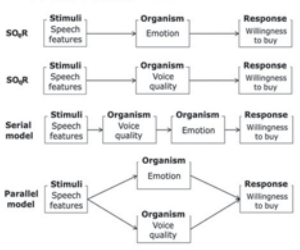


Fig.1：Hypothesized models.

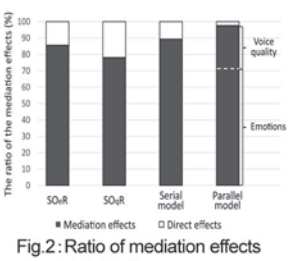


Fig.2：Ratio of mediation effects over total effect.

3-10-7

3-10-7　緊迫感知覚に寄与する変調周波数帯域の検討

Study on contribution of modulation frequency bands on the perception of urgency

○木谷俊介，劉小婷，郭太陽，磯山拓都（北陸先端大），
李軍鋒（中国科学院），赤木正人，鵜木祐史（北陸先端大）

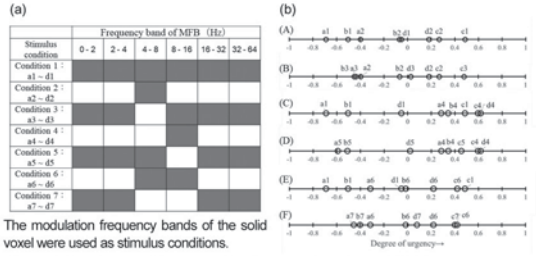
【背景】　音声コミュニケーションにおいて，非言語情報は重要である。我々のこれまでの研究から，音声の振幅包絡が緊迫感知覚に影響を与えていることが分かっている。しかし，緊迫感知覚に重要な振幅包絡線の変調周波数帯域は明らかになっていない。

【目的】　緊迫感知覚に寄与する変調周波数帯域を明らかにすること。

【手法】

1. 変調フィルタバンクを用いた雑音駆動音声の合成系を構築した。
2. 振幅包絡線の変調周波数を制御した雑音駆動音声 (Fig. 1(a)) を用いた聴取実験を行った。
3. 緊迫感知覚の心理スケールを算出した (Fig. 1(b))。

【結果】　4～16 Hz の変調周波数帯域が緊迫感知覚に重要であることがわかった。



The modulation frequency bands of the solid voxel were used as stimulus conditions.

Fig.1: (a) Stimulus conditions used in the experiment, and (b) Urgency scales under stimulus conditions.

3-10-8

3-10-8　異なる感情を想起する音声の組み合わせによる可笑しさの創出に関する一検討

Preliminary study on creating funniness by combination of voices evoking different emotions

☆渡邊悠希，坂本修一（東北大通研/院情科研），
星貴之，長谷芳樹（PxDT），中野学（PxDT，東北大）

◆ 本研究では，意外性を含む文章の前後半に異なる感情が想起されるように音響特徴量を操作し，生成された音声から想起される可笑しさの程度を評価した．

◆ 操作する音響特徴量は，癒やしの想起に関連すると示唆された時間長と F0 の変化幅とし，表 1 に示すように，音声の前後半にそれぞれ変換処理を施した．

◆ 図 1 に示すように，「平静」→「癒やし」に想起感情が変化する音声が概ね可笑しさを想起する程度が大きいという結果が得られた．

表 1 生成音声に適用した変換処理の内訳

前後半で設定した感情の変化（前半 － 後半：記号は図1と対応）		前後半で設定した音響特徴量の設定値（時間長，F0変化幅）	
		前半	後半
○	平静 － 平静	(1.0倍，1.0倍)	(1.0倍，1.0倍)
▽	癒やし(Δ1倍) － 平静	(1.1倍，1.2倍)	(1.0倍，1.0倍)
▼	癒やし(Δ2倍) － 平静	(1.2倍，1.4倍)	(1.0倍，1.0倍)
△	平静 － 癒やし(Δ1倍)	(1.0倍，1.0倍)	(1.1倍，1.2倍)
▲	平静 － 癒やし(Δ2倍)	(1.0倍，1.0倍)	(1.2倍，1.4倍)

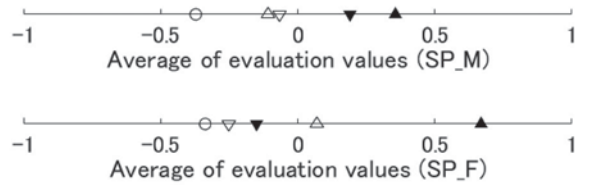


図 1　2名の話者（SP_M，SP_F）の発話に変換処理を行った音声に対する可笑しさの平均評価値

3-11-1

3-11-1　テキストマイニングによる自動車交通騒音に対する意見の分析

Analysis of opinions on road traffic noise using text mining

☆古味由惟（神奈川大・院），横島潤紀（神奈川大・工/神奈川県環境科学セ），
辻村壮平（茨城大院・理工研），山内勝也（九州大・芸工），山崎徹（神奈川大・工）

本報では，自由記述回答を分析することにより，自動車交通騒音に対する意見の把握および，既報研究の知見との比較を行った．抽出語に対して，単語の共起性や性別・住宅種別に関して分析を行った．図 1 に性別および住宅種別の対応分析を示す．図中の抽出語，抽出語の出現回数および語の位置関係から性別・住宅種別における特徴を読み解く．分析結果より，性別および住宅種別おける自動車交通騒音に対する意識の違いがあることが分かった．

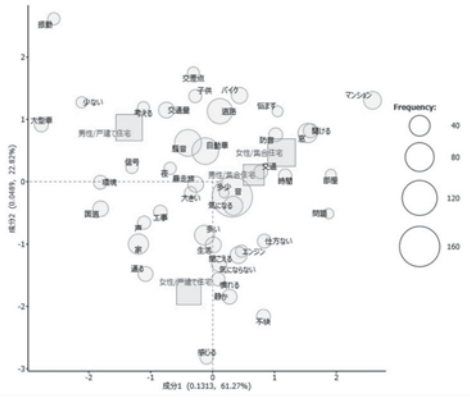


Fig.1 : Correspondence analysis of gender and housing type in open-ended questions

3-11-2

3-11-2　交通騒音による睡眠影響　その 5

Sleep effects on transportation noise Part 5

☆高橋達樹，高瀬ゆり，上田麻理(神奈川工科大)，廣江正明（小林理研）

◆ 睡眠時の交通騒音は，心理的影響や睡眠妨害だけでなく，健康影響も懸念されており，質の良い睡眠環境の確保が課題となっている．

◆ 本研究では，睡眠習慣の中でも，眠りの深さと寝つきをグループ分けして，自記式質問紙による主観的睡眠評価とアクチグラフ(体動)を用いた客観的な睡眠評価で，それぞれの群間の睡眠の差を統計的に分析した．

◆ 眠りの深さの群間では，主観的睡眠評価と客観的な睡眠評価で同様な睡眠影響の傾向が確認された．

◆ 眠りの深さの浅い群を対象に，呈示レベル $L_{Aeq,1min}$ と中途覚醒発生率(WASO)の関係の結果を示す(Fig.1)．

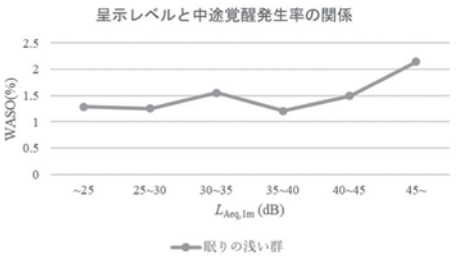


Fig .1　呈示レベルと中途覚醒発生率の関係（眠りの浅い群）

3-11-3

3-11-3　鉄道高架下保育施設における電車通過時の床及び柱の振動加速度レベルに関する実験的検討

Study of vibration acceleration level of floor and pillar of nursery schools located under the elevated railway

◎岡庭拓也，冨田隆太（日大・理工）

◆鉄道高架下保育施設の音・振動の評価・対策方法の基礎的な検討として，床及び柱の振動加速度レベルに関する実験的検討を行った．

◆橋脚側面(高さ 1 m～1.5 m)と橋脚から 70 cm 程度離れた床面(近傍点)及び近傍点から 1 m 離れた床面で，電車通過時の振動加速度レベルを測定した．

◆本報の調査の範囲では，振動加速度レベルの差に着目すると，橋脚と橋脚近傍点の床面の振動加速度レベルの差が大きいほど調査員の評価が良くなる傾向が見られた．

◆床面の振動と音を比較検討したところ，63 Hz 帯域では音と振動がある程度連動している可能性が示唆された．

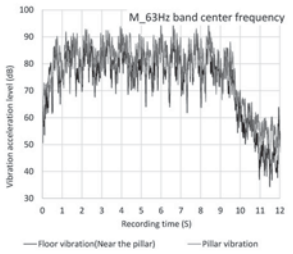


Fig.1　Vibration acceleration level of nursery school M (63Hz band center frequency)

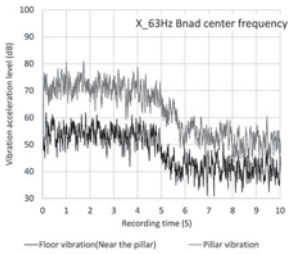


Fig.2　Vibration acceleration level of nursery school X (63Hz band center frequency)

3-11-4

3-11-4　航空機騒音暴露の日変動を考慮した長期間評価に関する検討

－ 日ごとの L_{den} の頻度分布と暴露反応関係 －

Long-term evaluation considering daily fluctuations in aircraft noise exposure - frequency distribution of daily aircraft noise exposure L_{den} and exposure-response relationships

○牧野康一（小林理研）、森長 誠（神奈川大・建築）

◆防衛施設飛行場周辺で行った社会調査で得られたデータを再整理し、当時の騒音モニタリングデータを用いて日々の L_{den} 測定値の統計量に着目して検討を行った。

◆日々の L_{den} 測定値の 80%レンジの上端値（$L_{den,10}$）を飛行場毎に推定し、暴露反応関係を再分析したところ、民間空港の暴露反応関係に近づく傾向がみられた。

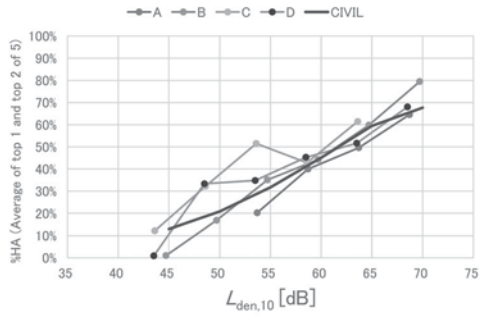


Fig. 飛行場別の L_{10} と%HA の暴露反応関係
(5 段階評価の上位 2 カテゴリと上位 1 カテゴリの平均値)

3-11-5

3-11-5　放射指向性を考慮したロボット掃除機の運転音評価方法の検討

Evaluation method of operating noise of robot vacuum cleaner considering radiation directivity

◎山内源太（日立）

◆ロボット掃除機の運転音は，従来直上 1 点で評価される。しかし，この測定点では側方の放射音を正しく評価できていない可能性がある

◆3 社 7 機種のロボット掃除機を対象に運転音の放射指向性を測定した結果，低周波数は直上方向のレベルが最大となる概ね全指向性に近い指向性，高周波数は側方のレベルが最大となる鋭い指向性であった

◆測定点は直上のみならず側方も必要であることから，評価量として音響パワーレベルを採用し残響室法により測定した結果，レベルが高い側方の影響が考慮された音源特性が得られた

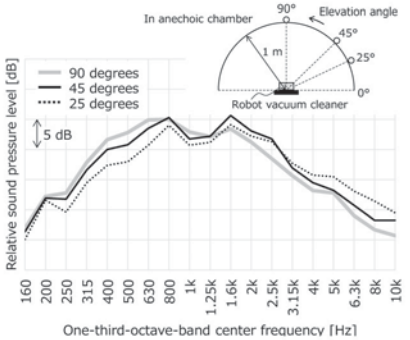


Fig.1 : Frequency characteristics at three elevation angles 1 m away from a robot vacuum cleaner.

3-11-6

3-11-6　ダミーヘッドを用いた VHF 領域の耳介周りの音響計測(その 4)音圧計測の誤差要因と頭部形状の違いの検討

Acoustic measurement around the pinna in the VHF region using a dummy head –Part4, factors that cause error of measurement, and influence on difference in head shape

☆春澤恒輝(神奈川工科大)，稲村祐美(エスピック)，廣江正明(小林理研)，中村健太郎(東工大)，長谷川英之(富山大)，神崎晶(慶應大)，桐生昭吾(東京都市大)，上田麻理(神奈川工科大)

◆ スピーカ呈示による可聴閾値計測では，音圧校正などが難しく，特に VHF 音の可聴閾値計測では，波長が短いため音が回折しづらいことから，頭の位置や向きのわずかな変化で聴取音圧が変化することがある．

➢ より正確な可聴閾値計測を行うには，頭の位置や向きによる聴取音圧の変化，誤差を把握する必要がある．そこで本研究では TSP 信号によるインパルス応答測定で，測定誤差について検討した後，頭部形状が HRTF の違いに及ぼす影響について検討した．

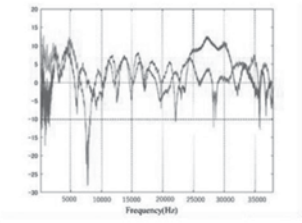


Fig.　HRTF of HATS measured in this study.

3-11-7

3-11-7　ダミーヘッド(HATS)を用いた VHF 領域の耳介周りの音響計測(その 5) 毛髪による VHF 音の音響的影響*

Acoustic measurement around the pinna in the VHF region using a dummy head –Part5, Acoustic influence of VHF sound by hair

☆志賀康祐，春澤恒輝(神奈川工科大)，稲村祐美(エスピック)，
廣江正明(小林理研)，長谷川英之(富山大)，中村健太郎(東京工業大)，
神崎晶(東京医療センター)，桐生昭吾(東京都市大)，上田麻理(神奈川工科大)

◆20kHz を超える非常に高い周波数の音は，どのような特性や，身体への影響があるのかを示す研究は少ないため，先行研究では、ＴＳＰ信号を用いたインパルス応答計測を用いて多角的かつ広帯域の周波数を調査した。

◆ＶＨＦ音は減衰しやすいため物体の影響を大きく受ける。本研究では、毛髪が HRTF に与える影響を調査した。

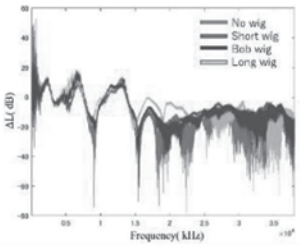


Fig. Comparison of four types of wigs at 30°

3-11-8

3-11-8　ダミーヘッド(HATS)を用いた VHF 領域の耳介周りの音響計測(その 6)－単純な耳モデルによる実験的検討とシミュレーション

Acoustic measurement around the pinna in the VHF region using a dummy head –Part6, Experimental examination and simulation with a simple ear model.

☆大石まなか，春澤恒輝(神奈川工科大)，廣江正明(小林理研)，
長谷川秀之(富山大)，中村健太郎(東京工業大)，神崎晶(東京医療センター)，
桐生昭吾(東京都市大)，上田麻理(神奈川工科大)

◆VHF 音の聴こえは聴覚器の大きさや形状等の影響を受ける．多数の聴覚器を検討するため，耳介形状の影響を数値解析で検討した．

◆本研究では成人と子供の外耳道の大きさ・形状について検討した．

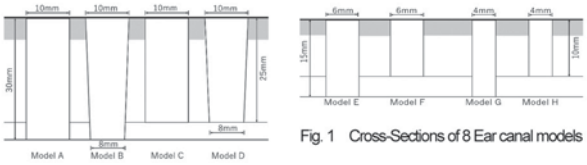


Fig. 1　Cross-Sections of 8 Ear canal models

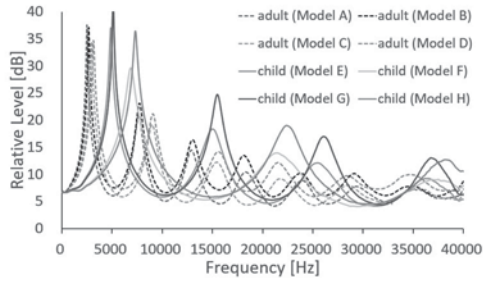


Fig.2 Frequency characteristics of relative levels at eardrum by reference to SPL in free field.

3-11-9

3-11-9　アジア地域における職場騒音の許容暴露制限

Permissible exposure level of occupational noise in Asia-Pacific region

○横山 栄，小林知尋（小林理研）

本研究では、アジア・オセアニア地域における労働安全衛生に関する法令について、インターネット検索によって調査した。対象とした 27 か国のうち 20 か国で法令が整備されていた。許容暴露騒音レベル（PEL）については、14 か国で 85 dB、6 か国で 90 dB、暴露時間と許容レベルの換算レート（ER）は 12 か国で 3 dB、7 か国で 5 dB を採用していた。暴露レベルが PEL を上回る場合には、聴覚保護具の着用や特殊健康診断の受診を義務付けている国も多くみられた。さらに、衝撃音についても 14 か国で許容レベルが規定されていたが、その測定方法は統一されていなかった。アジア地域における法令の概要は欧米諸国と同様であった。

Table　Permissible exposure levels and exchange rates for noise exposure.

Country	PEL [dB]	ER [dB]	Country	PEL [dB]	ER [dB]
China	85	3	Indonesia	85	3
Hong Kong	85	3	Laos	85	3
Japan	85	3	Malaysia	85	3
Korea	90	5	Philippines	90	5
Taiwan	90	5	Singapore	85	3
Bhutan	90	5	Thailand	90	5
India	85	5	Vietnam	85	3
Nepal	90	5	Australia	85	3
Sri Lanka	85	3	New Zealand	85	3
Cambodia	85	—	Fiji	85	3

3-12-1

3-12-1　反射音グループ化を用いたスパース等価音源法による初期インパルス応答推定

Estimation of multiple room impulse responses by sparse equivalent source method based on grouped image sources.

☆松橋遼，鈴木薫佳，津國和泉，池田雄介（東京電機大）

◆背景
- 多点の室内インパルス応答（RIR）推定手法が提案されている。
- 虚像法を用いた等価音源法は反射音次数が増加すると推定精度が低下する課題がある。

◆提案手法
- 反射音到来時間に基づき，反射音を時間領域でグループ化，分解
- スパース等価音源法でグループごとにモデル化

◆実験
- 2.5 次元条件での実測実験によって推定精度を評価
- 従来手法と比較し SNR 約 3dB 精度が向上，推定可能領域が拡大

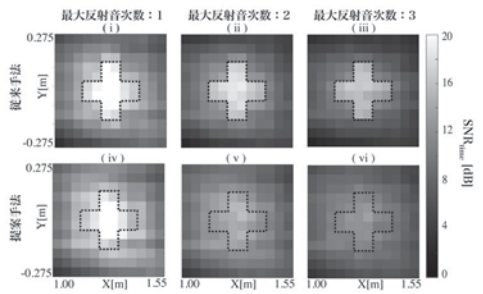


Fig.1: SNR distribution in time domain.

(The area enclosed by dotted line indicates microphone position)

3-12-2

3-12-2　音源の指向性と位置誤差を考慮した等価音源を用いた初期室内インパルス応答のモデル化

Modeling of early room impulse responses using equivalent source method considering errors of source directivity and position

◎津國和泉，松橋遼，鈴木佳薫，池田雄介（東京電機大）

◆問題点
- 室内インパルス応答の多点測定は困難
- 従来手法において音源の指向性によってモデル化精度が低下

◆目的
- 指向性を持つ音源からの室内インパルス応答のモデル化精度向上

◆提案手法
- スパース等価音源法を用いて音源の放射特性を事前にモデル化
- 事前の音源モデルを利用して室内インパルス応答をモデル化

◆シミュレーション実験
- 単一指向性を持つ音源からの一次反射音場をモデル化
- 提案手法は従来手法より約５dB の改善

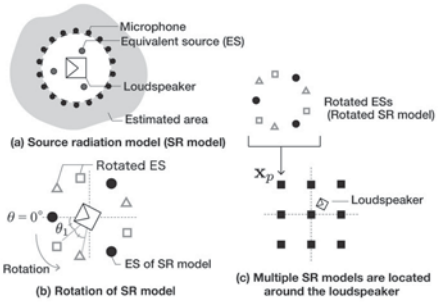


Fig.1 : The diagram of the proposed method

3-12-3

3-12-3　２時刻のマイクアレイ信号に基づくビームフォーミング

Beamforming Based on Microphone Array Signals of Two Time Snapshots.

◎岩見貴弘，尾本章（九大・芸工）

◆我々は既報で，帯域制限空間の再生核による音場表現に基づく瞬時アレイ入力を用いた軸ビームフォーミングを提案している。本研究では，この手法の欠点である前後同一視を，2 時刻のアレイ入力 (時刻情報) を用いることで解消した新たなビームフォーミングを提案する。

◆広帯域信号に対して適用可能であり，周波数領域手法のように狭帯域の結果を算出して統合する必要がないため計算コストが小さく，任意の次元及び素子配置で実行可能である。

◆遅延和ビームフォーマと比較して，計算量が小さくメインローブが鋭くサイドローブが良く抑制されていることを確認した。

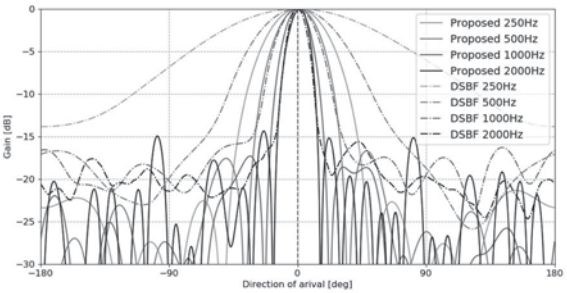


Fig.1　Beam patterns of proposed method and delay-and-sum beamformer for each frequency band.

3-12-4

3-12-4　マイクロホンアレイを用いた in-situ 吸音率計測法の評価

Evaluation of in-situ sound absorption coefficient measurement method using microphone array

○小柳慎一郎（竹中技研）

◆マイクロホンアレイで得られるクロススペクトルを利用した吸音率の in-situ 計測手法を提案している。本手法は指向性を使って入射波と反射波を分ける既存のアイデアに加え、エバネッセント波などの材料表面近傍に存在しうる吸音率と関係のない音場をキャンセルする機構を有している。時間領域での波形分割の必要がない一方で、係数が過大となることから SN 比の確保に問題があった。

◆本報告では係数の大きさの指標（ゲイン）を利用し、設計方程式で利用する正則化パラメータとゲイン、さらに設計誤差の関係を調べた。

◆正則化パラメータを動かした時のゲインと誤差のグラフにおいてトレードオフ関係を定め、折れ曲がり点から最適な正則化パラメータ σ を決定できることを示した。

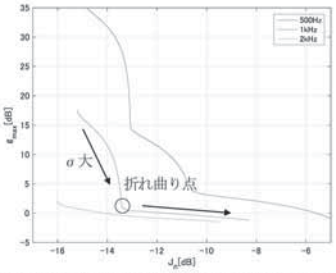


Fig.1 : Relationship between gain and error function for regularization parameter σ of Tikhonov regularization

3-12-5

3-12-5　一般的なスピーカを用いた全指向性音源構成の試み

An attempt of constructing omnidirectional sound source by fitting spherical harmonics coefficients.

☆加藤真木，岩見貴弘，尾本章（九大・芸工）

◆従来，室内インパルス応答測定において 12 面体スピーカアレイが使用されてきたが，12 面体スピーカアレイは高周波数帯域において全指向性を維持できていない。

◆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。

◆12 面体スピーカアレイの指向性測定を行い，指向性合成した音源との比較を行った。

◆全指向性との空間相関を取ることで定量評価を行い，指向性合成によって全指向性の模擬ができていることを確認した。

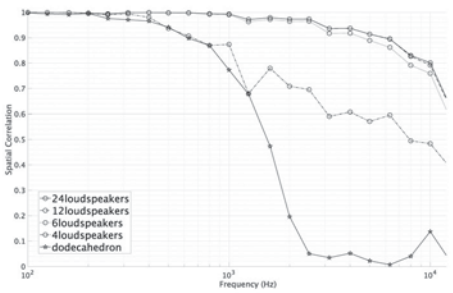


Fig.1　Spatial correlation with omnidirectioinal source.

3-12-6

3-12-6　室内音響測定における相反定理の検証

Investigation of room acoustic measurements based on reciprocity principle.

○後藤耕輔，山田祐生(竹中技研)

◆本報では室内音響測定に相反定理を適用するための嚆矢として，基礎的な実験による検討を実施したので，その結果を報告する。

◆測定は Fig. 1 で示す会議室で実施した。S1 地点に音源，P1 および P2 地点に受音点を設置した条件を直接法として，音源と受音点の位置を入れ替えた条件を相反法として，測定を行った。

◆室内インパルス応答の時間波形を評価することで，周波数やスピーカとマイクロホンの指向特性が与える影響を確認した。

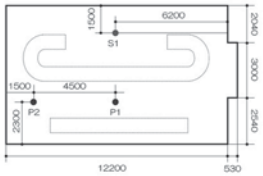


Fig. 1 Measurement arrangement.

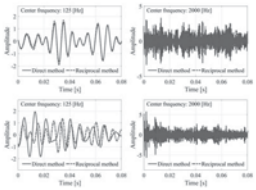


Fig. 2 (top) Waveforms by NL-52.

(bottom) Waveforms by MKE600.

Table 1 Normalization cross-correlation. (P1)

Frequency [Hz]	NL-52		MKE600	
	0-80 ms	80-500 ms	0-80 ms	80-500 ms
125	0.99	0.99	0.40	0.51
500	0.82	0.86	0.51	0.44
2000	0.29	0.27	0.42	0.26

3-12-7

3-12-7　反射音到来方向を考慮したステージ音響評価に関する研究　その1 −方向別 ST の計測方法−

Stage acoustic evaluation considering arrival direction of reflection sound, Part 1 - Measurement method of directional ST

☆板垣大稀(東大・工)，橋本悌(ヤマハ)，佐久間哲哉(東大・工)

◆筆者らは、反射音の到来方向を考慮した舞台音場評価のため、1 次のアンビソニックスマイクを用いた方向別インパルス応答計測に基づく ST の測定手法を構築した。

◆本報では、二種類のスピーカを用いた測定システムを用いて ST の基準エネルギーが十分な再現性を持つことを確認した。また、反射音エネルギーの時間窓の妥当性について検討し、方向別 ST の測定システムの検証を行った。

◆Fig.1 に音響測定を行なった測定位置を、Fig.2 に測定位置 4 における水平面内の $ST_{Early,dir}$ を示す。壁面付近での ST_{Early} は初期反射音の開始時刻によって値が大きく変動し、方向別で見ると壁面側の $ST_{Early,Left}$ で 4 dB 程度の差が見られた。

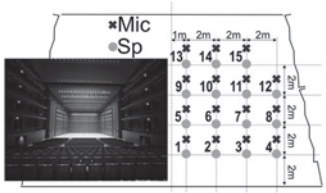


Fig.1: Measurement points

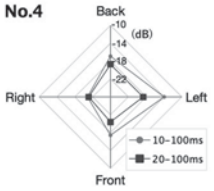


Fig.2: Directional ST_{Early} measured at point no.4

3-12-8

3-12-8　反射音到来方向を考慮したステージ音響評価に関する研究 その 2 − 各種ホールの測定事例 −

Stage acoustic evaluation considering arrival direction of reflection sound, Part2 - Measurement examples of various venues -

◎橋本悌(ヤマハ)、板垣大稀、佐久間哲哉(東大・工)

◆筆者らは、反射音の到来方向を考慮した舞台音場の評価のため、方向別インパルス応答計測に基づく ST の測定手法を構築した。1 次のアンビソニックスマイクを用いて、直交 6 方向で ST_{Early} および ST_{Late} を算出した。

◆本報では、室形状・空間規模の異なる 7 つのホールの多数点で方向別 ST を測定し、その結果について考察を行なった。Fig.1 に音響測定を実施したホールの内観を示す。

◆$ST_{Ealry,dir}$ は舞台の大きさやホールの音響特性によって値が変化した。また、舞台上の測定位置による方向別の偏差も確認された。$ST_{Late,dir}$ は方向別や舞台上の測定位置による偏差も小さい傾向にあり、平均吸音率が低く、残響時間が長い空間で値が高い傾向にあった。

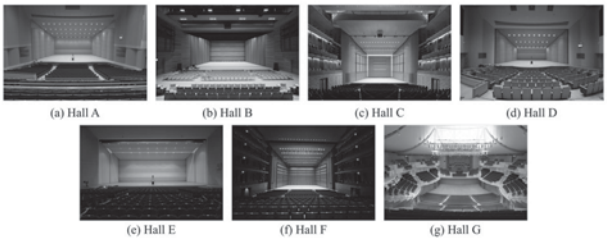


Fig.1 : Outline of halls where the acoustic measurement were performed

3-12-9

3-12-9　等価音源法と機械学習を用いた少数計測による壁面吸音率の推定

Estimation of sound absorption coefficients by small number of measurements based on equivalent source method and machine learning

☆大川祐貴子，松橋遼，津國和泉，池田雄介（東京電機大），
及川靖広（早大理工）

◆目的：
機械学習を用いた吸音率推定における学習用の計測データの削減

◆提案手法：
- 等価音源法に基づく擬似マイクロホンの生成による計測点の増加
- グリッド状の計測データから MLP と CNN を用いて吸音率を推定

◆結果：
- CNN を用いた推定が最も推定精度が高かった。
- CNN を用いた場合，擬似マイクロホン信号によって若干の推定精度改善が見られた。

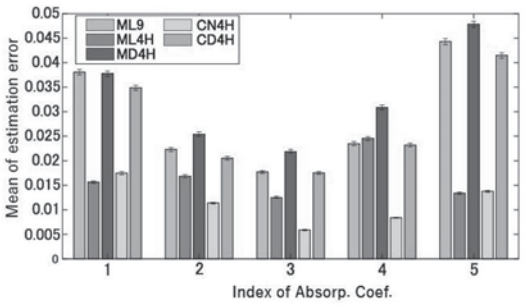


Fig. 1　Means of estimation error for absorption coefficients using MLP and CNN

3-12-10

3-12-10 建設現場環境音を用いた作業管理システムの開発 －音源定位と深層学習による作業推定技術の適用－

Work-monitoring in construction site using ambient sounds based on sound localization and deep learning

○諸橋俊大（鹿島建設），井本桂右（同志社大学），吉田康仁，岡尚人（鹿島建設）

◆音源定位と深層学習による作業推定技術により建築現場の“どこ”で“何の作業”が行われているかを環境音から推定する手法を提案する。

◆音により作業管理をするメリットは、映像が苦手な、障害物による死角が多い場所や、暗所でも問題なく機能する点にある。

◆音源定位は、複数のマイクロフォンアレイを用いて、音源の位置を推定する手法を提案する。反響や残響の影響が大きい建築現場特有の環境下でも、作業管理システムに必要な精度を確認した。

◆建築現場で実際に収録した５種類の環境音を深層学習により、分類した。結果は、F1 値が、90.69%となり、作業管理システムに必要な分類性能を確認した。

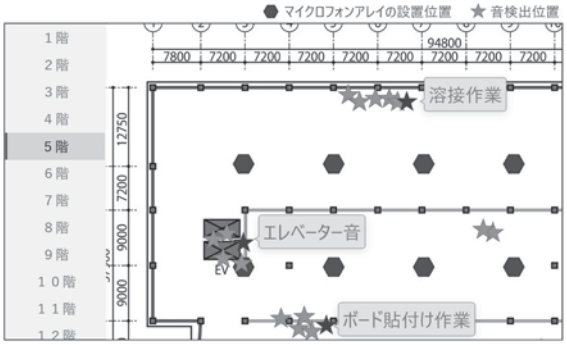


Fig.1：System image

3-12-11

3-12-11 仮想空間における響きの印象に関する研究 －仮想空間の視覚情報に調和する残響時間の検証－

Study on the impression of reverberation in virtual space －Verification of reverberation time in harmony with visual information in virtual space－

○風岡翔太，石川あゆみ(岐阜高専)

◆本研究では、仮想空間の視覚情報に調和する残響時間および、その調和に影響する視覚情報の要素を明らかにすることを目的とし、Fig.1 に示す室内空間の 3D モデルを用いて主観評価実験を実施する。

◆仮想空間の視覚情報を見て被験者が予想した残響時間を「予想残響時間」として実験結果を分析したところ、マテリアルやオブジェクトといった視覚情報がある空間条件(Fig.1 の D～I) で予想残響時間が想定する残響時間に近づき、そのばらつきは小さくなる傾向を確認できた。この理由として、マテリアルやオブジェクトに空間のスケール感を把握しやすくする効果があるためと考えられる。

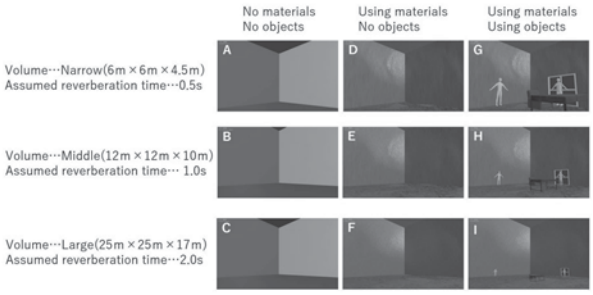


Fig.1：Screenshot of each spatial condition

3-12-12

3-12-12 仮想空間による実空間の再現性に関する研究 -実空間と 3D モデルで表現された仮想空間に対する主観評価実験-

Research on the Reproducibility of Real Space by Virtual Space -Subjective evaluation experiment for real space and virtual space represented by 3D model-

○木村健伸，石川あゆみ(岐阜高専)

◆本研究では、岐阜高専建築学科第 3 学年の教室(以下、教室)と教室を再現した 3DVR の仮想空間に対する主観評価実験を行い、実空間と視聴覚情報からなる仮想空間の印象差を明らかにするとともに、仮想空間のリアルさや実空間の再現性に寄与する要素を明らかにすることを目的とする。

◆教室の写真および 3D モデルを Fig.1 に示す。

◆主観評価実験の結果、仮想空間による実空間の再現性を向上させるためには、距離感や現実感、解放感といった視覚印象、音像位置や響き、力強さといった聴覚印象に関わる部分に注力して仮想空間を制作することで、再現性を高くすることが可能と考えられる。

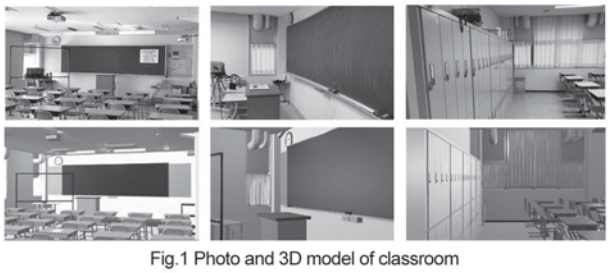


Fig.1 Photo and 3D model of classroom

3-12-13

3-12-13 室内音場のインテンシティと粒子速度を用いた側方エネルギー率の定義

Definition of lateral energy fractions using the sound intensity and particle velocity of sound field in a room

○羽入敏樹，星和磨，鈴木諒一（日大･短大）

◆演奏空間における ASW，LEV など音の空間印象の評価指標として ISO3382-1 に規定されている側方エネルギー率 J_{LF} と J_{LFC} の定義の問題点を取り上げ，その解決を試みた。

◆問題を解決するため，以下のような音圧，粒子速度，音響インテンシティなど音響物理量による J_{LF} と J_{LFC} の定義を提案した。

$$J_{LF}=\frac{\int_{0.005}^{0.080}\left[\left\langle \mathbf{u}(t),\mathbf{e}_x\right\rangle \rho c\right]^2 dt}{\int_0^{0.080}p^2(t)\,dt} \qquad J_{LFC}=\frac{\int_{0.005}^{0.080}\left|\left\langle \mathbf{I}(t),\mathbf{e}_x\right\rangle \rho c\right| dt}{\int_0^{0.080}p^2(t)\,dt}$$

$$=\frac{\int_{0.005}^{0.080}\left[u_x(t)\rho c\right]^2 dt}{\int_0^{0.080}p^2(t)\,dt} \qquad =\frac{\int_{0.005}^{0.080}\left|I_x(t)\rho c\right| dt}{\int_0^{0.080}p^2(t)\,dt}$$

ただし，$\mathbf{u}(t)$はインパルス応答の瞬時粒子速度ベクトル，$\mathbf{e}_x(t)$は x 軸+向きの単位ベクトル，$u_x(t)$は$\mathbf{u}(t)$の x 成分，ρc は空気の特性インピーダンス，$p(t)$はインパルス応答の瞬時音圧，$\mathbf{I}(t)$はインパルス応答の瞬時インテンシティベクトル，$I_x(t)$は$\mathbf{I}(t)$の x 成分である。ベクトルの座標系としては，y 軸を音源に向け，x 軸を側方（リスナーの両耳軸）とする右手系の 3 次元直交座標系を想定する。

◆従来の定義に基づく両指向性マイクによる測定結果と，新定義に基づいて CC 法で測定した結果を比較した。その結果，ISO3382-1 に基づく測定では大きな誤差が生じる場合があったが，新定義による測定は大きな誤差は生じなかった。

3-12-14

3-12-14　ソロ演奏受聴時の音響状態の判別に関する研究

A study on discrimination of acoustic condition listening to solo musical performance

〇徳永泰伸（舞鶴高専），寺島貴根（三重大）

◆本報の目的はインパルス応答受聴時とソロ演奏受聴時の音響状態の知覚の違いを明らかにすることである。

◆ホールで採取したインパルス応答や，これらのインパルス応答をソロ演奏に畳み込んだ楽曲を呈示刺激とし，被験者に差異の知覚についての回答を求めた。

◆Fig.1，Fig.2 は多次元尺度法（MDS）による被験者の回答結果の布置図であり，Fig.1 では dim.1 に対する布置に規則性が見られるが，Fig.2 ではその規則性が失われている。

◆ホール全体を対象範囲とした場合にはインパルス応答に対する判断と楽曲に対する判断の傾向は似ているが，対象範囲が小さくなると，楽曲を伴う音響状態の判別は，困難になる傾向にある。

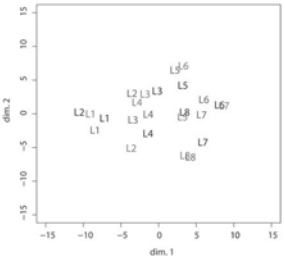


Fig.1：Result of MDS (Large area)

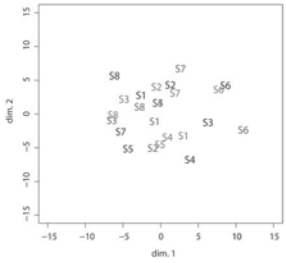


Fig.2：Result of MDS (Small area)

3-12-15

3-12-15　オンラインコミュニケーションにおける飛沫防止パーティションの影響に関する実験的検討

Experimental study on the effect of sneeze shield on online communication

〇秋山あさひ(東大大学院)，米村美紀，坂本慎一（東大生研）△長井達夫（東京理科大）

◆本検討では、飛沫次感染防止として設置されているパーティションが持つ、オンラインコミュニケーション利用者、その周囲にいる同室者への音響的影響を検討した。

◆パーティションの有無、配置、高さ、厚さを変更し、音響測定と心理評価を行った。音響測定では挿入損失、残響時間、STIr を評価項目とし,心理評価では、話者の話しやすさ、聴者の聞き取りやすさ等を評価項目とした。

◆残響時間と STIr についてはパーティション有無による違いはあまり見られなかったが、挿入損失は 1kHz 以上の帯域で周囲騒音を防ぐ効果が期待できる結果となった。

◆心理評価では、パーティションがある場合には騒音による不快感を取り除くことができることが分かった。

Fig.1：Condition of the psychological evaluation

Fig.2：Ease of hearing
(A:No sneeze shield B:H=600,T=3,1 C:H=600,T=3,2 D:H=600,T=5,1 E:H=900,T=3,1 1: 2:)

3-P-1

3-P-1　音声の時間周波数表現に対する嗄声判定の試み - Transformer ベースの画像認識の応用 -

An attempt to determine hoarseness for time-frequency representation of speech: Applications of Transformer-based image recognition

☆石原一樹，荒井隆行（上智大）

◆GRBAS 尺度は嗄声の重症度を表すものである．本研究では，機械学習を用いて，G 尺度の 4 段階を自動で推定することを試みた．

◆画像認識のモデルとして Vision Transformer (ViT)を使用した．オーバーラップ率を 50%に固定し，窓の長さのパラメータを変化させ，各パラメータでのスペクトログラムを生成した．G 尺度がラベル付けされたスペクトログラムを学習し，テストデータに対して G 尺度の推定を行った．

◆結果として，窓の長さ 20 ms，フレームシフト幅 10 ms のスペクトログラムを使用した際，正解率が 75%と最も高くなった．

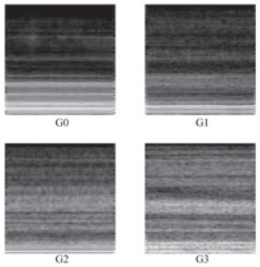


図 1: G 尺度のクラス毎のスペクトログラム
(窓の長さ 20 ms，フレームシフト幅 10 ms)

3-P-2

3-P-2　少量ラベル付きデータを用いた病的音声の声質自動評価

Automatic Evaluation of Pathological Voice Quality using a Small Set of Labeled Voice Samples

☆日高駿介，李庸學，△中西萌，若宮幸平，△中川尚志，鏑木時彦（九州大）

◆音声医学の臨床で行われる病的音声の声質の聴覚印象評価（GRBAS 尺度）には，評者間・評者内変動が原因で再現性に欠ける問題がある。機械学習に基づく声質自動評価は再現性の問題を解決するが，ラベル付きデータを大量に集めることは困難である。本研究では，ラベル付きデータが少ない状況下での高精度な GRBAS 自動評価の実現を試みた。

◆本研究では，持続母音/a/の音声波形から 3 種類の時間周波数表現(対数振幅スペクトログラム，短時間フーリエ変換の位相の時間微分である瞬時周波数と周波数微分である群遅延)を計算し，GRBAS 尺度の点数の確率分布を出力するニューラルネットワークを構築した。データ増強技術として，音声波形のランダム速度摂動，音声波形のランダムクロップ，時間周波数表現の周波数マスキングを検討した。8 名の専門家によって GRBAS 評点が付与された，300 の病的音声サンプルから構成されるデータセットを構築し，ネットワークの性能評価実験を行なった。

◆最良の推定条件の場合，自動評価の性能は，GRBAS 全項目で専門家の評者間信頼性に匹敵し，項目 G, B, A, S で専門家の評者内信頼性に匹敵した。データ増強技術に関して，ランダム速度摂動が全体的に最も有効であった。時間周波数表現に関して，項目 G, R, A では対数振幅のみを用いた時に最良の結果が得られたのに対し，項目 B（気息性）と S（努力性）では対数振幅と位相微分の併用がさらなる性能向上をもたらした。本研究で得られた知見は，喉頭疾患の自動検出など他の課題にも応用可能であると考えられる。

3-P-3

3-P-3　MRI モデルを用いた声帯ポリープの発声への影響

Effect of vocal fold polyp on the oscillation properties of the MRI 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。

◆ポリープのある物理モデルとポリープのない物理モデルを作成し、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た。

◆2 種類の大きさ（3mm と 5mm）を持つポリープを、MRI モデルの 3 種類の異なる位置に付加した。

◆ポリープがあることでオンセット圧は高くなるが、発声効率にはあまり影響しないことが分かった。また、ポリープの位置よりも大きさの方が発声に影響することが分かった。

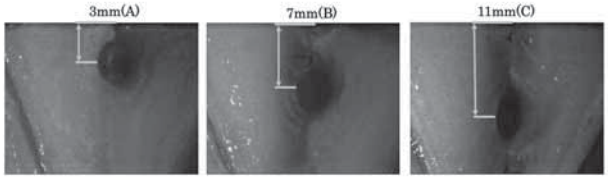


Fig.1：Physical model of vocal fold polyp

3-P-4

3-P-4　感音性難聴のある若年者による歌唱支援システムの評価 —視覚と振動による音高のフィードバック—

Evaluation of singing training system by young deaf and hard-of-hearing person: Visual and tactile feedback of voice pitch

○安啓一，坂尻正次(筑波技大)，藪謙一郎(東大)，
三浦貴大(産総研)，△片桐淳(プッシュ・ポップ)，伊福部達(東大)

◆若年の聴覚障害のある参加者を対象として歌唱時の音高フィードバックを，視覚及び触覚(Fig. 1)を通して行った。

◆ド（C3）およびラ（A3）の音を 4～5 秒持続発声させ，ターゲットの音高と自声の音高をそれぞれ視覚および触覚でフィードバックした。

◆視覚フィードバックでは 2 秒程安定している区間があったが，触覚では開始 1s 後に音高にずれが生じた。

◆先行研究の盲ろう者を対象とした実験に比べ、聴覚障害者は普段から触覚デバイス（点字ディスプレイなど）に慣れていないため，実験前の訓練が重要であることが示唆された。

ディスプレイ（視覚）　ピンデバイス（触覚）

Fig.1：視覚および触覚による音高のフィードバック

3-P-5

3-P-5　第二言語学習における音声生成能力の変化が音声知覚能力に与える影響

The effects of speech production change on speech perception in second language learning

○末光厚夫(札保大)，能田由紀子，前川喜久雄(国語研)

◆本報告では，音声生成能力の変化が音声知覚能力に与える影響を再検証するために，著者らが構築した WAVE を使用した第二言語発音学習システムを用いて，日本語母語話者（4 名）にアメリカ英語（AE）母音/æ/の調音訓練を行い，訓練によって生じる音声生成能力と音声知覚能力の変化を比較した.

◆訓練前後で/æ/を含む単語を発音した際の音声を計測し，/æ/区間の F1 と F2 を LPC 分析により抽出し，AE 母語話者の/æ/との間のマハラノビス距離を計算して平均した（Fig.1）．また，訓練前後で/æ/，/ɑ/，/ʌ/が最少対立となる CVC 単語を刺激として用いた弁別課題を行い，/æ/が含まれる刺激が用いられた試行の正答率を求めた（Fig.2）.

◆被験者 S1 と S3 において，調音訓練によって/æ/の発音が音響的に向上し，/æ/の弁別成績も向上していることがわかる．この結果は先行研究と同様の結果であり，音声生成能力の向上が音声知覚能力の向上に寄与することを示唆している.

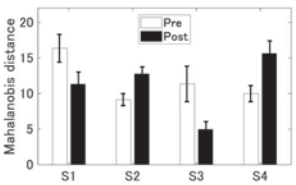


Fig.1：Mahalanobis distance between the produced and reference sounds at the pre- and post-test phases.

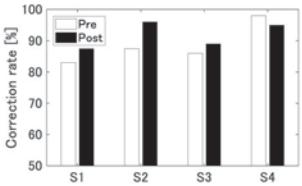


Fig.2：Correction rates of /æ/-/ɑ/, /ʌ/ before and after articulatory training.

3-P-6

3-P-6　「○ッグ」の発音と表記

Study on orthography and pronunciation of /Qgu/ in Japanese

○白勢 彩子(東京学芸大)

目的・概要　小さい「っ」で表記される促音は，物理的特性としては主に子音の閉鎖あるいは摩擦区間の伸長として現れる。後続が無声の破裂・破擦音，摩擦音の場合に限られ，表記としては清音になることが知られている。近年，外来語において有声子音に前接する促音（有声促音；「ビッグ」「バッグ」など）が生じているとの指摘がある。現代の日本語における有声促音の分布について，後続を「グ」とする促音を対象に，コーパスによって表記と対応づけつつ，現状の生起傾向について分析した。

手続　発音については日本語話し言葉コーパス(CSJ)，日本語日常会話コーパス(CEJC)，表記については現代日本語書き言葉均衡コーパス(BCCWJ)，国語研日本語ウェブコーパス(NWJC)を用いた。語彙素に「ッグ」を含み，かつ書字形または発音形出現形に「ッグ」/「ック」を含むものを検索した。

結果　有声促音は書きことばにおいて圧倒的な出現頻度が確認できた一方で，話しことばでは/Qgu/と/Qku/が共存していることがわかった（/Q/は促音拍）。

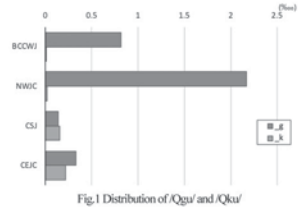


Fig.1 Distribution of /Qgu/ and /Qku/

/Qku/の発音に世代間の相違はないこと，一方，「ッグ」の表記は 1995 年以降，特にウェブ媒体のデータで大幅に増加していることが考えられた。これらのことから，有声促音の発達については，表記が先行して生じたのではないかと考えられた。

3-P-7

3-P-7　Sequential VAE 潜在変数による日本語アクセント型の判定

Classification of Japanese accent types based on static latent variables of Sequential VAE

○勝瀬郁代（近畿大・産業理工）

◆日本語単語音声のピッチアクセント型の特徴表現を得るために，Sequential VAE による F0 のモデル化を試みている．

◆モデルでは，潜在変数は，アクセント型の特徴表現を担う静的な潜在変数 Z_f と，有声／無声の特徴表現を担う動的な潜在変数 $Z_{t:T}$ から成る．

◆Fig. 1 は，F0 のテストデータをモデルに入力して得られた潜在変数 Z_f の分布を示す．Fig. 2 は，0 型アクセントの F0 に対し，抽出された $Z_{t:T}$ をそのまま使用し，Z_f の値を(-3, -3)から(3, 3)まで変化させて F0 を再構成した例を示す．Fig. 1 で観察されたアクセント型の分布に従い，Fig. 2 では，F0 の形状が変化している．

◆提案モデルと GRU モデルと全結合モデルで，アクセント型識別実験を行った結果，各正答率は，提案モデル：0.962，GRU：0.863，全結合モデル：0.907 となった．

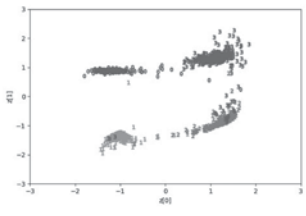


Fig.1：Distribution of latent variables Zf corresponding to each input F0.

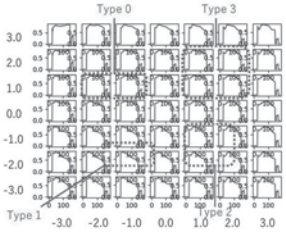


Fig.2：F0s decoded with Z_f from (-3, -3) to (3, 3).

3-P-8

3-P-8　授業内音声イベント検知による授業活性度の推定

Estimation of class activity level by acoustic events during class

中野魁人（滋賀大・院），高瀬壮彦（滋賀大・院），○市川治（滋賀大）

◆本研究では，授業の活性度という指標を導入し，客観的な授業評価の一助とすることを目標としている．ここで「授業活性度が高い」とは，教師が生徒とインタラクティブな授業を行うスタイルのことを想定しており，良い授業かどうかという観点には踏み込んでいない．

◆教員に装着したピンマイクを用いて，実際の中学校にて 21 コマの授業の音声を収録した．この音声を入力し，板書，教師発声，生徒発声，グループワーク，笑い声，拍手，沈黙，個別対話といった音イベントの発生を推定する．これについては前回に報告した．

◆本報告では，正確な音イベントが得られたとして，それらを入力として対象授業の活性度指標を算出する機械学習モデルの構築を試みる．

◆21 個の授業データについて 21 fold のクロスバリデーションを行った．モデルの学習には LightGBM を使用し，3 クラスの分類問題として学習を実行した．

◆教師データとなる評点は，２～４名の評価者の評点を補間・標準化し，評価者の平均をとっている．実験結果を表 1 に示す．評点の大きい授業ほど，活性度の高い授業である．人の評価には揺らぎがあるので，発表当日までに，評価者を増やした実験結果を用意する予定である．

Table. 1 Experimental result

		Predicted			
		0	1	2	recall
Actual	0	4	2	0	0.67
	1	2	8	0	0.80
	2	1	3	1	0.20
	precision	0.57	0.62	1.00	

3-P-9

3-P-9　ささやき単語のアクセントの差異がスペクトル特徴に及ぼす影響

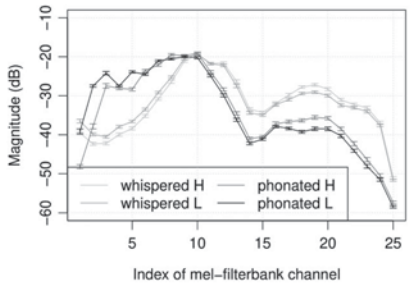
Effect of Whispered Word Accent on Spectral Feature.

○今野英明（北海道教育大）

◆音素列は同じでありながらアクセントによって意味の異なる 90 単語について，ささやき声と通常発声の母音部のメルフィルタバンク分析
- 話者6名; アクセントあり1068 母音, アクセントなし 1128 母音

◆アクセントの有無でスペクトルの低域／高域のバランスが変化

◆全体に 500 Hz 付近以下の周波数領域において，アクセントのある母音のエネルギーが小さい傾向
- ピッチの高い孤立発声のささやき母音と同じ傾向
- 日本語のささやき声もピッチアクセント？
- 通常発声も同様の変化：ピッチに関わる共通の特徴？



Mel filterbank analysis of whispered and phonated /a/ uttered by 6 speakers. (H: accentuated, L: not accentuated)

3-P-10

3-P-10　高齢者に対して強調すべきと判断された単語のテキストからの予測と分析

Prediction and analysis of words to be emphasized for elderly adults

○中嶋秀治（日本電信電話(株)，NTT コミュニケーション科学基礎研究所），水野秀之（公立諏訪東京理科大学）

◆**背景**：
- 高齢者にとって聞いて分かりやすい音声や話し方を調べている
- 本研究に先立ち，高齢者が聞いて分かりやすいと判断した模範話者から，話す内容の重要箇所で音声表現上の**強調を高齢者向けに行なう**というインタビュー結果を得ていた

◆**目的**：
- 音声表現上の強調を加えるために必要な**テキストからの強調位置予測方式**の評価を行なう

◆**手法**：
- 単語系列から強調／非強調の系列への**系列予測問題として設定**
- 深層学習器のひとつである BERT を適用
- 文書単位のモデルと文単位のモデルを比較

◆**結果**：
- 強調／非強調の精度は 80%前後
- 文書単位モデルの精度が若干高かった．

◆**今後にむけて**：
- 精度改善に向けて，強調／非強調ラベル付けの揺れの心配される細かな単語単位ではなく，少し大きな文書タイトルや箇条書きといった**文書レイアウトの単位と強調／非強調との間の関係分析を行なった結果**を示す

3-P-11

3-P-11　日本人ドイツ語学習者による留学経験が分節的・超分節的特徴における発音に与える影響

The effects of study abroad on segmental and suprasegmental features by Japanese German learners

◎粕谷麻里乃（東邦音大・音），荒井隆行（上智大・理工）

◆目的

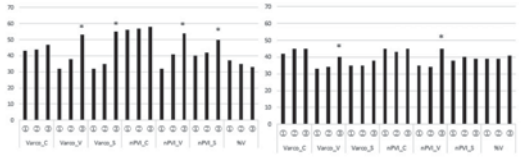
留学経験は，ドイツ語学習者の発音を母語話者に近づけるのか？被験者の発音を音響特徴量の観察により定量的に評価する．

◆手法

日本人ドイツ語学習者 2 群 (ドイツ留学経験の有無) を対象に，リズム指標 PVI (Pairwise Variability Index) と Varco (Variation Coefficient) を用いて発音の韻律と弱音節母音を観察した．発音上の変化は，いつ頃どのように出現するのか，計 3 回 (滞在直前，1 か月後，2 年後) にわたり，被験者のドイツ語朗読音声を評価した．

◆結果

滞在初期の変化は，ドイツ滞在 (SA) 群の弱音節母音の調音 (特に F2) と，母音区間長にみられた．その後，2 年間目標言語環境下で生活することで，音節長にも変化がみられたことは，非ドイツ滞在 (NSA) 群と大きく異なった．一方，母音間子音の変化は乏しく，子音の扱いに母語の影響が残りやすさが観察された．



Fig：SA群 (左) と NSA群 (右) の韻律評価 (留学前①, 1ヵ月後②, 2年後③) (*=p<0.05)

3-P-12

3-P-12　声門幅を狭めた声帯物理モデルにおける声門内圧測定: 正弦波非定常流の周波数が与える影響

Measurement of intra-glottal pressure with narrowed glottal width: frequency effect of the unsteady flow

○中村和史，井上卓哉，△中川栞里，徳田功（立命館大）

◆アクリル製 M5 型声帯物理モデル（Fig.1）を作成した。

◆声門幅を 0.20 mm に固定した。

◆定常流を流す実験と流量が正弦波状に変動する非定常流実験を行い，声門内圧，声門下圧を測定した。

◆非定常流の周波数を 1 Hz と 0.4 Hz と設定した。

◆定常流実験と非定常流実験における声門内圧力分布の比較を行った。

◆周波数が 1 Hz のとき，定常流と非定常流で計測された声門内圧力分布には有意差が見られたのに対し，0.4 Hz のときには，ほとんど有意差が見られなかった（Fig.2）。

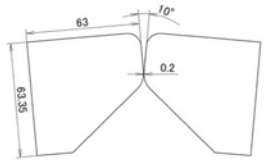


Fig.1：Configuration of M5 Model.

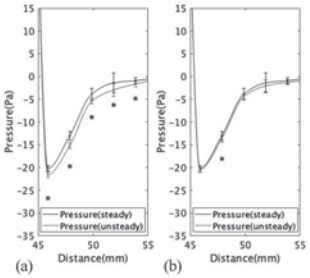


Fig.2：Comparison of the intraglottal pressure distributions between steady and unsteady flow experiments. Frequency was 1Hz in (a), 0.4 Hz in (b).

3-P-13

3-P-13　周波数を変えた声帯物理モデルの吸気発声実験

Effect of fundamental frequency of the vocal fold physical model on the inspiratory phonation

☆長谷川寛人（立命館大），△足立恭平（立命館大），徳田功（立命館大）

◆硬さの異なる二種類の MRI 声帯モデルを用いて、周波数を変えた呼気・吸気発声の比較実験を行った。

◆呼気発声と比較して、吸気発声では、基本周波数が高くなり、オンセット流量およびオンセット圧は下がった。これらの傾向は、声帯の硬さ、すなわち声の高さに依存しないことが分かった。

◆得られた結果は、人の呼気発声に関する先行研究と矛盾しないことが確認できた。

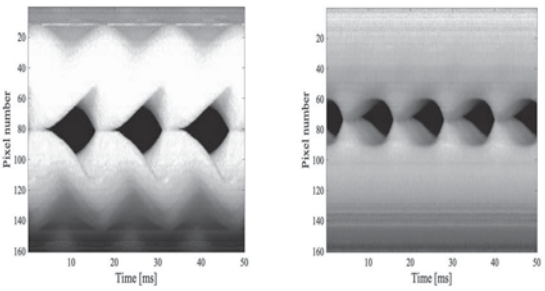


Fig.1 Kymogram of expiratory (left) and inspiratory (right) phonations.

3-P-14

3-P-14　声帯と声帯膜の間に存在する溝が声帯膜物理モデルに及ぼす影響について

Influence of sulcus on physical model of vocal fold and vocal membrane

☆槌田悠希，金谷麻由佳，吉谷友紀，△小畑大樹，徳田功（立命館大）

◆動物喉頭に見られる声帯膜下の溝を模擬する物理モデルを構築した。

◆溝の大きさを Large と Small の二種類、声帯膜の硬さを Hard と Soft の二種類作成した。

◆吹鳴実験を行い、声帯と声帯膜の反相振動が起こる割合を調べたところ、溝よりも声帯膜の硬さの方が、膜振動に大きい影響を与えることが分かった。

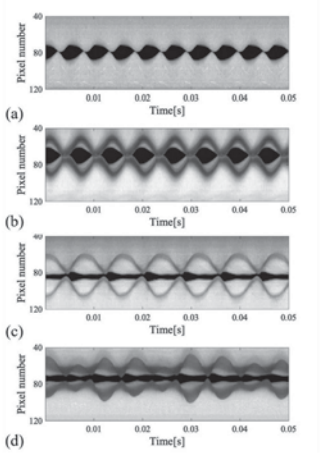


Fig. 1 (a)Kymograms without VM, (b)with hard VM/large sulcus, (c)with soft VM/large sulcus, (d)with soft VM/small sulcus

3-P-15

3-P-15　声帯モデルと仮声帯モデルの固有振動周波数を考慮した片側喉頭吹鳴実験

Hemilarynx experiment of vocal and ventricular fold models

☆松島大輔（立命館大），金谷麻由佳（立命館大），韓聰（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、片側喉頭吹鳴実験を行った。
◆声帯モデル単体の振動周波数は 340 Hz、仮声帯モデル単体の振動周波数は 135 Hz であった。
◆声帯と仮声帯を重ねて吹鳴実験を行ったところ、二つのモデルの同期的な振動は見られなかったが、声帯と仮声帯の周波数比が整数に近い振動を示すことが確認できた。
◆仮声帯間距離を広げると、周期と振幅は不規則になった。

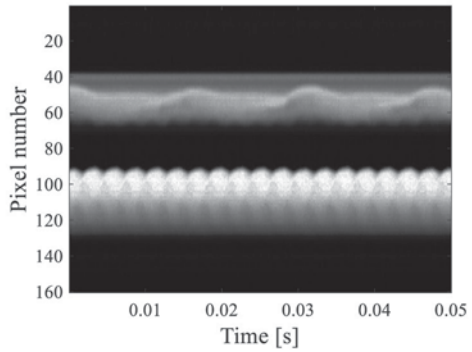


Fig.1 Kymogram observed when ventricular fold distance was 0 mm

3-P-16

3-P-16　アカゲザルのマイクロ CT 画像を用いた声帯物理モデルの構築

Construction of vocal fold physical model using micro-CT images of rhesus monkeys (*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（*Macaca mulatta*）の喉頭をマイクロ CT 撮影し，生理データに基づく物理モデルを構築した。
◆吹鳴実験を通して，声門下圧，音声，高速度映像を計測し，仮声帯のある物理モデルとない物理モデルを比較した。
◆摘出喉頭でも同様の実験を行い，振動特性を比較した。
◆仮声帯があることで基本周波数が低くなることがわかった。
◆摘出喉頭で見られた 1:2（仮声帯：声帯振動数）の同期振動が，本モデルでも観測された。

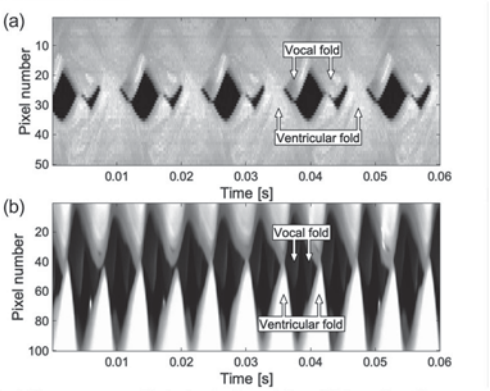


Fig.1:Kymograms of (a) physical model and (b) excised larynx.

3-P-17

3-P-17　磁気センサシステムによる調音運動計測のための口蓋・咬合面の計測法

Measurement of 3D palate shape and occlusal plane for observation of articulatory movements using EMA

○能田 由紀子(国語研)，北村 達也(甲南大)，
竹本 浩典(千葉工大)，前川 喜久雄(国語研)

◆磁気センサシステム（EMA）による調音運動計測では口蓋の三次元形状や咬合面を精密に計測することは難しい。
◆前報では，バイトプレートを用いて口蓋の精密な形状を歯科印象材で取得すると同時に EMA でその位置計測を行い、光学計測した口蓋形状を EMA の計測座標系に写像する方法を提案した。
◆本報では，写像の誤差を軽減するため、センサ間距離やセンサの位置などバイトプレートの形状を再検討した。
◆その結果、写像後の光学計測した口蓋と，EMA で計測した口蓋の位置の誤差が１ミリ以下となり、精密な口蓋形状を正確に調音運動の計測座標系に組み込むことができた。

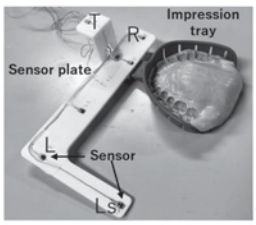


Fig.1 Bite plate consisting of a sensor plate with 4 sensors and an impression tray

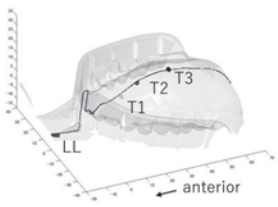


Fig.2 Sensors on the tongue and lower lip during the rest position are overlaid on the palate with midsagittal line.

3-P-18

3-P-18　声区に着目した歌唱発声時の声道形状の観測

Observation of the vocal-tract shape during singing vocalization focusing on the vocal register.

☆木下珠子，李庸學，鏑木時彦（九州大）

◆主要３声区である地声声区・裏声声区・ミックス声区について，声区ごとに声帯振動パターンが異なることがわかっている。本研究では，ポピュラー歌唱者について、特にミックス声区に着目し、MRI による調音器官及び声道形状の可視化を行った。
◆ポピュラー歌唱者の成人男性１名・成人女性１名が主要３声区と話声を発する際の声道形状を３次元 MRI 撮像し，声区ごとの正中面画像の比較と声道断面積関数の算出を行った。
◆検討の結果，ミックス声区において軟口蓋の挙上が小さくなり，鼻腔への音波伝搬が発生している可能性が示唆された。また，歌唱者が声帯振動パターンに加えて声道形状をも変化させて声区のコントロールを行なっていることが示唆された。

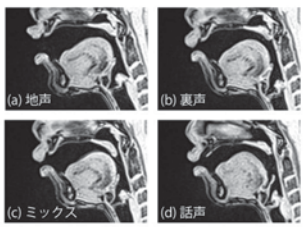


図 1 女性被験者が音高 G4（392 Hz），母音/o/を発声した際の声区ごとの声道形状

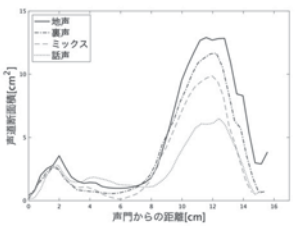


図 2 女性被験者が音高 G4（392 Hz），母音/o/を発声した際の声区ごとの声道断面積関数

3-P-19

3-P-19　高速度ディジタル撮像に基づいたホイッスル声区の声帯振動パターンの分類

Classification of vocal fold vibration patterns based on high-speed digital imaging in whistle register

☆加藤日花里　李庸學　若宮幸平　鏑木時彦（九州大）

◆歌唱声区の中で最も基本周波数の高いホイッスル声区における声帯振動メカニズムを明らかにするため、本研究では、10000 fps を超えるフレームレートの高速度ディジタル撮像(HSDI)によって、4 名の歌手の喉頭を観察し、声門面積波形とフォノバイブログラム(PVG)からスペクトルと声門開放率を算出して分析を行った。

◆観測された声帯振動パターンは声門の閉鎖状態から(a)完全閉鎖、(b)不完全閉鎖、(c)部分閉鎖の3パターンに分類することができた。

◆PVG は声帯長方向の声帯振動情報を表すことができ、声帯振動様態の分析に有効であった。

◆先行研究ではそれぞれのパターンは単独で確認されていたが、本研究では、3 つの振動パターンは同時に存在することが明らかになり、ホイッスル声区は単なる超高音域の裏声ではなく、声帯音源の生成に多様性を持つ声区と考えられる。

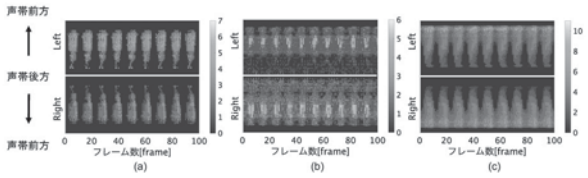


Fig.1 : PVG of each vibration pattern.
(a) Complete closure, (b) Incomplete closure, (3) Partial closure

3-P-20

3-P-20　表情を用いたボイストレーニングの試みと考察

Attempts and consideration of vocal coaching using facial expressions

〇高野佐代子、長塚全（Zen Voice Factory）、土田義郎（金沢工大）

◆ 防災放送を行う一般職員の方に対し、短時間かつ効果的なトレーニング法を模索している。

◆ 手本の顔の表情（PC1）と自身の顔の特徴点（PC2）を同時に観察して発話するシステムを開発した。手本として怒り・悲しみ・喜びの演技を収録したものを用い、防災放送を無意味語で置き換えた「近くの川がソビサンしました。非常にケーメな状態です。直ちに指定の小学校にセタンしてください。」を発話させた。

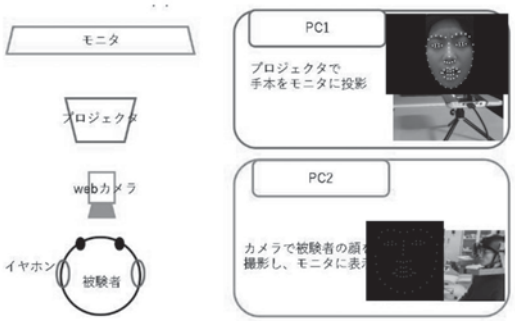


Fig. The use of facial expressions for vocal coaching

3-P-21

3-P-21　非言語的な発声によるポジティブ感情の表出と知覚

The expression and perception of positive emotions through nonverbal vocalization

☆大屋里佳（東京女子大・学振），田中章浩（東京女子大）

◆様々なポジティブ感情を非言語的な発声（e.g., 笑い声）から伝達させたところ，９種類もの感情が伝達された。

◆感情表出の音響解析を実施し，解析結果に基づいて，当該感情が知覚されやすい「典型」を見出した。

◆声の高さと大きさの「典型」で感情を表出すると，意図感情が伝わりやすいことを示した。

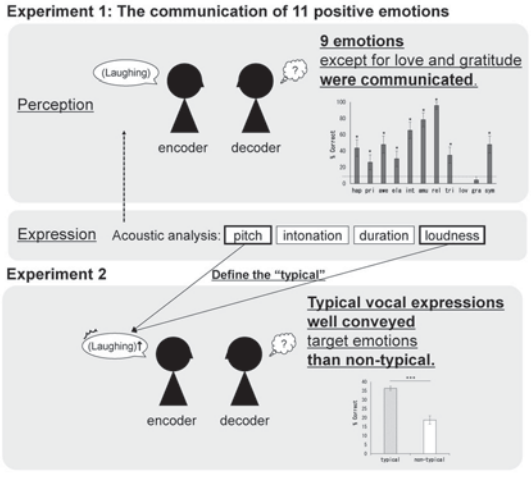


Fig 1. The illustration of this study

3-P-22

3-P-22　雑音下における音声聴取成績と皮質の時間情報処理の関係

Relationship between speech perception and cortical temporal information processing in noise

☆齊官重樹，大塚翔，中川誠司（千葉大）

◆ 聴力検査が正常である人でも，特に騒音や部屋の残響がある時にコミュニケーションの困難さを訴えることがある．この困難さが隠れた難聴であり，現状では実用的な診断手法は提案されていない．

◆ 皮質レベルの時間情報処理については，劣化度合いの異なる合成音声を用いて，了解度の上昇に伴ってδ波およびθ波と音声の同期性が上昇することが報告されている．一方で，競合音存在下における聴取成績とδ波およびθ波の関係は明らかになっていない．

◆ 本研究では，SN 比を段階的に変えて雑音下での聴取成績を測定し，同時に計測したδおよびθ帯域の脳活動との関係を調べた．

◆ SN 比が低下することに伴って聴取成績が低下することを確認した（Fig. 1）．被験者を増やすことで，PLV と聴取成績の相関関係の有意性を検証する必要があるもの，後頭部付近の電極で測定したδおよびθ帯域の PLV と聴取成績との間の相関が最も高かった(Fig. 2)．しかし，PLV と SN 比との関係は明確ではなく，PLV を指標として雑音下での聴取成績を推定するに至らなかった．

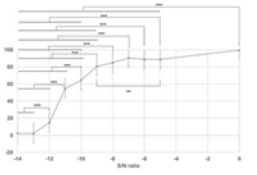


Fig.1 Correct answer rate of degraded speech with different formant bandwidth. **p<0.01, ***p<0.001 (Corrected for multiple comparisons with Ryan's method).

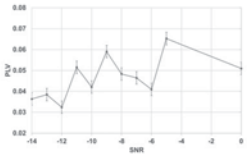


Fig.2 PLV of δ − θ band as a function of SNR at O2

3-P-23

3-P-23　基本周波数の周波数変調に対する発声の不随意応答：刺激音の種類・変調量による影響

Involuntary vocalizations corresponding to modulated fundamental frequency: Effects of the type and modulation of stimulus

☆廖嘉慧(豊橋技科大)，河原英紀(和歌山大)，松井淑恵(豊橋技科大)

◆ランダムな基本周波数変調のある刺激音を作成し，それを聴取しながら発声した際の発声の不随意応答を調査した．

◆刺激には，純音 SINE とその倍音を重ねた複合音 SINES (1~20 倍音)，MFND (2~20 倍音)，MFNDH (8~20 倍音)を用いた．

◆SINE 条件以外では刺激音のパルスの後に，刺激と反対方向への発声ピッチの変化（補償応答）が見られた．

◆変調量を変えても応答の大きさは変わらなかった．

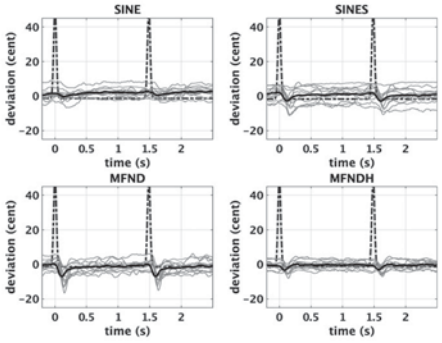


Fig. 1. Results of pulse recovery analysis for each type of stimulus sound. Dotted line, stimulus sound; black line, subjects' average response; gray line, each subject's response.

3-P-24

3-P-24　Perception of Mandarin and English Liquids by Japanese Native Speakers: The Effects of L2 and L3 Experience

○ Peng Yongzhe (Sophia University)

◆This study investigates the perceptual performance of Mandarin liquids /l/ and /r/ by native speakers of Japanese, focused on the effect of L2 (English) experience.

◆Three groups of speakers with different language experience participated in an oddity discrimination task involving liquids from both Mandarin and English.

◆The results showed that the perceived dissimilarities between the liquids leading to the similar perceptual performance among the three contrasts.

◆Language experience is an important factor for learning these consonants. Participants with more learning experience of Mandarin are easier to distinguish these consonant pairs, even the English contrasts.

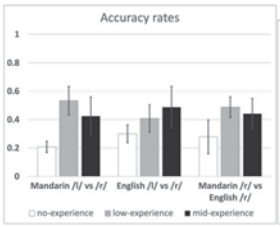


Fig.1 : Average accuracy rates.

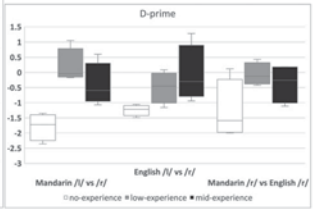


Fig. 2 : D-prime scores.

3-P-25

3-P-25　自己聴取音における音色と音高の印象に関する調査

Study on timbre and pitch differences in self-perceived own voices

◎森田翔太(福山大)，鳥谷輝樹(金沢大)，鵜木祐史(北陸先端大)

◆目的：音高・音色の事前スクリーニングテストにより，録音音声と自己聴取音（自身が発話しながら聴いている音声）の音高や音色について，統制を取って調査することの効果を検討する．

◆実験：実験参加者の母音 /a/，/i/，/u/，/e/，/o/と鼻音 /n/を録音し，実験協力者に音高の高い・低い，音色の違いの有無について録音音声と自己聴取音を聴き比べて回答させた．

◆結果（スクリーニング後）：

- 音高の違いについての結果を図１に示す．
  - ✧ 母音すべてで違うと判断されて安定した結果が得られたが，高い・低いの判断では回答が割れて個人差があった．
- 音色の違いについての結果を図２に示す．
  - ✧ /e/の同じと回答した１名を除いてすべてで違うと判断された．
  - ✧ スクリーニング前に比べ，安定した結果が得られた．

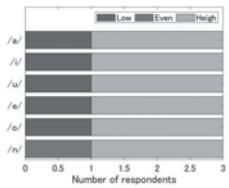


Fig.1 : Results of pitch difference for self-perceived own voices relative to recorded speech.

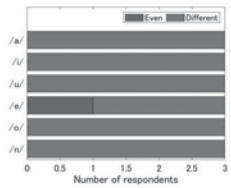


Fig.2 : Results of timbre difference for self-perceived own voices relative to recorded speech.

3-P-26

3-P-26　頭皮上で検出される骨伝導音声の調音素性情報伝達特性

Transmission characteristics of phonetic features of bone-conducted speech detected on the scalp

☆南里聡志，大塚 翔，中川誠司(千葉大)

◆骨伝導マイクロホン (BCM) には，周囲騒音を検出しにくいという特長があるが，装用性に問題があった．対して，我々はヘルメットに BCM を組み込んだ新型骨伝導デバイスを提案し，頭皮上から検出される骨伝導音声 (以下，頭皮上骨伝導音声) の評価に取り組んでいる．

◆前報では，頭皮上骨伝導音声を用いた明瞭度試験を行い，前額部・頭頂部が骨伝導マイクロホンの検出部位として有用である可能性を示した．しかし，音声特徴量ごとの伝達特性は明らかになっていない

◆Sequential information analysis (SINFA) を用いて，頭皮上骨伝導音声の調音素性情報伝達特性を検討した．

◆平均すると骨伝導音声の伝達率は気導音を下回るが，検出部位間の差異が大きかった．特に前頭極正中部 (Fpz) では気導音と同等，もしくは気導音を凌ぐ伝達率を示すことが多かった．

◆調音素性のうち，調音位置の情報の伝達率が最も低かったが，前頭極正中部 (Fpz) の骨伝導音は気導音を凌ぐ伝達率を示した．一般に骨伝導検出音は 1 kHz 以上の高周波成分の伝達に不利とされるが，口腔に近い前頭極正中部 (Fpz) では高周波成分がよく保存されていることを示唆している．

Table 1 SINFA results summary for response to consonants. (Transmission rate (%) (Transmitted information (bits))

Phonetic feature	AC	Lar	Mas	Che	Fpz	Cz	Oz
Voicing	61.0(0.51)	52.5(0.44)	47.9(0.39)	54.0(0.45)	61.1(0.50)	55.6(0.46)	66.3(0.64)
Nasality	79.4(0.51)	58.9(0.38)	66.8(0.43)	70.9(0.45)	70.4(0.45)	70.9(0.45)	45.2(0.22)
Manner	36.0(0.18)	25.9(0.29)	31.4(0.36)	32.0(0.37)	33.7(0.17)	31.3(0.36)	22.2(0.11)
Position	34.8(0.58)	16.6(0.17)	22.8(0.24)	26.4(0.28)	43.0(0.72)	25.4(0.27)	23.1(0.39)
Contracted sound	57.0(0.56)	37.5(0.37)	58.0(0.58)	48.4(0.47)	72.2(0.72)	62.1(0.62)	51.8(0.51)
Articulation(%)	55.89	40.07	44.96	52.33	62.81	54.52	45.67

3-P-27

3-P-27　高齢難聴者の音声了解度客観評価を目指した GESI の開発

－ 強調音声と模擬難聴音声による評価 －

Development of GESI to predict speech intelligibility of elderly hearing-impaired listeners

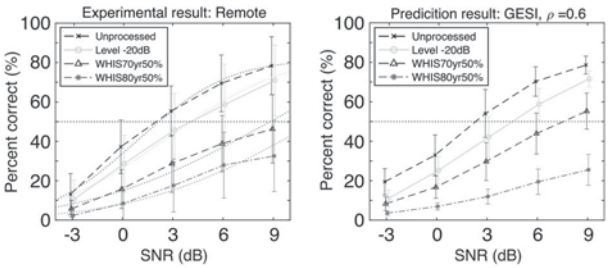
- Evaluation with speech enhancement and simulated hearing loss -

☆山本絢子，入野俊夫(和歌山大)，荒木章子(NTT)，△田丸萌夏(和歌山大)，新井賢一，小川厚徳，木下慶介，中谷智広(NTT)

◆高齢難聴者の音声了解度予測を目指して Gammachirp Envelope Similarity Index (GESI) を開発

◆強調処理音声と模擬難聴音声の主観評価実験を防音室とクラウドソーシングで実施 ➔ GESI で予測できるか検証・従来手法と比較

- 強調処理音声
  - マスクベース MVDR ビームフォーマによる強調処理：予測可能
  - Ideal ratio mask (IRM) による強調処理：過剰評価
- 模擬難聴音声
  - 従来予測できなかった聴取環境による了解度の違いを予測可能



Speech intelligibility (%) of simulated hearing loss sounds.
Left: subjective evaluation in crowdsourcing, Right: prediction using GESI.

3-P-28

3-P-28　拡声環境を想定した音声了解度指標 GESI と従来手法との比較

Comparison of GESI and conventional speech intelligibility indicators by actual environment

☆渡邊健太郎(室蘭工大)，小林洋介(室蘭工大)，入野俊夫(和歌山大)

◆ガンマチャープ聴覚フィルタバンク (GCFB)と変調周波数フィルタバンク(MFB) による特徴分析と拡張コサイン類似度を計算して統合した客観音声了解度指標 Gammachirp Envelope Similarity Index (GESI)が提案された。

◆GESI は拡張コサイン類似度の重みρにより，評価信号と基準信号の音圧レベル差に対応する。この重みρは語音聴取閾値 (SRT) と関係することが報告されている。

◆本報告では，実環境を想定した８条件の評価音声信号とその主観評価値を用いて，GESI の推定性能を評価する。

◆評価に用いた８条件の主観評価結果は，SRT が異なるため，重みρの値を 0.55 と 0.60 を比較した。また従来型了解度指標として ESTOI とも比較した。

◆結果より，８条件中５条件で GESI は ESTOI よりも主観評価との SRT 差が 0 に近く，SRT が 0 より大きい３条件では，ρの値が 0.55 より 0.60 の方が主観評価値に近い傾向にあった。

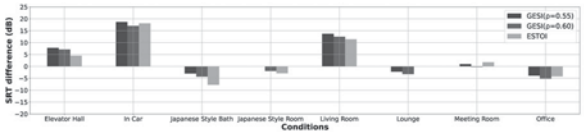


Fig.1：Results of SRT differences by environment

3-P-29

3-P-29　演技未経験者の感情音声の演技発話における台本の影響

The influence of the script on the acting speech of the emotional voice by non-actors

☆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)

◆複数の話者による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収録を行う際の台本の影響を調査した。

◆台本は２種用意し、台本１では表現する感情と感情が起きる状況、台本２では話し手の年齢・職業・人物像・感情・発話背景、聞き手の年齢・職業、聞き手との関係を設定した。

◆大学生の 20 代男性６名を対象に収録し、WORLD を用いて F0 解析を行った。

◆感情指示によって台本１と台本２の F0 平均に差があった。感情間の差としては、喜びの音声が他の感情と比べて F0 平均が高かった。

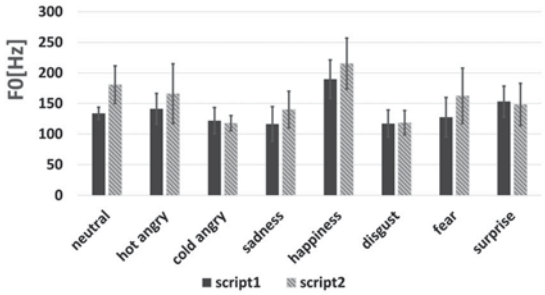


Fig. 1: F0 average of six male speakers by each emotion and script. Error bar: SD between speakers. The black bar corresponds to script 1 (simple situation instruction). The gray bar corresponds to script 2 (speaker details instruction).

3-P-30

3-P-30　ラジオ聴取経験が及ぼす音声感情知覚への影響

The effect of radio-listening on perception of emotional prosody

☆鍬真衣，田中章浩(東京女子大)

◆音楽経験のような音に関する経験によって音声感情知覚の精度が変化することが明らかになっている。本研究では「音声のみを聞く経験」といえるラジオ聴取経験が音声感情知覚に及ぼす影響について検討した。

◆ラジオリスナーと非リスナーに感情表出強度の異なるモーフィング音声（喜び・怒り・悲しみの３種類）を呈示し，感情ごとにその感情が読み取れるか否か二肢強制選択させた。

◆その結果，リスナーの方が怒りモーフィング率の小さい音声から怒り感情を知覚した（Fig. 1）。ラジオリスナーは音声からわずかな怒り感情を読み取れる可能性が示された。

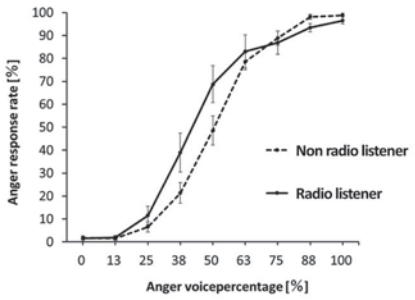


Fig.1：Anger response rate for continuum [from left to right, neutral-anger]

3-Q-1

3-Q-1　話者年齢埋め込みベクトルを用いた若年話者判別

Identification of young speakers using speaker age embedded vector

☆奥本佑哉，西村竜一(和歌山大)

◆発話が「大人っぽい声」か「子供っぽい声」かの自動判別を検討

◆クラウドソーシングで収集した年齢・性別ラベル付き実環境発話を対象とした評価

- 学習済みモデルを用いて音声から x-vector を抽出
- 大人と若年者の境界年齢として年齢閾値を設定し，大人の男性・大人の女性・若年者の３クラスにラベリング
- x-vector を入力としてラベルを予測する全結合 NN を構築
- 年齢閾値を 9~18 歳に設定して交差検証

◆先行研究の GMM-HMM 及びスペクトログラムを入力とした CNN, LSTM による３クラス分類結果と比較（Fig. 1）

◆実験結果を踏まえ，話者年齢埋め込みベクトルの導入を提案

- 話者年齢を推定するようにニューラルネットワークを学習
- 学習したネットワークを用いて発話音声から話者年齢埋め込みベクトルを抽出

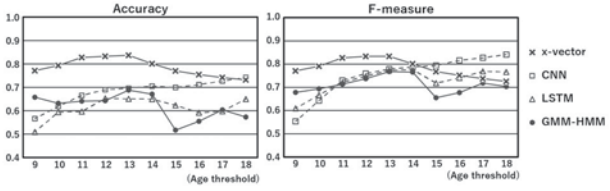


Fig.1：Accuracy and F-measure for each age threshold

3-Q-2

3-Q-2　End-to-End 音声認識の継続学習における部分パラメータ更新による破滅的忘却の防止

Preventing Catastrophic Forgetting by Partial Fine-tuning for Continual Learning of End-to-End ASR

◎高島悠樹，堀口翔太(日立)，
渡部晋治(CMU/JHU)，Paola Garcia(JHU)，川口洋平(日立)

◆実環境へのドメイン適応のために，運用時に実環境データの蓄積とそれを用いたドメイン適応を繰り返すことでモデルを徐々に高精度化していくことができる。

◆しかし，従来のドメイン適応は目標ドメインでの精度向上を目的としており，元ドメインでの精度劣化を招く。この破滅的忘却を防ぐための従来手法はメモリコストが大きい傾向がある。

◆そこで単純で効果的な手法として，一部のパラメータのみを更新することによる end-to-end 音声認識モデルの継続学習を検討する。

◆Transformer に基づくモデルと recurrent neural network transducer (RNN-T)を用いた実験により，エンコーダのパラメータのみを更新しその他を固定することで破滅的忘却を防ぎながら，ファインチューニングと同等の適応性能が得られることが明らかになった。

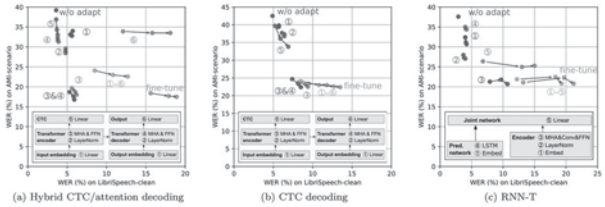


Fig.1：WERs (%) with different adapting modules.

3-Q-3

3-Q-3　器質性構音障害者向け音声認識モデルにおける発話辞書適応方式の比較検討

A comparison of adaptation methods of a pronunciation dictionary 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，音素列を単語列に変換する発話辞書を構音障害者の発音スタイルに合わせて適応させることにより，誤りを軽減することを検討する．

◆初めに，各話者の音素認識モデルを学習し，誤認識率の高い音素に関して発話辞書へ発音パターンを追加することで適応辞書を作成する．

◆適応辞書を用いて単語認識タスクによる評価を行った結果，話者や条件により差異はあるものの，手法の有効性が確認された．

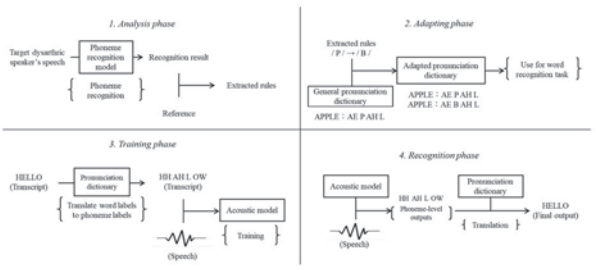


Fig.1：Overview of our proposed method.

3-Q-4

3-Q-4　wav2vec 2.0 によるラベル無し音声を用いた脳性麻痺患者の音声認識

Speech recognition for cerebral palsy patients using unlabeled speech on 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，音声認識を用いたアシスタントシステムが期待されるが，構音障害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，読み上げ発話(ラベル付き音声)の収録は患者にとって負担が大きく，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自由発話音声に着目し，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，脳性麻痺患者のラベル無し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ことで，脳性麻痺患者の音声認識において音素誤り率の性能向上を確認した．

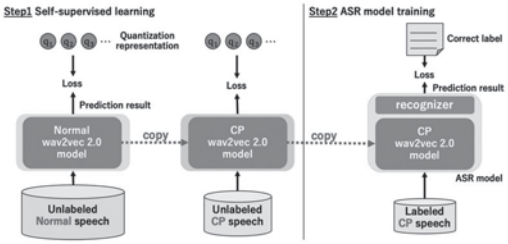


Fig.1：Training procedure using wav2vec 2.0 for cerebral palsy (CP) speech recognition.

3-Q-5

3-Q-5 Non-autoregressive 型音声認識モデルにおける音声合成を用いた話題適応の検討

Text-only domain adaptation using speech synthesis in a non-autoregressive ASR model

◎佐藤裕明 (NHK), △三島剛 (NES), △河合吉彦, △望月貴裕 (NHK)

◆筆者らは，intermediate-CTC 音声認識モデルに対し，テキストをモデルの中間層が出力する潜在ベクトルに変換することで，テキストのみで intermediate-CTC モデルを学習可能にする話題適応手法を提案した．この手法では，テキストから発話長を推定する必要がある．

◆本稿では，学習済みのモデルが出力する実際の CTC 記号列に則り，テキストから統計的に妥当なCTC 記号列を生成することで発話長を推定する方法 (a) と，音声合成手法 Fastspeech で推定する方法 (b) で比較実験を行った．

◆結果，方法 (a) の方が話題適応の認識精度が上回った．本話題適応手法においては発話長を事前に推定した後，記号列から潜在ベクトルに変換した方が高い精度が出るという結果となった．

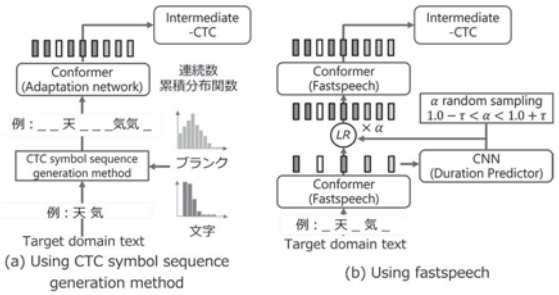


Fig.1 : Duration length estimation methods.

3-Q-6

3-Q-6　学習データの拡張による聴覚障害者音声認識の検討

Automatic Speech Recognition for Deaf Speech Based on Data Augmentation

○ 小林 彰夫，安 啓一 (筑波技大)

◇ 聴覚障害者の音声認識
- 聴覚障害者のうち，音声による意思疎通を行う者は多い (50%以上)
- 障害者がコミュニケーションツールとして音声認識を使う可能性
  → 発話が不明瞭で認識性能が劣化
  → 声質変換でデータ拡張を行い，認識性能改善を図る

◇ 声質変換に基づくデータ拡張
- Voice-Transformer[Kameoka, 2020] によるパラレルコーパス (ATR バランス文) を用いた声質変換
- 声質変換データ：聴者 5/10 名 (JNAS)，聴覚障害者 4 名（B,C セットを除くバランス文）

◇ 実験
- 学習データ：聴者（ASJ/JNAS）＋声質変換による拡張データ
- 評価データ：聴覚障害者 4 名（クローズド，C セット），5 名（オープン，C セット）
- モデル：Conformer-Transducer を用いた音素認識による評価

◇ 実験結果
- 話者クローズドな評価データに対してはデータ拡張の効果あり
- 話者オープンな評価データに対してはデータ拡張の効果はみられず，既存のモデルとの混合を行う必要がある

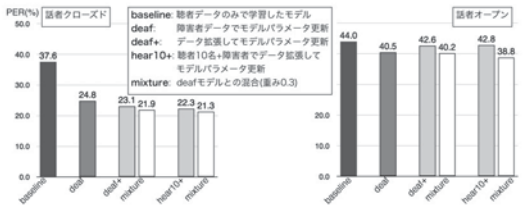


図 1: 実験結果（音素誤り率）

3-Q-7

3-Q-7　CTC 音声認識モデルにおける中間層ロスと条件付けが与える影響の考察

Investigating Effects of Intermediate Conditioning for CTC-based ASR

○市村 収太, 中込 優, 小松 達也, 藤田 雄介, 木田 祐介(LINE)

◆非自己回帰型音声認識である Self-Conditioned CTC のモデルの最適な構成について調査及び考察を行った。

◆中間層の予測結果を条件付けする回数を増やすことにより、条件付き独立性の仮定を緩和し、エンコーダモデル内に学習データの言語知識を獲得できる可能性を示した。また、Greedy デコードで高い性能を得られることを示した。

◆日本語話し言葉コーパス(CSJ)評価セットにおいて、自己回帰型音声認識である Conformer-Transducer モデルを上回る性能を示した。

◆eval1、eval2、eval3 においてそれぞれ、Greedy デコードで Character Error Rates **4.0%**、**3.0%**、**3.4%**を達成した。

Table 1　Character Error Rates(%) on CSJ

Method	Decoder	Conditioning Indices $\mathcal{N}$	#Params	Testset eval1	eval2	eval3
Conformer CTC	greedy	None	120M	4.95	3.65	4.10
+ 6-gram LM fusion	beam search	None	120M	4.70	3.51	3.92
Self-Cond. Conformer CTC	greedy	{6, 12}	123M	4.21	3.19	3.45
	greedy	{3, 6, ..., 15}	123M	4.17	2.98	3.42
	greedy	{2, 4, ..., 16}	123M	4.15	3.04	**3.35**
	greedy	{1, 2, ..., 17}	123M	4.06	2.97	3.43
+ speed perturbation	greedy	{1, 2, ..., 17}	123M	3.97	3.00	3.36
+ 6-gram LM fusion	beam search	{1, 2, ..., 17}	123M	**3.93**	**2.96**	3.37
Conformer	Transducer [3]	None	120M	4.1	3.2	3.5

3-Q-8

3-Q-8　An experimental study of E2E speech recognition with speech invariant structure

☆Runqiang WANG, Daisuke SAITO, Nobuaki MINEMATSU (UTokyo)

◆With the rapid development of artificial intelligence and machine learning, end-to-end speech recognition become more and more popular. However, end-to-end models are always requiring large amounts of data and spend a lot of time in training.

◆We aim to reduce the training time by combining the end-to-end model with invariant structure, which is invariant to any invertible transformations.

◆In this work, we choose Transformer model as the baseline, and we change the encoder part by replacing the attention module to invariant structure module.

◆We compare performance (Fig. 1) and training time (Fig. 2) between baseline and our proposed model.

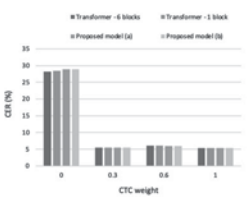


Fig. 1 : Comparison of different models in character error rate (CER).

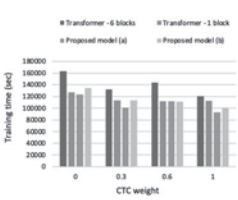


Fig. 2 : Comparison of different models in training time.

3-Q-9

3-Q-9　話し言葉音声認識における主観的評価と客観的評価の分析

Relationship between objective and subjective evaluation for spontaneous speech recognition

☆加藤拓，山縣将貴，石川澄，中島悠輔，菊入圭，吉村健（NTT ドコモ）

◆音声認識の客観的評価指標として，一般的に WER や CER が用いられるが，主観的な感覚と乖離する場合がある

◆会議等の話し言葉音声を用い，文意の捉えやすさの観点で主観的評価と客観的評価の相関を分析し，正解文におけるフィラー等の扱い方及び，より良い客観的評価指標について検討した

◆主観的評価実験として，認識結果のみを見た実験Ａと，認識結果と正解文を比較した実験Ｂを実施した

◆文意を捉える上では，フィラー等を削除した正解文を用いると，主観的評価との相関が弱まることを示した

◆WER よりも CER の方が，主観的評価との相関が強く，内容語文字誤り率（ContentCER）が最も文意の捉えやすさを反映していることを示した

Table 1 各指標におけるスコアと相関係数

指標	正解文内のフィラーの有無	スコア[%]	相関係数	
			実験Ａ	実験Ｂ
WER	なし	16.69	0.401	0.525
	あり	11.74	0.425	0.605
CER	なし	13.01	0.410	0.554
	あり	6.82	0.453	0.676
ContentWER	-	14.78	0.481	0.613
ContentCER	-	8.00	**0.506**	**0.678**

3-Q-10

3-Q-10　Investigation on sub-character tokenization for RNN-Transducer

A fully end-to-end Mandarin speech recognition system

〇沈 鵬，Lu Xugang，河井 恒（NICT）

◆We focus on building a fully end-to-end ASR system that can use pronunciation information-based modeling units.

◆Firstly, a pronunciation-aware unique character encoding method is used for encoding the Chinese character into a unique code.

◆Then, tokens are extracted based on the proposed encoding for building RNN Transducer-base ASR systems.

◆Experiments on the Aishell-1 Mandarin dataset showed the effectiveness of the proposed method.

Encoding type	Example	
Word	语音	
Character	语	音
Syllable w/o tone	yu	yin
Syllable w/ tone	yu3	yin1
Sub-char code	yu3#0	yin1#1

"语音" means "speech"

Fig.1：Encoding types of a Chinese word.

Modeling units	Dev	Test
char(4232)	6.09	6.49
char-word(5000)	5.90	6.48
char-word(8000)	6.30	6.98
Proposed (43)	5.71	**6.25**
Proposed (100)	**5.70**	6.31
Proposed (300)	5.83	6.29

Fig.2：Experimental results (CER %) of the baseline and proposed systems.

3-Q-11

3-Q-11　実音声を用いた音声言語獲得エージェントの評価

Evaluation of spoken language acquiring agent using real voices

☆小松亮太（東工大），岡本拓磨（NICT），篠崎隆宏（東工大）

◆人間と共生するロボットの実現には環境の観察や様々な人との対話を通じて自律的に音声言語を学習する能力が必要不可欠である．

◆先行研究では，教師ラベルを用いずに音声単語を学習するエージェントが提案されている．しかし，音声画像接地に基づくエージェントに関する実験ではノイズを加えた合成音声が用いられていた．そのため，話速や声の高さが異なる複数の話者との対話において同様な学習が行えるのか不明であった．

◆本講演では，31 名から新たに実音声を収録し，音声言語獲得エージェントの学習アルゴリズムの有効性を検証する．

◆Table 1 は 6 名の被験者が音声対話を主観評価した結果である．
食べ物タスクにおいて，1,000,000 対話学習したエージェントは 0.8 の確率で適切な応答を返しており，MOS は 4.33 であった．
6 名の被験者の音声は学習データに含まれていないことから，エージェントの話者性に対する頑健性を確認した．

Table 1：Mean opinion score and accuracy of the agents after 50,000 and 1,000,000 dialogues.

Dialogues	Mean opinion score		accuracy	
	Trial 1	Trial 2	Trial 1	Trial 2
50,000	1.67	1.67	0.233	0.200
1,000,000	4.33	4.33	0.800	0.800

3-Q-12

群論を用いた解析的声道長正規化処理と音声認識への応用

Analytical Vocal Tract Length Normalization by Group Theory and Application to Automatic Speech Recognition

〇宮下敦志，戸田智基（名古屋大）

音声認識システムには、話者の違いによる発声の揺らぎに対して予測が不変であることが求められる。声道長変換はそのような揺らぎを模倣する変換の１つであり、声道長正規化は入力データから揺らぎを取り除く手法である。本報告では、全域通過フィルタによるワーピングで表される声道長変換について、群論を用いて別の変換式を与え、解析的な正規化処理 VTLN-G を導く。これにより、従来の声道長正規化手法よりも厳密な不変性の取り扱いが可能となる。提案手法の有効性を検証するため、TIMIT データセットを用いた実験的評価を行った。その結果から、提案手法による入力データの正規化が音素識別モデルの汎化性能を向上させることを確認した。

Tab.1：Accuracy for phoneme discrimination task

Method	TIMIT	TIMIT-VTL
DNN	$\mathbf{0.571} \pm 0.002$	0.520 ± 0.002
CNN	0.565 ± 0.002	0.520 ± 0.002
VTL-CNN	0.546 ± 0.002	0.512 ± 0.003
VTLN-G+DNN	0.560 ± 0.002	$\mathbf{0.551} \pm 0.002$
VTLN-G+CNN	0.542 ± 0.002	0.534 ± 0.002

3-Q-13

3-Q-13 マルチモーダル対話システムにおける対話破綻時のユーザの個人差

Analysis of Individual Differences in User Response to Dialogue Breakdown in Multimodal Dialogue System

☆坪倉和哉，入部百合絵（愛県大），北岡教英（豊橋技科大）

◆現状の対話システムは，ユーザに対して不適切な発話をしてしまう「対話破綻」が生じているため，言語・非言語情報を用いて対話破綻を検出する技術が求められている．

◆対話破綻時のユーザの反応には個人差が存在すると考えられる．本研究では，個人差の中でもユーザの内的状態に関する性格特性に着目し，**対話破綻時の非言語特徴（韻律情報と顔表情）と個人差（性格特性）との関係を分析**した．

◆性格特性の分類には BigFive を利用した．各 BigFive 性格特性の得点を上位群と下位群に大別し，その２つの群間でマン＝ホイットニーの U 検定を行った．その結果，**協調性，神経症傾向，開放性について，破綻時の非言語特徴について有意差が認められた**（Table.1）．

◆今後は，個人差を考慮した対話破綻検出手法の提案や対話破綻からの回復方法の検討を行う．

Table.1：Results of Mann-Whitney U test.
(+: p < 0.1, *: p < 0.05, **: p < 0.01, ***: p < 0.001)

BigFive	Features	P-value
Agreeableness	RMSenergy	0.0614 +
Neuroticism	AU02 (Outer Brow Raiser)	0.0004 ***
Openness to experience	AU06 (Cheek Raiser)	0.0891 +
	AU12 (Lip Corner Puller)	0.0152 *

3-Q-14

3-Q-14　音声認識率を報酬とした深層強化学習による特徴量変換法

Feature transformation method using of deep reinforcement training by speech recognition

◎石田涼（阪工大・情），△鈴木基之（阪工大）

◆近年では音声認識システムの精度が向上しつつあるが、雑音や激しい訛りが混じる音声に対しては認識精度が低い。

◆このような問題に対し世間ではモデルの改良などで対処されるが、モデルの種類ごとに最適な改良を必要とするため手間と費用がかかる。

◆本研究では音声認識システムではなく音声の特徴量を変換するシステムを深層強化学習によって作成する。

◆深層強化学習が必要とする報酬には「音声認識精度を与える」方法と「音声認識結果に基づいた音素フレームの正誤情報を与える」方法の２種を提案し、実験を行った。

◆結果として学習データの認識精度が低いものが含まれる方が認識精度を維持されやすくなり、音声認識結果に基づいた正誤情報を報酬とした場合に認識精度が低下しやすくなった。

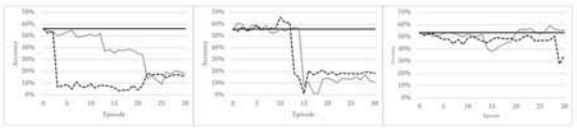


Figure 1　学習データによる認識精度の違い

Orijinal　Accuracy　Frame

3-Q-15

3-Q-15　音声と顔画像を同時に用いたマルチモーダル年齢推定

Multi-modal age estimation using face and speech information

〇俵直弘，小川厚徳，北岸佑樹(NTT)

◆ 音声と顔画像の両方を用いて年齢を推定するマルチモーダル年齢推定モデルとその学習法を提案(図1)

1. 顔画像と音声についてそれぞれ年齢推定モデルを個別に学習
2. 各モデルの中間特徴量を入力としたマルチモーダル年齢推定モデルを学習

◆ AgeVoxCeleb を用いたマルチモーダル年齢推定実験において、提案手法は音声または顔画像のどちらか一方のみを用いたモデルよりも高い年齢推定精度を達成(表1，図２)

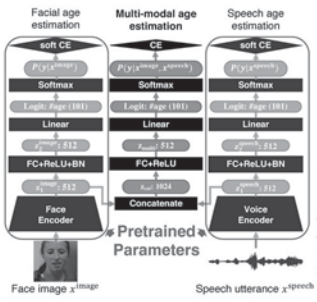


Fig. 1 Proposed multimodal age estimation model

Tbl. 1 Experimental result of multimodal age estimation

Modality	Mean absolute error
Speech	8.09
Image	5.23
Multi-modal (proposed)	**4.78**

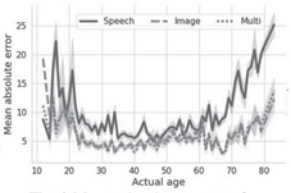


Fig. 2 Mean absolute errors of each model in each age

3-Q-16

3-Q-16　話者情報を利用したマルチタスク学習に基づく叫び声の強度推定

Intensity estimation of shout based on multi-task learning using speaker information

☆石田 泰都，福森 隆寛，山下 洋一（立命館大・情報理工）

■ **【背景】**

叫び声を用いた監視システムを実運用する際には，より強く叫ばれている音声を優先的に検出できることが望ましい．先行研究では，叫びの度合い（叫び声の強度）の推定タスクが検討されている．

■ **【概要】**

本稿では，叫び声の強度推定における推定精度の向上を目的として，強度推定と性別分類のマルチタスク学習を提案した．提案手法は，性別分類の補助タスクによって性別間のデータの特徴の違いをモデルが捉え，強度推定の精度が向上することを期待する．

■ **【結果】**

従来手法・提案手法で，３種類の入力特徴量（スペクトログラム，ケプストログラム，スペクトログラムとケプストログラム両方）ごとにネットワークを構築し評価実験を行った．実験の結果（Table 1），マルチタスク学習に基づく提案手法は，強度推定タスクのみで学習する従来手法に比べ，RMSE が最大 0.04 ポイント改善した．

Table 1：Experimental results

Speech features	Method	
	Conventinal	Proposed
Spectrogram	0.86	0.83
Cepstrogram	0.75	0.74
Spectrogram + Cepstrogram	0.85	0.81

3-Q-17

3-Q-17　大規模事前学習モデルを用いたマルチモーダル感情認識

Multimodal emotion recognition using large-scale pre-trained models

○安藤厚志，高島瑛彦，増村亮，鈴木聡志，牧島直輝(NTT)

◆画像，音響，言語の 3 つのモダリティを用いて発話から人物の感情認識を行う，マルチモーダル感情認識に取り組む．

◆従来のマルチモーダル感情認識ではヒューリスティック特徴量や行動推定モデルの出力，単語の分散表現が入力特徴量に用いられており，近年着目されている大規模事前学習モデルの利用例は少ない．

◆本研究では，マルチモーダル及びユニモーダル感情認識において，大規模事前学習モデルに基づく特徴量の有効性検証を行う．

◆感情認識モデルとして，ユニモーダルエンコーダとゲート付きデコーダで構成されるモデルを用いる．大規模事前学習モデルには，画像モダリティでは CLIP ViT，音響モダリティでは WavLM，言語モダリティでは BERT に基づく Transformer モデルを用いる．

◆マルチラベル感情認識実験の結果，大規模事前学習モデルはユニモーダル及びマルチモーダルの双方で有効性があること，言語モダリティに比べて画像及び音響モーダルで事前学習モデルの効果が高いことが明らかとなった．

Table 1: Unimodal / multimodal emotion recognition performances.

	modality			Hap.		Sad		Ang.		Sur.		Dis.		Fea.		Average	
	V	A	L	WA	F1	WA	F1	WA	F1	WA	F1	WA	F1	WA	F1	WA	F1
Conv.	✓			.673	.673	.673	.689	.699	.701	.675	.739	.696	.721	.747	.795	.694	.720
		✓		.608	.608	.663	.679	.560	.595	**.854**	.845	.631	.675	**.881**	**.867**	.700	.711
			✓	.654	.648	.675	.688	.704	.718	.660	.730	.755	.778	.783	.821	.705	.731
	✓	✓	✓	.706	.706	.673	.694	.729	.734	.742	.788	.799	.808	.789	.825	.740	.759
P/T Model	✓			.681	.680	.671	.684	.729	.725	.782	.807	.765	.778	.828	.848	.743	.754
		✓		.656	.655	**.704**	**.710**	.744	.742	.799	.823	.804	.810	.772	.814	.747	.759
			✓	.645	.645	.670	.685	.720	.728	.751	.795	.793	.808	.792	.827	.728	.748
	✓	✓	✓	**.720**	**.719**	.684	.701	**.760**	**.758**	**.854**	**.855**	**.829**	**.831**	.859	.865	**.784**	**.788**

3-Q-18

3-Q-18　BiLSTM-CTC モデルを使用した自発的な笑い声と叫び声の End-to-End 検出モデルの構築

BiLSTM-CTC End-to-End model for automatic detection of spontaneous laughter and scream

☆松田匠翔，有本泰子(千葉工大)

◆FBANK，ComParE 特徴量，両方をそれぞれ入力，笑い声，叫び声を出力とした BiLSTM-CTC モデルを構築し，検出実験を行い比較した。

◆Corpus-closed の実験では笑い声，叫び声の検出性能の検証を行い，Corpus-open の実験では他コーパスに対する汎用性の検証を行う。

◆Corpus-closed では F-measure が 73%, 60%程度(笑い声，叫び声)を示した。笑い声・叫び声検出の先行研究より精度 30%-40%程度向上。

◆Corpus-open では Corpus-closed より精度が大幅低下。ComParE が 62.97%(笑い声)と比較的高く，韻律情報が笑い声検出に有効の可能性。

◆一方で，叫び声の F-measure が 20%未満と低く，本研究の入力に叫び声に有効な音響的特徴量が含まれていなかった可能性がある。

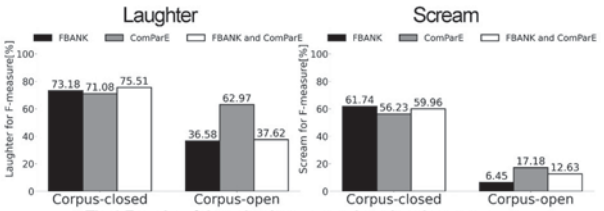


Fig.1 Results of detection by corpus-closed and corpus-open

3-Q-19

3-Q-19　Sequence-to-sequence歌声合成のための発声タイミングのモデル化に関する検討

A study on vocal timing modeling for sequence-to-sequence singing voice synthesis

☆西原美玖，法野行哉，橋本佳，南角吉彦，　徳田恵一 (名工大)

- 本稿では，様々な音素境界の推定手法による楽譜特徴量のフレーム化を行い，フレーム駆動型seq2seq歌声合成におけるアテンション機構に基づく時間構造の制御に関する振る舞いを検討し，どのように発声タイミングがモデル化されているか明らかにする．
- 実験結果から，学習時に使用する音素境界による発声タイミングが実際の歌唱時と比較して遅延が少なくかつ安定したタイミングであり，合成時は実際の発声タイミングとのずれを考慮して柔軟に推定された音素境界を使用することで，発声タイミングを適切にモデル化可能であることを示した．

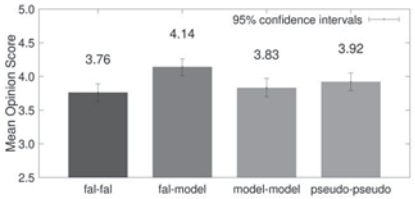


Fig. 1: Mean opinion scores.

3-Q-20

3-Q-20　教師なし事前学習を用いた対話エージェントの誤反応防止技術

Self-Supervised Pretraining for Device Directed Speech Detection

◎佐藤宏，芦原孝典，安藤厚志，森谷崇史(NTT)

◆概要

- 音声対話エージェントに入力された音声が，ユーザーが意図して機械に話しかけた音声か否かを識別する**応答義務推定問題**を検討
- DNN ベースの応答義務推定モデルの構築には，大規模なペアデータが必要であり，アノテーションコストが高かった
- 本稿では、ペアデータの限られた条件で有効性が報告されている**教師なし事前学習**を応答義務推定に適応し，有効性を検証した

◆評価実験

- 教師なし事前学習を用いた手法は，音響特徴量を使用する従来法に対して性能向上を実現
- 本タスクでは日本語データで学習した HuBERT [1]モデルが最も効果的であった

事前学習モデル/特徴抽出器	事前学習データ	パラメータ数	EER [%] (↓)	ROCAUC [%] (↑)
FBANK	-	1.2 M	3.9	99.15
HuBERT Base	LS 960 hr	95.8 M	2.9	99.34
Wav2Vec 2.0 Base	LS 960 hr	96.2 M	3.7	99.35
WavLM Base	LS 960 hr	95.6 M	3.3	99.11
HuBERT Base	CSJ 520 hr	95.8 M	**2.7**	**99.55**
Wav2Vec 2.0 Base	CSJ 520 hr	96.2 M	4.5	99.12

[1] W-N. Hsu, B. Bolte, Y-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, "Hubert: Self-supervised speech representation learning by masked pre-diction of hidden units," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 3451–3460, 2021.

3-Q-21

3-Q-21　Pause prediction using BERT-based features for long-form text-to-speech synthesis

○ † Dong Yang, ‡ Tomoki Koriyama, †Hiroshi Saruwatari
(† The University of Tokyo, ‡ CyberAgent, Inc.)

◆Silence pauses always appear in the natural sounding speech. In text-to-speech (TTS) systems, pause prediction is an essential part that can influence naturalness and intelligibility significantly. Pauses appear at punctuations which we call punctuation pauses. Besides pauses are also inserted between the words to utter long sentences fluently, which we call “respiratory pauses”. We focus on the prediction of the respiratory pauses.

◆We explored whether BERT can accomplish pause prediction in a large-scale multi-speaker English corpus, and illustrated the performance of this transformer-based model compared with the conventional LSTM-based one.

◆According to the precision-recall curve in the experimental results, our proposed model performs better than conventional method.

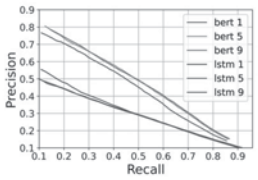


Fig.1 Precision-recall curve

3-Q-22

3-Q-22　Encoder-Decoder 型音声合成におけるピッチ同期分析適用の検討

An investigation on applying pitch-synchronous analysis to Encoder-Decoder speech synthesis

◎蛭田宜樹，田村正統（東芝デジタルソリューションズ）

◆Encoder-Decoder 型音声合成モデルである FastSpeech2 に，音声波形のピッチ周期に基づいて分析・合成するピッチ同期分析を適用する方法を検討した．

◆評価実験では，提案法は従来法の音質スコアを上回らなかったものの，FastSpeech2 と WaveNet の組み合わせ(FS2+WN)よりも合成速度や韻律の制御性について上回る結果が得られた．

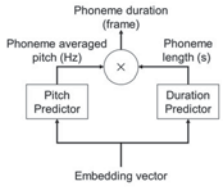


Figure 1 Proposed method for phoneme duration.

Table 1 The evaluation result of calculation time (s).

Method	RTF
Proposed(CPU)	**0.28**
FS2+WN(CPU)	332.88

Table 2 The evaluation result of prosody controllability (Hz).

Method	Male	Female
Proposed	**8.22**	**6.91**
FS2+WN	9.18	9.09

3-Q-23

3-Q-23　アクセント辞書を適用可能なラベル推定 Transformer とアクセント結合 Transformer を用いた音声合成言語処理部の開発

Text analysis for speech synthesis using a transformer for estimating phonetic and prosodic features that can apply an accent dictionary combined with an accent sandhi transformer

◎栗原清（NHK）

◆音声合成における言語処理部について、アクセント辞書を適用できる Transformer を用いた言語処理部を開発したため報告する

◆音声合成の言語処理部は、近年 Deep Neural Network(DNN)による手法が登場し始めた。ただし、これらの手法は DNN を用いるため一対一の変換はできるが、単語単位の辞書変換はできなかった

◆また、Transformer を用いた手法は、予備実験より固有名詞・数詞・アルファベットを正しく変換する事ができない傾向があった

◆これらの問題を解決するため、Transformer を用いても辞書を適用できる手法を開発した。また、アクセント結合を考慮したアクセント結合 Transformer も合わせて開発した。Fig.1 に全体概要図を示す。

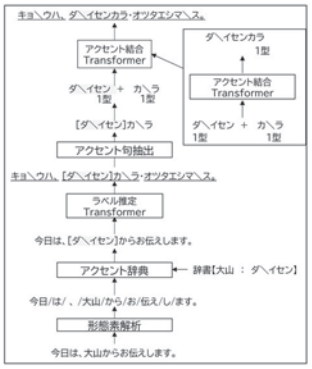


Fig.1：全体概要図

3-Q-24

3-Q-24　笑い声合成における音声記号表現と音響特徴量の感情次元による制御

Generating laughter components and acoustic features with emotion manipulation.

☆木村駿野，森大毅（宇都宮大）

◆笑い声を合成するために必要な笑い声の音声記号表現を、どのように作り出すかが課題となっている。

◆感情次元(快-不快、覚醒-睡眠)と自然笑い声の音声記号表現、および音響特徴量との関係を機械学習によりモデル化する。

◆快-不快次元が、笑い声の長さに関連することが確認された。

◆感情次元が、音声記号表現の種類に影響を与える可能性がある。

◆感情次元により、音声記号表現および音響特徴量を制御することで、合成笑い声から知覚される感情状態を制御できることが示された。

快-不快，覚醒-睡眠	音声記号表現の生成例
中立，中立	hu hu H hu he ha hu hu hu hu
非常に快，非常に覚醒	hu ha a ha H hu h hu u hu ha H hi H hu H hu u H H hi he Na ha H hu H he U hu ha H hi H hu H hu u H

3-Q-25

3-Q-25　微分可能な信号処理に基づく音声合成器を用いたDNN音声パラメータ推定の検討

A study of DNN-based speech parameter estimation with speech synthesizer based on differentiable digital signal processing

☆松永裕太 (LINE 株式会社，東大)，寺島涼，橘健太郎 (LINE 株式会社)

- 微分可能な音声合成器を用いた，高品質な音声を合成可能な音声パラメータを推定する DNN の学習法を提案．

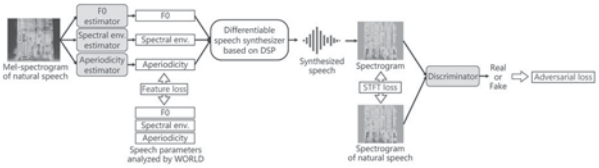


Fig. 1: Overview of proposed method

- 客観評価と主観評価で，提案手法による分析合成音声が WORLD による分析合成音声よりも高品質．
- 敵対的学習の導入によりさらに品質が改善．

	MCD	PESQ	FDSD
Base.(WORLD)	5.273	2.819	0.595
Prop.(STFT)	**5.113**	**3.263**	**0.452**
Prop.(GAN)	5.480	3.171	0.548

Tab. 1: Objective evaluation results

Method A	Method B	A	B	Neutral
Base.(WORLD)	Prop.(STFT)	0.220	**0.330**	0.450
Base.(WORLD)	Prop.(GAN)	0.050	**0.490**	0.460
Prop.(STFT)	Prop.(GAN)	0.090	**0.400**	0.510

Tab. 2: Subjective evaluation results

3-Q-26

3-Q-26　jaCappella コーパス：重唱分離・合成に向けた日本語アカペラ歌唱コーパス

jaCappella corpus: Japanese *a Cappella* Singing Voice Corpus for Chorus Singing Voice Separation and Synthesis

◎中村 友彦，高道 慎之介，丹治 尚子（東大院・情報理工）
深山 覚（産総研），猿渡 洋（東大院・情報理工）

◆日本語のアカペラ歌唱からなる jaCappella コーパスを構築
- 著作権保護期間が終了した童謡・唱歌を，6 声（lead vocal, soprano, alto, tensor, bass, vocal percussion）のアカペラへ編曲．
- ジャズ，パンクロック，ボサノバ，ポップス，レゲエ，演歌アレンジをそれぞれ 5 曲作成．原曲の雰囲気を保った曲も 5 曲作成．
- 全曲に関して楽譜を MusicXML 形式で作成．
- 各声部の録音を標本化周波数 48 kHz の WAVE フォーマットで作成．

◆研究用途無償・商用利用有償での公開に向けて準備中

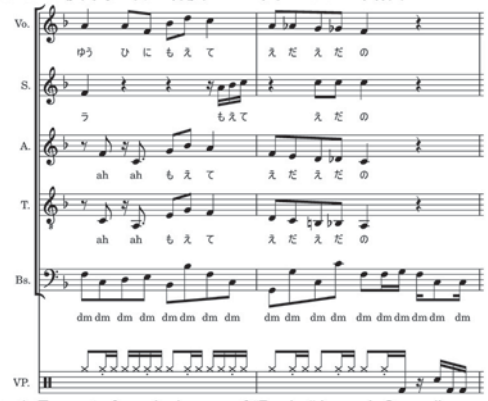


Fig. 1: Excerpt of musical score of "Poplar" in our jaCappella corpus.

3-Q-27

3-Q-27　高齢者向け発話の韻律推定

Estimation of prosodic features of talking style speech for elderly person

○水野秀之（公立諏訪東京理科大学），
中嶋秀治（NTT コミュニケーション科学基礎研究所）

◆高齢者にとって聞き取りやすい話者による高齢者を意識した発話（高齢者向け発話）の韻律的特徴（F_0最大値，F_0最小値，F_0 レンジ，話速）について，同一話者が同一文書を読み上げた発話（読み上げ発話）の韻律的特徴との比較と推定を行い以下の点について確認した．
- 読み上げ発話(Reading)と高齢者向け発話(Elderly)の韻律的特徴の相関係数は全て 0.6 以下と低く相関性が見られない。特に F_0 最小値は，ほぼ無相関な分布となっている（Figure1 参照）
- 韻律的特徴の推定精度（決定係数）は全て 0.4 以下と低く推定困難
- 相関分析より F_0 最小値は任意性が高いと考え，高齢者向け発話の F_0 最小値も条件として加え F_0 レンジの推定を行った結果，決定係数が 0.75 となり推定精度が大幅に向上　（Figure 2 参照）

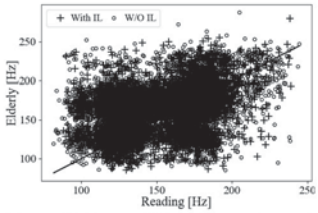


Figure.1 Relation of F0 minimum of accent phrase with or without important label (IL)

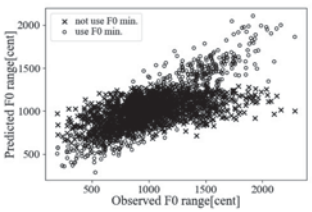


Figure.2 Comparison of prediction result

3-Q-28

3-Q-28　深層ガウス過程音声合成における畳み込み・self-attention・リカレント構造の評価

Evaluation of convolution, self-attention, and recurrent structures in deep Gaussian process speech synthesis

☆中村 泰貴（東大院・情報理工），
郡山 知樹（東大院・情報理工/サイバーエージェント），
猿渡 洋（東大院・情報理工）

◆以前の報告で，ガウス過程回帰に基づく self-attention 構造を有する Sequence-to-Sequence (Seq2Seq) 深層ガウス過程音声合成を提案
◆本稿では，畳み込みガウス過程回帰に基づく畳み込み構造 (CONVGPR) を導入した Seq2Seq 深層ガウス過程音声合成を提案
◆さらに Seq2Seq 深層ガウス過程音声合成において最適な構造（畳み込み・self-attention・リカレント）を評価

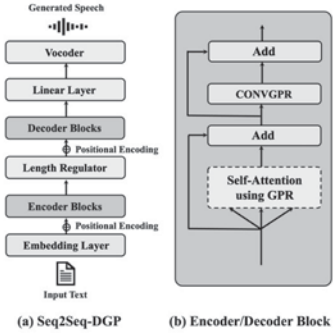


Fig.1 : Framework of Seq2Seq DGP-based speech synthesis with self-attention and convolution structure. [ ] is replaced by a recurrent structure, SRU, in Decoder Blocks.

3-Q-29

3-Q-29　BERT を用いた雑談対話音声からの認知症疑い検出

Dementia Detection with BERT from Chat Dialogue

☆岡田智哉，入部百合絵(愛県大)，北岡教英(豊橋技科大)

◆発話音声から軽度認知症を検出する研究は，言語情報を人手による書き出し文から取得していることが多く，実用的ではない。

◆本研究では，音声認識器から得たテキスト文（自動書き起こし文と呼ぶ）を用いて軽度認知症疑いを検出することを目的とする。そこで，BERT(Bidirectional Encoder Representations from Transformers)を用いた検出精度を調査するとともに，BERT を特徴量抽出器として利用し，音響的特徴量と組み合わせた検出手法を提案する。

◆音響的特徴量のみを用いた場合と，３種類(自動書き起こし文，自動書き起こし文から記号（句読点他）削除，方言修正)のテキスト文を入力とした BERT 出力層前の全結合層の値（BERT 特徴量）と音響的特徴量と組み合わせた結果を Table1 に示す。

◆BERT 特徴量と音響的特徴量を入力とした５つの識別器を用いて比較したが，１つの識別器を除いて，自動書き起こし文からの BERT 特徴量と音響的特徴量を組み合わせた場合が最も高い結果を得た。この結果から，認識誤りを含むテキスト文であっても BERT 特徴量と音響特徴を組み合わせることで識別精度が向上することが明らかとなった。

Table1 Precision, Recall and F-measure with LGR(N-CG)

データ名	適合率	再現率	F 値
音響特徴のみ	0.6538	0.6800	0.6666
音響特徴 + 自動書き起こし文	0.7727	0.6800	0.7234
音響特徴 + 記号削除	0.6521	0.6000	0.6250
音響特徴 + 方言修正	0.6071	0.6800	0.6415

3-Q-30

3-Q-30　筋萎縮性側索硬化症者の音声合成を目的としたモデル適応と声質変換の比較評価

Comparative Evaluation of Model Adaptation and Voice Conversion on Speech Synthesis for a Person with Amyotrophic Lateral Sclerosis

☆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、外部雑音を含み、かつ病状の進行により明瞭性が低下した少量の筋萎縮性側索硬化症（ALS）者の音声を用いて音声合成モデルを生成することを考える。

◆雑音の問題に対しては音声強調を、少量データおよび明瞭度の問題に対しては声質変換（VC）と TTS モデルの適応（MA）の２つのアプローチについて検討し、各手法による音声合成品質を比較評価する。

◆学習にノイズ抑制音声を用いたものの、VC による合成音声にはノイズが大きく含まれ聞き取りづらい音声となった一方、MA ではノイズが小さく ALS 者本人の声質に似た明瞭な音声が合成できた。

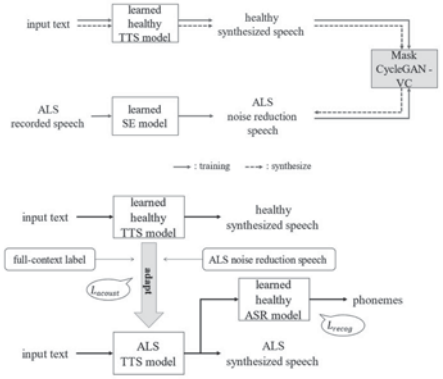


Fig. 1 各手法の概要（上：VC 下：MA）

3-Q-31

3-Q-31　深層学習に基づく歌声合成における歌唱表現のモデル化の分析

Analysis of modeling of singing expression in DNN-based singing voice synthesis

☆高谷 恒輝，能勢 隆，今井 柊平，伊藤 彰則（東北大院･工学研）

◆本研究は非自己回帰構造の歌声合成システム Sinsy をベースとした F0 予測器による，特にビブラートに着目した歌唱表現のモデル化に関する分析を行う.

◆LSTM でビブラートの位相に関連する情報を受け取り，過学習により歌唱表現の詳細な特性を捉えることによって，F0 が過剰平滑化されずにビブラートをモデル化することが可能であると考える.

◆過学習と LSTM によるビブラートのモデル化の分析を行うため，Sinsy の LSTM を CNN で代替したモデルや Sinsy から CNN を除いたモデルを過学習させ，歌唱表現のモデル化性能を分析した.

◆客観評価ではビブラートの周波数特性を示す FT loss が学習の進行につれて低下した．実験から，過学習により LSTM を必要とせずにビブラートのモデル化が可能な一方，LSTM の導入によりモデル化性能が向上することが示された.

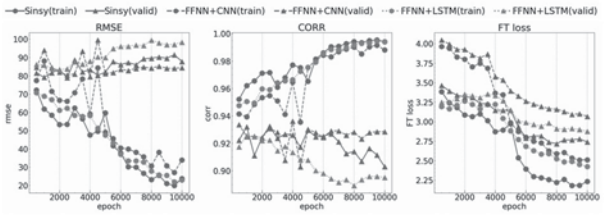


Fig. 1 : Objective Evaluation for each epoch.

3-Q-32

超高齢者コーパスと S-JNAS を用いた高齢者音声の音響的特徴の分析

Acoustic characteristics of the elderlies in the super-elderly speech corpus EARS and S-JNAS

☆福田芽衣子，杉山雅和，西村良太(徳島大)，
入部百合絵(愛知県立大)，山本一公(中部大)，北岡教英(豊橋技科大)

◆高齢者音声コーパスの S-JNAS および超高齢者音声コーパス EARS の話者（60～98 歳）ついて基本周波数（Fo），フォルマント周波数，喉頭雑音，MFCC（0～12 次）などの年齢による推移を調査した.

その結果，男女ともに母音 /i/ の F1 の上昇および F2 の低下，一部の MFCC の平均値と標準偏差，シマの標準偏差の低下が認められた.

しかし，Fo と上記以外のフォルマント周波数，母音中心化の指標，MFCC の大部分およびジッタは男女で異なる動態を示した.

高齢者音声の性差は、調音器官の形態学的な違いや性ホルモン作用に加え，加齢性の変化と個人差が加わり多様な様相を呈すると思われる.

◆ これらの音声の性差を考慮に入れることで，今後，音声認識や疾患判別の精度向上を図れる可能性が考えられた.

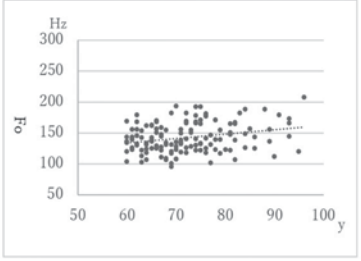


Figure 1: Fundamental frequencies of male speakers

3-Q-33

講演取消

3-Q-34

3-Q-34　プロ声優を対象とした大規模コーパス朗読におけるリテイク数の比較

Comparison of the number of retakes in a large-scale corpus reading by professional voice actors

☆山本泰我，小口純矢，森勢将雅（明治大）

- ◆ プロ声優による大規模コーパスの収録におけるリテイク数で文の読みやすさを検証
- ◆ 全体の約 33%の文章では，リテイク無し
- ◆ 多くの文章で，一方のリテイク数が多いともう一方は少なく，特定の傾向は認められず
- ◆ 両者ともリテイク数は概ねモーラ数に比例

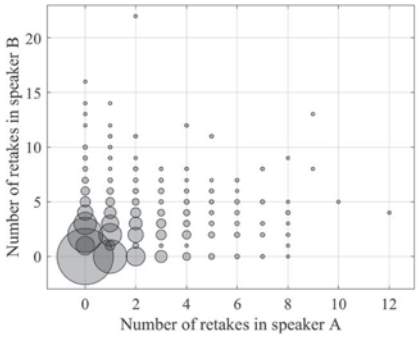


Fig 1: ２名のプロ声優のリテイク数を比較するバブルチャート

3-Q-35

3-Q-35　中国語の舌尖音の鼻音の特徴と発話の正確さの関係

Relationship between features of Chinese velar of nasal sound and utterance accuracy

○星野朱美(富山高専)

◆背景と目的：日本語話者は中国語の「前鼻音」と「後鼻音」の聴き取りと発話が困難な上，自習する際にも自分の発音に対する正確な評価手段がない。CAI のための発話自動学習システムの開発を試みる。

◆解析手法：中国語話者と日本語話者の舌尖音の「前鼻音」と「後鼻音」の発話の対，dan[tan]-dang[taŋ]，den[dən]- deng[təŋ]，と dan[t'an]-dang[t'aŋ]を対象にして，まず，それらの発話のスペクトログラムを用いて，「前鼻音」と「後鼻音」の相違点、有声期間中のパワー、フォルマントを比較することにより，学生の「前鼻音」と「後鼻音」の発話の問題点を分析した。更に有声期間中のパワーを比較することにより，「前鼻音」と「後鼻音」の特徴的パターンを見出した。さらにそれぞれ「前鼻音」と「後鼻音」発話のフォルマントを解析し，正確な発話の特徴付けとしてF1～F3 を抽出した。

◆今後の展望：「前鼻音」と「後鼻音」の自動判別を目的として，各発話のパワーの特徴の自動測定システムを開発し，発話の有声期間中のパワーを自動計測し，さらに今回は抽出した発話の特徴として F1～F３と今までの研究成果を用いて正確な発話の判定基準を確立し，自習教材の開発を目指す。

3-Q-36

3-Q-36　差分メルケプストラムを用いた声質変換による喉締め歌唱音声改善方式の検討

Study of choking singing voice improvement method based on voice-conversion using mel-cepstral differential

☆植田遥人，原直，阿部匡伸(岡山大院･HS 統合科学研)

◆喉締め歌唱音声を複式歌唱音声に変換する方式を検討する.

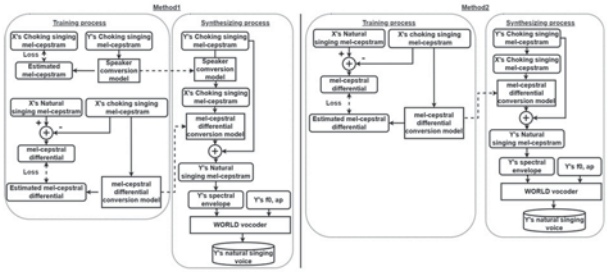
◆本稿では，腹式歌唱から喉締め歌唱の差分メルケプストラムに任意歌唱者の喉締め歌唱メルケプストラムを足し合わせることで，変換を実現する.

- 方式１：元歌唱者の話者性を考慮しないモデル
- 方式２：元歌唱者の話者性を考慮したモデル

◆客観評価実験のメルケプストラム歪みから，方式１の変換後歌唱が腹式歌唱に近づいていることを示した.

◆主観評価実験では方式１の変換後音声についてABXテストを行った.

- 学習に用いていない歌唱者の喉締め歌唱の改善度は，学習に用いた歌唱者に比べ半分程度であった.
- 話者性に関しては維持できているという結果であった.



3-Q-37

3-Q-37　自己教師あり学習による音声特徴表現をクロスリンガル声質変換に適用する実験的検討

An Experimental Study of Applying Self-supervised Speech Representations to Cross-lingual Voice Conversion

☆ Pin-Chieh Hsu, Nobuaki Minematsu, Daisuke Saito (東京大学)

◆Recently, the state-of-the-art cross-lingual voice conversion (CLVC) has been implemented with phonetic posteriorgram (PPG), which is supposed to be a speech representation of spoken contents. The strategy is to extract the spoken linguistic contents from the source speech and then synthesize the converted speech based on the extracted spoken content features and the target speaker's voice characteristics.

◆In this work, we proposed to use self-supervised speech presentation to replace the PPG in the CLVC system. The representation is learned by the wav2vec 2.0 XLSR model, which is trained on a cross-lingual unlabeled speech dataset, and we experimented with it on the Japanese-to-English cross-lingual voice conversion. Furthermore, we also proposed to use two training techniques to improve further the speaker similarity of converted results from the proposed systems.

◆According to the results of objective evaluations, we show that the results of the proposed systems are comparable in terms of quality, demonstrating that self-supervised learning features show great promise for the CLVC field and, in the future, can be further improved on the ability to disentangle speaker information.

3-Q-38

3-Q-38　ポップアウトボイスの個人内変化に伴う声帯振動特性と声道音響特性の変化

Changes in glottic and vocal tract acoustic properties due to intra-speaker variation of "pop-out" voices

○北村 達也(甲南大), 能田 由紀子 (国語研), 榊原 健一 (北海道医療大)
河原 英紀 (和歌山大), 天野 成昭 (愛知淑徳大)

◆ プロのナレーター6 名, 一般の話者 1 名のポップアウトボイス (PV) と通常音声 (NV) の音声と EGG データをスタジオ収録

◆文章音声の EGG データから基本周波数と声門開放率を計算

◆PV と NV とでは基本周波数と声門開放率が異なる (Fig. 1)

◆母音のスペクトル包絡も PV と NV とで大きく異なる

◆ポップアウトボイスは単なる「大きな声」ではない

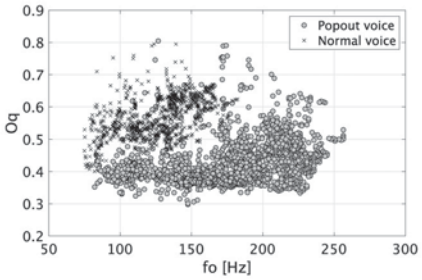


Fig.1 : Voice profile of pitch frequency (fo) and open quotient (Oq) of a professional male speaker.

3-Q-39

3-Q-39　声質変換で目標歌手の声質を再現する歌声合成システムにおける特有の問題点とその解決方法に関する検討

An examination of the specific problem and its solution in the singing voice synthesis system whose output is used for voice conversion system in order to reproduce the target's voice

☆井沼望, 坂野秀樹, 旭健作(名城大院)

◆歌声合成システムに声質変換を適用する時, 歌声ライブラリに女声が多いことから, 目標歌手が男声の場合, 女声から男声への異性間変換を行う機会が多く, 変換品質の低下が起こりやすいと想定される.

◆異性間変換における声道長の差異吸収のため, 学習前に複数の伸縮率でスペクトル包絡の周波数軸伸縮を行った.

◆本実験では, 変換元歌手には, 歌声合成システム NEUTRINO の女声ライブラリ 7 種を, 声質変換システムには GMM 最尤推定による変換を行う sprocket を使用した. 目標歌手は 20 代の一般男性 1 人とした.

◆変換品質の客観評価指標には, K 分割交差検証を行い, メルケプストラム歪み MCD の全評価データにおける平均値を用いた.

◆結果として, 変換前の周波数軸伸縮による変換品質の向上は確認できなかった.

Table.1 : Average MCD when the frequency axis is not contracted and the amount of improvement at each contraction rate [dB]

Contraction rate	1 (No contraction)	0.794	0.630	0.500
Voice Library	Average MCD [dB]	The amount of improvement [dB]		
Itako Tohoku	5.769	0.018	0.074	0.077
JSUT	5.754	-0.034	-0.011	-0.158
Kiritan Tohoku	5.636	-0.050	0.039	0.036
Merrow	5.913	-0.053	-0.010	0.034
No.7	5.795	-0.069	-0.156	-0.291
Yoko	5.672	-0.033	0.080	-0.001
Zunko Tohoku	5.819	0.111	0.113	0.163
Average	5.765	-0.016	0.018	-0.020

3-Q-40

3-Q-40　Robustness of Noisy-to-Noisy Voice Conversion against Variations of Noisy Condition

雑音環境下音声変換の雑音環境種別に対する頑健性

☆Chao Xie, Tomoki Toda (Nagoya University)

We have proposed a novel Noisy-to-Noisy Voice Conversion (N2N VC) framework converting the speaker characteristic of a noisy speech while retaining the background noise. The first "noisy" means no clean data from source/target speakers are necessary for training. The second "noisy" means the noise in the converted samples is controllable. Previous experiments involved speaker-independent noisy conditions for the training set. In this paper, we extended it to more realistic conditions where noise and speaker are correlated and found the performance degraded significantly. We then proposed a noise augmentation strategy to improve the framework for speaker-dependent noisy environments. Subjective evaluations were conducted, and the results show the effectiveness of the augmentation.

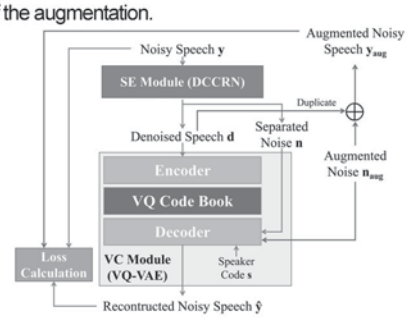


Fig.1 : The overview of the proposed method.

3-Q-41

3-Q-41 ゼロショット学習に基づく声質変換のための雑音に頑健な話者エンコーダ学習

Noise robust training of speaker encoder for voice conversion based on zero-shot learning

☆竹内太法，立蔵洋介（静岡大・創造科学技術大学院）

◆ゼロショット学習に基づく VC では，変換目標の話者を表現するためにいくつかの短い発話を必要とする．

◆雑音が含まれる目標話者の発話から頑健な話者表現を得るための話者エンコーダ構成法を検討する．

◆手法 1 (Enhancement) 音声強調手法で得られる音声をクリーン音声のみで学習された話者エンコーダに入力

◆手法 2 (Noisy) 雑音を含む音声で話者エンコーダを学習

◆2 つの手法とクリーン音声のみから学習された話者エンコーダで得られる話者表現を用いて AutoVC による VC を評価

◆Mel-CD で SNR の高低によって手法 1,2 に優位な違いが現れたが，Clean モデルに対して優位な差がみられなかった．

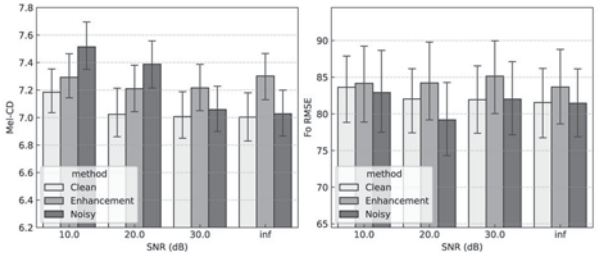


Fig. 1: Mel-cepstral distortion (Mel-CD) and Fo root mean squared error (Fo RMSE) of converted speech for each SNR

3-Q-42

3-Q-42 Fine-Tuning を用いた少量学習データによる電気喉頭音声変換の基礎的検討

Basic study of electric laryngeal speech conversion using small amount of learning data based on Fine-Tuning

☆佐藤元，池田雄介(東京電機大)

◆背景

- 電気喉頭音声の明瞭性や自然性を改善するため、深層学習を用いた自然音声へ変換が試みられている
- 電気喉頭音声データの収集の負担を軽減するため，少量データによる学習が望ましい。

◆提案手法

- 大量の自然音声(JVS コーパス)で学習した ACVAE を少量の電気喉頭音声で Fine-Tuning を行う。

◆実験

- Fine-Tuning により，MCD が 0.2 dB 改善されることを確認。

条件	学習データ量 64 文	32 文	16 文
F-NA	8.38±0.11	8.20±0.10	8.31±0.11
F-SS	7.94±0.08	8.12±0.06	8.42±0.10
N-NA	8.29±0.08	8.21±0.11	8.21±0.10
N-SS	8.13±0.11	8.11±0.12	8.30±0.11

Table 1 MCD of generated and target voices

3-Q-43

3-Q-43 パーキンソン病の重症度に基づく音響的特徴量の分析

Analysis of acoustic features based on severity of Parkinson's disease

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

◆48~75 歳(平均年齢 63.0 歳)のパーキンソン病(PD)患者，計 84 名を対象に音読と自由会話音声を収集して症状の違いを判別し，音響的特徴を明らかにする．

◆重症度の指標である PD 評価スケールの運動機能検査 UPDRSⅢを用いた．UPDRSⅢの点数を閾値に，PD 患者の音声を 2 群に分けて音響的特徴量から 2 クラス分類を試みた．2 値分類モデルの精度の評価指標 MCC（Matthew’s Correlation Coefficient）をもとに分類精度を閾値毎に評価した．その結果，自由会話音声で 16 点の場合に MCC が最も良い結果であった（Fig.1）．

◆上記閾値をもとに，閾値超と閾値以下の 2 クラス間で音響的特徴量の検定を行った結果，NHR（Noise to harmonic ratio）の標準偏差と DFA（Detrended fluctuation analysis）に有意差が認められた（Fig.2）．NHR の標準偏差と DFA は音声障害が強くなるほど値が大きくなる．臨床でも 16 点以下は軽症であると判断されるため，運動障害の初期を判別する上で，NHR の標準偏差と DFA は有効であると考えられる．

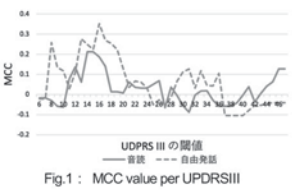


Fig.1： MCC value per UPDRSIII

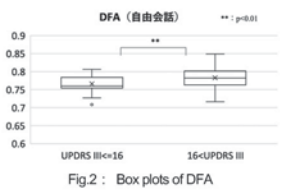


Fig.2： Box plots of DFA