2-1-1

2-1-1　「音空間デザイン」の提案 —楽器を使った音空間デザインの提案と実践—

Proposal of "Sound Space Design"-Proposal and practices of Sound Space Design using musical instrumentals-

☆水口聖菜、石川あゆみ(岐阜高専)

◆音環境デザインの手法の一つに、空間に音を加えるデザインがある。この手法では、単に演出的な音や音楽を空間に付け加えるのではなく、ある空間に対して目指すべき音環境を見つけ出し、その実現に向けて順序建てて作り上げていくことが必要とされているが、その実践例は少ない。

◆空間に新たな魅力を生み出そうとするデザインを「音空間デザイン」、それによって生み出された空間(既存の空間及び音環境に音空間デザインで加えた音を合わせた場)を「音空間」と定義する。

◆本研究では、楽器を使って音空間デザインを行う場合にどのような音風景が生み出され、それによりどのような効果があるのかを明らかにするために実践を行った。

Fig.1：A scene from the trial at school festival

Fig.2：A scene from the practice at ennoie midoribashi

2-1-2

2-1-2　アコーディオンにおける空気の圧縮性を考慮したピッチベンド発生原理の解明

Elucidation of the principle of pitch bending in an accordion considering the compressibility of air.

☆岡田日向(九大・芸工)，西宮康治朗，△武田真和，△長秀雄(青山学院大)，鮫島俊哉(九大・芸工)

◆フリーリード楽器であるアコーディオン（Fig. 1, 2）にはピッチ（音高）を連続的に下げるピッチベンドと呼ばれる奏法が存在する。本研究ではピッチベンド発生原理の解明を目的とする。

◆ピッチベンドの測定結果（Fig. 3）より，ピッチベンドには空気の圧縮性の関与が考えられた。

◆アコーディオンリードの物理モデルによる検証から，リードの発振メカニズムを明らかにした。

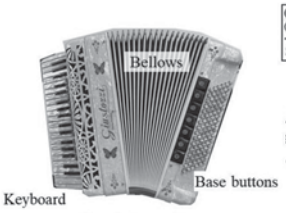


Fig. 1 Accordion

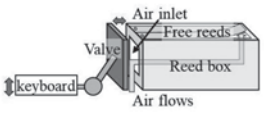


Fig. 2 Structure around accordion reeds.

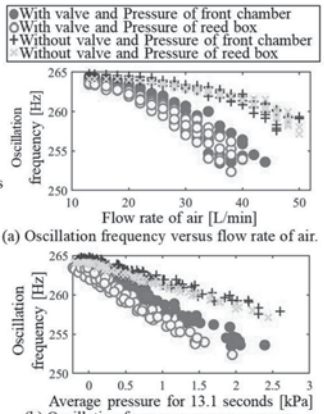


Fig. 3 Pitch bending measurement results.

2-1-3

2-1-3　口笛吹鳴時の声道可視化の検討

A Consideration of vocal tract visualization during whistling

○森　幹男，△菅原真一(福井大院)

◆音高を変化させて口笛吹鳴したときの声道の正中矢状面の画像を用いて，音高変化時の舌の動きについて考察した。

◆低音域（Ⅰ）では音高が高くなると舌の重心は前方へ移動する。

◆中音域（Ⅱ）高音域（Ⅲ）では舌の重心は下方へと移動する。

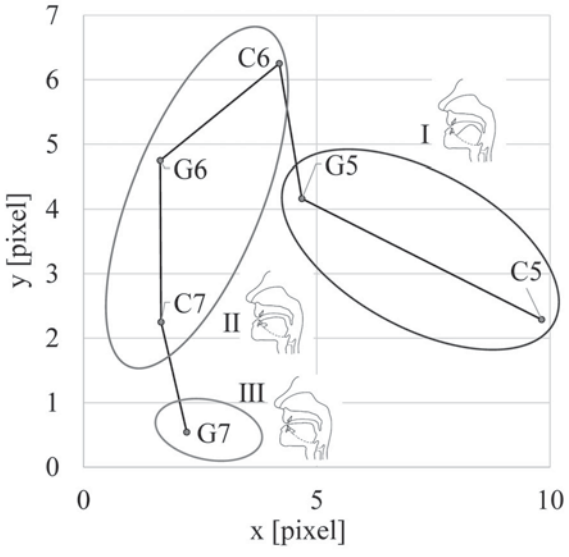


Fig.1：Movement of the tongue's center of gravity

2-1-4

2-1-4　声道模型を用いた口笛の重音奏法の検討

A Consideration on multiphonic whistling by using the vocal tract model

☆川原田美月，森　幹男(福井大院)

◆共振周波数が協和音となる声道形状の口笛声道模型（Type Ⅰ～Ⅵ）を用いて，口笛の重音奏法について検討した。

◆発生音と口唇部における流速との関係を実験的に調べた結果，今回提案した声道形状の口笛声道模型では，重音は発生しなかった。

◆柔らかい素材を用いて声道模型を構成し，ダンピングの影響も考慮することで今回提案した声道形状で重音が発生するか今後検討を行う。

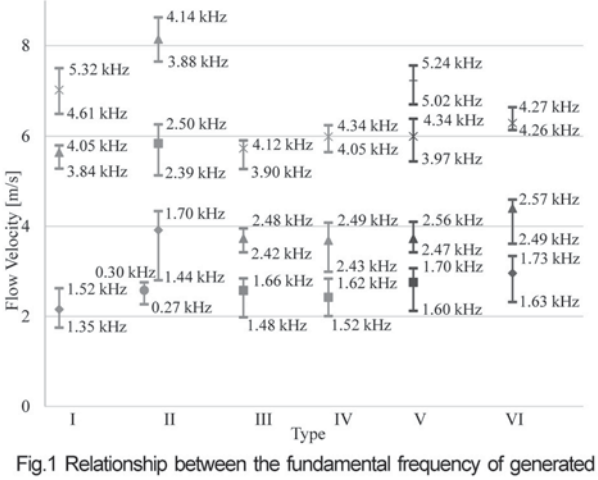


Fig.1 Relationship between the fundamental frequency of generated sound and flow velocity

2-1-5

2-1-5 声楽歌唱の主観的評価と音響特徴の関連性

Relationship between subjective performance ratings and acoustic features of vocal singing

☆近藤はるか(慶應大院・政メ)，川原繁人(慶應大・言文)，△藤井進也(慶應大・環境)

音楽コンクールは音楽家にとって、才能や練習の成果を評価され、キャリアを左右する重要な機会である。コンクールの結果は一般的に審査員の主観的評価によって決まるが、声楽を客観的に評価するため、歌声を定量化する様々な指標が提案されてきた。本研究の目的は、音響指標とプロの声楽家による主観的評価との関係を探求することである。コンサートホールで声楽コンクールを開催し、本研究には、22 人の歌手（女性 12 人、平均年齢 27.41±10.40 歳）と 5 人の審査員（女性 4 人、平均年齢 49.60±10.59 歳）が参加した。歌手は『Caro mio ben』を歌い、審査員は 100 点満点で歌声の印象を評価した。音声は課題曲の最高音を抽出し、Praat と MATLAB で音響解析し、統計解析は R を用いて順序ロジスティック回帰（審査員の得点 ～ SPR + SC + SPL + jitter + shimmer + HNR + CPP）を行った。その結果、SPL、jitter、shimmer、CPP と主観的な演奏評価に強い関連性がみられ、音量が大きく変化の多い声が高く評価され、安定性の低い声が低く評価されることが明らかになった。

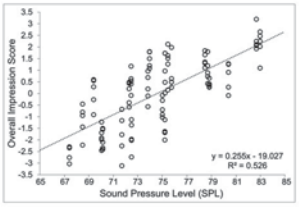


Fig.1：Vocal competition experiment. Fig.2: Relationship between overall impression score and SPL.

2-1-6

2-1-6 画像認識 AI による歌唱表現と顔の表情の定量分析

Quantitative Analysis of Singing Expression and Facial Gestures with Image Recognition AI

○高橋純(大阪芸大)，阪本英駿，北村達也(甲南大)，川下由理

◆歌唱において歌い手は，単に楽譜に書いてある音符通りに歌声を発するのではなく，歌詞の内容や旋律に合わせて**表現**を行なっている。

◆しかし，**歌唱表現**がどのような物理的事象を示すのかは明らかになっていなかった。

◆そこで本研究では，歌唱における顔の表情を画像認識 AI を用いることにより定量的に示し，歌唱表現との関係を検討した。

◆実験参加者は、音楽大学を卒業して声楽経験があり、現在も演奏活動を続けている女性の歌唱者７名である。

◆その結果，童謡チューリップにおいて表現をつけて歌唱する際には，**常に口角が高い位置**にあり，それにより頬があがって**口唇が開く**ことを定量的に示すことができた。

◆さらに表現ありでは、表現なしに比べて**頭部の動きが大きくなる**ことを明らかにした。

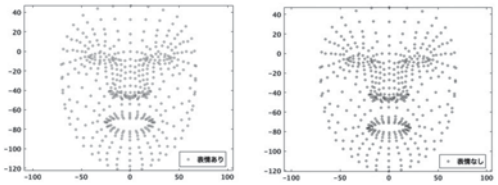


Fig.1：表現あり（左）・なし（右）における平均顔の比較

2-1-7

2-1-7 ４本マレット奏法におけるフィンガーコントロールと演奏音の関係

A Relationship Between Finger Control and Performance Sound in Four-Mallet Technique

☆竹下和秀(国立音楽大学大学院)，三浦雅展(国立音楽大学)，
小幡哲史，和田洋平，森　隆志(ヤマハ株式会社)

◆マリンバの４本マレット奏法の持ち方であるニュースタンダードグリップとインディペンデントグリップでは，打叩動作に指が使用されているものの，指の動きについて触れられている先行研究は存在しておらず，指の使い方に関する指導は指導者や演奏家の感覚的な判断から抜け出せていない。

◆本研究では，指の動作の収録に特化したモーションキャプチャシステムを使用して指の動作を収録し，指と手首関節の角度と演奏音の音圧レベルがグリップによってどのように異なるかを調査した。

◆調査の結果，インディペンデントグリップでは，打叩に環指と小指の DP 関節を使用していたが，ニュースタンダードグリップでは，環指の PIP 関節と DIP 関節を使用していることが分かった。演奏音については，インディペンデントグリップがダイナミクスの変化が大きい楽曲演奏に適しており，ニュースタンダードグリップはダイナミクスの小さい楽曲演奏に適していることが分かった。

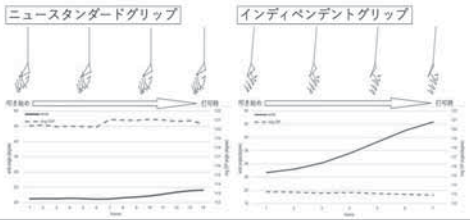


Fig.1 Example of striking motion differences between the New standard grip and Independent grip: angles of the right wrist and ring finger PIP joint when striking with the outer mallet of the right hand, with corresponding frame images."

2-1-8

2-1-8 クラリネット演奏の口腔内におけるリード挙動とリード品質の関係

A Relationship Between Reed Behavior Inside the Player's Mouth During Clarinet Performance and Reed Quality

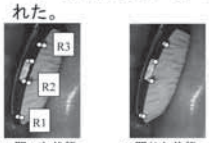
☆鈴木真琴(国立音大大学院)，若槻尚斗(筑波大学)，三浦雅展(国立音大)

◆クラリネット演奏におけるリード挙動は，人工吹鳴装置によって記録された事例はあるものの，奏者の口腔内を対象とした記録は実験環境の構築が困難であることから行なわれてこなかった。そこで，筆者らはストロボスコープを用いて口腔内を観察する方法を提案した[1]。

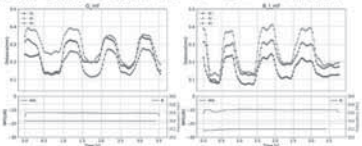
◆本研究はクラリネット演奏中のリード挙動を明らかにすることを目的とする。そのため，本報告ではクラリネット奏者が複数通りのリードを演奏した際のリード挙動とリード品質の関係について調査した。

◆本報告では試奏の評価で演奏適性が高いもの(G)，演奏適性が低いリードのうち，息が通りにくいもの，息が通りやすいもの(B_l)，そして演奏適性が高かったリードが演奏によって変化した結果古いと評価されたものを用いる。極小のカメラとストロボスコープとして LED を咥えた状態でロングトーンを演奏し，リード変位の記録と演奏音の収録を行なった。撮影した動画の例として，図 1(a)に口腔内でリードが開いた状態，図 1(b)にリードが閉じた状態のリードの様子を，図 2 に取得したリードの変位，演奏音の RMS および f0 周波数の一部を示す。

◆実験の結果，リード先端の最大変位量と演奏音の RMS には強い正の相関があることが確認された。さらに，演奏適性が高いリードの先端は見かけ上正弦波に近い振動をし，演奏適性の低いリードはマウスピースとの距離が大きくなった点で微細な振動をしていることが確認された。



(a)開いた状態　(b)閉じた状態
図 1 演奏中のリードの様子



(a)G における *mf* の演奏　(b)B_l における *mf* の演奏
図 2 リード変位，演奏音の RMS および f0 周波数

[1]鈴木真琴，若槻尚斗，三浦雅展，"クラリネット演奏の口腔内におけるリード挙動の撮影方法"，
日本音響学会音楽音響研究会資料，Vol.42，No.9，pp.33-38 (2023)

2-1-9

2-1-9　統計モデリングと機械学習に基づく音楽進化解析

Music Evolution Analysis Based on Statistical Modeling and Machine Learning

○中村栄太（九大・シス情）

本講演では、音楽文化の発展原理を理解するためのデータ科学的研究について議論する。特に、創作知識の伝達・変化・選択の過程の統計モデリングと機械学習に基づき、進化の駆動力や創作知識の参照プロセスを分析する研究などを紹介する。

【主な関連文献】

[1] E. Nakamura and K. Kaneko, “Statistical evolutionary laws in music styles,” Scientific Reports, 9(15993), pp. 1–11, 2019.

[2] R. Singh and E. Nakamura, “Dynamic cluster structure and predictive modelling of music creation style distributions,” Royal Society Open Science, 9(220516), pp. 1–18, 2022.

[3] E. Nakamura, “Computational analysis of selection and mutation probabilities in the evolution of chord progressions,” Proc. International Symposium on Computer Music Multidisciplinary Research (CMMR), pp. 462–473, 2023.

[4] E. Nakamura and Y. Saito, “Evolutionary analysis and cultural transmission models of color style distributions in painting arts,” Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 493–500, 2023.

2-1-10

2-1-10　天井に円形配置した6チャネル分散マイクロホンによる歌唱者の発声方向推定

Estimation of the direction of the singer's voice using circular six-channel microphones around overhead on the ceiling.

☆福田光太郎, 上林大輝, 津田龍星, 中山雅人, 高橋徹（阪産大院・工）

【目的】歌唱者の発声方向を天井に円形配置したマイクロホンによる観測音をもとに推定

【手段】五母音の平均放射特性と天井高さでの観測音から求めた 1/3 オクターブバンドレベルの比で発声方向を推定

・平均放射特性と歌唱音の計測時の配置 (Fig.1)

【評価】(1) 平均放射特性の発声者と，歌唱者が同一人物の場合 (Matched) と別人の場合 (Mismatched) で方向推定精度を比較

(2) 使用する周波数帯域で方向推定精度を比較

(3) マイクロホンの設置高さで方向推定精度を比較

【結果】(1) Matched, Mismatched 条件で方向推定精度は同程度(Fig.2)

→ 平均放射特性は本人のものである必要がない

(2) 推定精度は中心周波数 4000Hz の方が高精度

(3) 設置高さが方向推定精度に与える影響は小さい

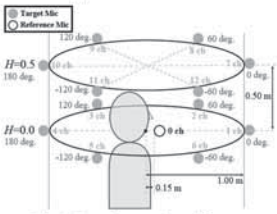


Fig.1 : Experimental condition.

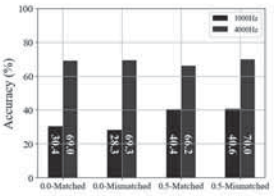


Fig.2 : Estimation accuracy.

2-1-11

2-1-11　オペラ歌唱における子音の発声技術の検討

A study on consonant production skills in operatic singing

☆深澤実紅, 北原玲奈, △久保田新菜, 志々目樹, 戸田菜月, 竹本浩典（千葉工大）, 高橋純（大阪芸大）

◆オペラ歌唱では，聴衆に明瞭に歌詞を伝えるために子音にも高度な発声技術があると考えられるが，その技術は明らかになっていない。

◆本研究では予備調査により選択した武島羽衣作詞・滝廉太郎作曲の「花」を用いてプロのオペラ歌手（プロ）と歌唱指導を受けていない学生（一般学生）の歌唱音声を録音し，その音声を分析して音素長や発声タイミングにどのような違いがあるか分散分析により検討した。

◆プロは語頭の子音を学生より有意に長く発声していた。

◆その一方，プロは破裂音の閉鎖区間が有意に短いが無声破裂音では気音部，有声破裂音では母音への過渡部を有意に長く発声していた。

◆さらに，プロは小学節の頭の部分の子音を有意に早く発声することで，子音を長く発声し，歌詞を印象付けていると考えられる。

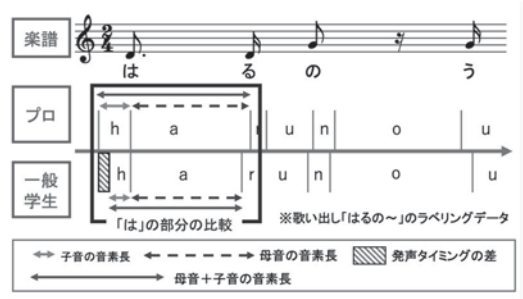


Fig1.プロと学生の比較例：歌い出しの「は」の部分

2-1-12

2-1-12　子音や母音の時間長が歌声の印象評価に与える影響

Effects of duration of consonants and vowels on impression rating for singing voice

☆北原玲奈, △久保田新菜, 深澤実紅, 志々目樹, 戸田菜月, 竹本浩典（千葉工大）, 高橋純（大阪芸大）

◆子音と母音の音素長が歌のうまさの評価に与える影響については明らかになっていない。

◆そこで本研究では，歌のうまさが平均的な学生（男性 M，女性 F）が武島羽衣作詞・瀧廉太郎作曲の「花」を歌唱した音声の音素長をプロの声楽家の平均的な音素長に変換して印象評価を行なった。

◆22 名の学生が歌のうまさの評価を行い，評価値からシェフェの一対比較法による分析を行って歌のうまさの尺度図 (Fig. 1) を算出した。

◆その結果，音素長を変更するだけでは印象評価は大きく向上しなかったが，音高を楽譜通りに修正すると大きく向上し，音高を修正した上で音素長を変換するとさらに向上することが明らかになった。

M00・F00：原音声　　M10・F10：音高修正

M01・F01：音素長伸縮　M11・F11：音高修正＋音素長伸縮

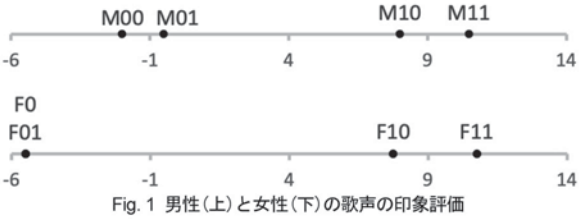


Fig. 1 男性（上）と女性（下）の歌声の印象評価

2-1-13

2-1-13　二重唱の歌い出しタイミングに対する同時性知覚の刺激閾調査

Investigation of stimulus threshold to simultaneity perception for onset of duet singing

☆兵藤弘明（東大院･情報理工），高道慎之介（慶應義塾大/東大院･情報理工），猿渡洋(東大院･情報理工)

◆重唱は，複数の歌唱者が互いに異なる声部を担当し，声を合わせて歌唱する歌唱形態であり，多くの人に親しまれている．

◆一体感のある重唱の実現には，楽曲内の然るべき箇所において，重唱グループを構成する者の歌い出しが互いに揃うことが必要である．

◆重唱における歌い出しの時間差の許容度合いがわかれば，重唱練習の指針となるだけでなく，許容度合いを歌声の客観評価指標として応用することも可能である．

◆本研究では二声部からなる重唱（二重唱）の歌い出しに注目し，聴取者が「揃っている」と判断する歌い出しの時間差の閾値の推定を行う．

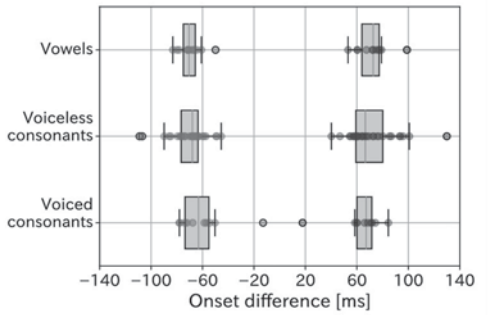


Fig.1：Estimated stimulus thresholds to simultaneity perception for onset of duet singing.

2-1-14

2-1-14　高齢者歌唱における独唱と集団歌唱でのフォルマントの違い

Differences in formants between solo and group singing in elderly

○太田 公子（同志社大学 文化情報学部）

♦高齢者による集団歌唱感覚を調査するため，歌唱データを収集し分析している。集団で歌唱を行うと，独唱の時より第1フォルマント周波数(F1)と第2フォルマント周波数(F2)のフォルマント面積の範囲が広がった。この結果を受け，本報告では集団歌唱感覚の要因の一つとしてフォルマントの変化を挙げ，歌い易さとフォルマントがどう関わっているか実験で検証した。

♦女声 22 名（平均年齢 70 歳）による童謡《紅葉》の合唱を 3 種類に加工したもの（歌詞あり斉唱，ハミング，2.5 kHz より高域を遮断）を用意し，77 歳の被験者に加工合唱音を聞きながら歌う実験を行った。

♦結果，加工合唱音のうち，歌唱あり斉唱は大きな声が出せていた。また，高い声を音程よく出せていたのはハミングであった。実験後被験者に快く歌唱できたものを問うと，ハミングが一番良かったという。歌詞ありの場合，子音後の母音部分で安定した音程が知覚される。ハミングではオンセット部分で有声音が出ているため，その声をガイドとしてオンセットで音程を合わすことができた結果，テンポよく歌えたことにつながったと考えられる。

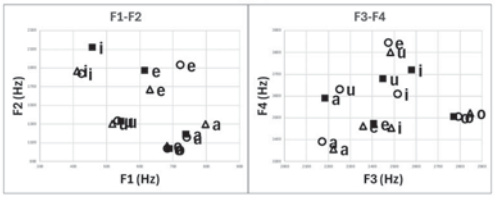


Fig.1：Distribution of 5 vowels in singing while listening to choral processing sounds. (■:Singing with lyrics, △:Cut-off frequency 2.5 kHz, ○:Humming singing).

2-1-15

2-1-15

微分可能なコンプレッサーの深層展開に基づくパラメータ推定と信号復元

Parameter estimation and signal recovery based on deep unrolling with differentiable compressor

☆王檉，赤石夏輝，山田宏樹，矢田部浩平(農工大)

◆ 背景
- コンプレッサー処理信号から元信号を復元する手法が望まれる
- それに対して，我々はこれまで
  - ✧ コンプレッサーのパラメータ推定 DNN を提案した
  - ✧ 処理信号を元信号に復元できる反復アルゴリズムを提案した

◆ 本研究
- 推定 DNN と復元アルゴリズムを組み合わせた深層展開に基づく手法を提案
- 復元アルゴリズムの性質を反映させた適切なパラメータを推定することで，元信号のより良い復元ができると考えられる

◆ 実験と結果
- T-Unet と提案手法の復元性能を比較調査した
- 全体的に T-UNet には及ばなかったが，深層展開による元信号の推定精度の向上を確認した

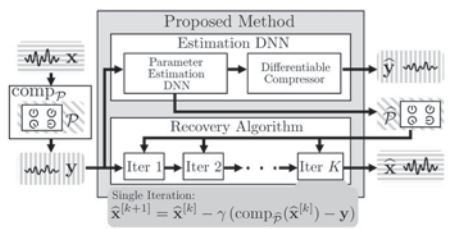


Fig.1: The structure of the parameter estimation signal recovery model based on deep unrolling with differentiable compressor

2-1-16

2-1-16　蝸牛遅延特性を模擬したIIRオールパスフィルタのリアルタイム実装

Real-time implementation of IIR all-pass filter simulating cochlear delay characteristics

☆飯田大貴，山田宏樹，矢田部浩平(農工大)

背景
- 我々はこれまで，蝸牛の群遅延特性を模したオールパスフィルタを設計した
- 蝸牛遅延を考慮した信号処理を音楽に適用することで，特徴的な音色を付加することができる
- フィルタをリアルタイムで適用することができれば，より広い範囲での活用が期待できる

本研究
- 蝸牛遅延フィルタの VST プラグインを作成した
- 楽器の演奏音に蝸牛遅延フィルタを適用することで，演奏にどのような影響を与えるか調査した

結果
- DAW 上で蝸牛遅延フィルタが正常に動くことを確認した
- 5倍の蝸牛遅延と120 msの遅延を与えた場合を比較すると，同程度にタイミングに遅れが生じることが分かった

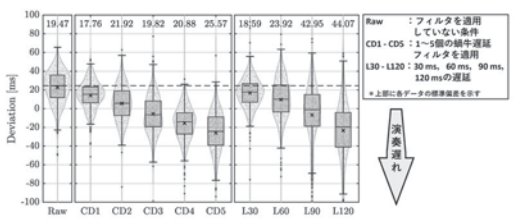


Fig. 1 The difference between the metronome beats and the guitar beats

2-1-17

2-1-17　moises による楽器分離と自動採譜のための主旋律推定

Instrument separation using moises and main melody estimation for automatic transcription

☆宮武碧，下倉良太，△飯國洋二（大阪大）

◆機械学習によるピアノ楽譜生成は複雑で人間には演奏不可能な楽譜を生成する．主旋律のみの音源を入力して，よりピアノ演奏に適した楽譜を生成することを目指す．

◆音源分離サービス moises を利用して混合音源を 10 個の音源に分離する．それぞれの音源に対して主旋律抽出モデルで F0 列を獲得する．混合音源の F0 列と，各分離音源の F0 列を比較し，最も類似度の高い分離音源が主旋律の音源であると推定する．

◆主旋律が ギター → トロンボーン → ギター と推移する楽曲で実験．最後のギター以外は正しく推定できている．（Fig. 1）

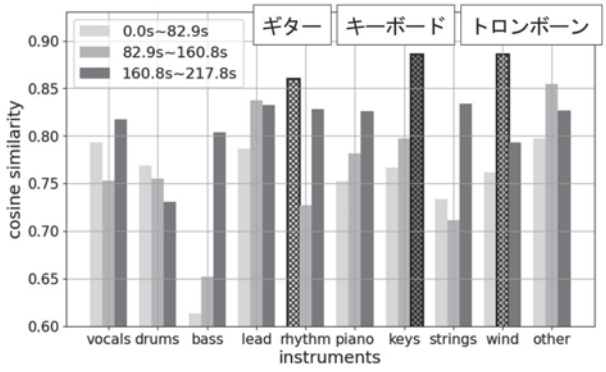


Fig. 1: Flying/Aimee Norwich の実験結果．各時間の最大に網掛けを付与．

2-1-18

2-1-18　各楽器パートに焦点を当てた知覚的楽曲間類似度の調査

Investigation of perceptual music similarity focusing on individual instrumental parts

☆橋爪優果，戸田智基（名古屋大）

◆ABX テストを行い，のべ **451 人**による有効な回答を得た．

◆被験者は**楽曲とそれを構成する 5 種類の楽器音**を個別に聴取し，それぞれで A, B どちらが X に似ているか回答

◆**音色，リズム，メロディ，総合**という 4 つの観点で回答

◆楽器間での**類似度基準が異なる**こと，**リズムやメロディ**が重視されること，先行研究の手法の**特徴や改善点**が示唆された．

Fig.1: Experimental Screen

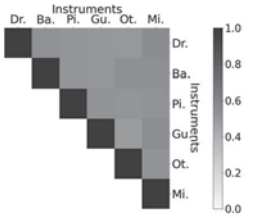


Fig.2: Percentage agreement of answers between each instrument. Dr.: drums, Ba.: bass, Pi.: piano, Gu.: guitar, Ot.: others, Mi: mix.

	drums	bass	piano	guitar	others	mix
音色	73.6	73.1	74.7	75.3	74.8	74.1
リズム	**85.1**	**80.3**	78.7	**80.4**	79.7	**83.2**
メロディ	56.2	**79.7**	**81.1**	**81.3**	**83.1**	**83.8**

Table 1: Percentage agreement with “総合” for each perspective. *(%)*

2-1-19

2-1-19　コード進行の類似度と遷移に着目した音楽推薦システムの提案

A Music Recommendation System Based on Chord Progression Similarity and Transition

☆小野坂 浩，△常盤 達司（広島市立大院・情報科学研）

◆本研究では、楽曲の雰囲気を構成する「コード進行の時間遷移」を用いた音楽推薦システムを提案する。

◆従来の推薦システムでは、一曲全体の特徴を用いた推薦（楽曲間推薦）が主流であったが、本提案システムでは、一曲の一部分のコード進行に基づき、コード間類似度を用いた推薦（楽曲内推薦）を行う。

◆本システムにより、類似したコード進行を用いている異なるアーティストの曲を発見することができるので、楽曲同士の意外な関係性に気づく可能性がある。

Song to Search： Cherry / Spitz

Chord progression to search: ['C', 'G', 'Am', 'Em', 'F', 'C', 'F', 'G']

Chord distance: ['6.33', '8.944', '6.33', '9.429', '6.33', '6.33', '9.191']

Songs with similar chord transitions:

・Marigold / Aimyon :

['C', 'G', 'Am', 'Em', 'F', 'C', 'F', 'G'], Similarity: 0.0

・Magic Carpet / Takaya Kawasaki :

['F', 'C', 'Dm', 'G', 'F', 'C', 'G', 'Am'], Similarity: 0.361

・Sora Mo Toberu Hazu / Spitz :

['C', 'G', 'Am', 'F', 'G', 'C', 'F', 'G'], Similarity: 0.507

Fig.1 : Proposed Music Recommendation System

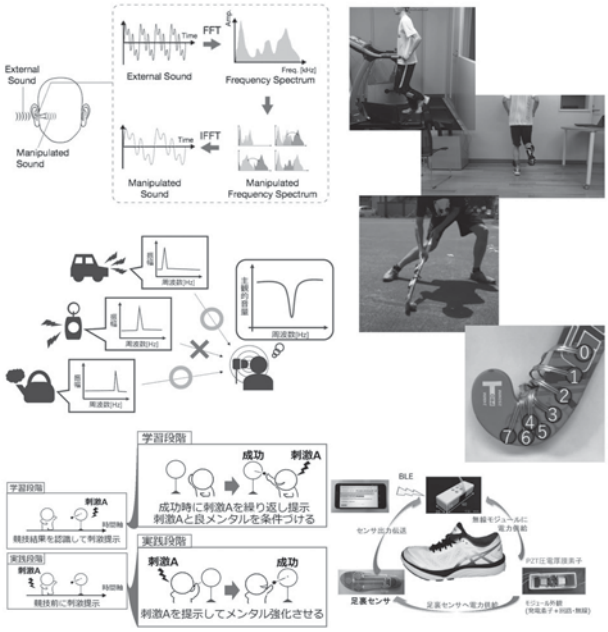
2-2-1

2-2-1　ウェアラブル音響フィードバックとスポーツ

Wearable Auditory Feedback for Sports

〇寺田 努（神戸大）

本講演では，ウェアラブル音響フィードバック技術を用いてスポーツの支援(上達支援，行動制御，メンタル支援)を行う事例について講演者らの取組みを解説します．



2-2-2

2-2-2　センサデータを活用したスポーツ支援

Sports Support Utilizing Sensor Data

○内山彰（大阪大）

我々のグループでは，ウェアラブルセンサやサーモグラフィ，慣性センサ，映像などのセンサデータを活用したスポーツ支援に関する研究を実施している。本講演では，ウェアラブルセンサやサーモグラフィを用いて深部体温を推定する取り組みや，選手やボールの位置データからサッカーのプレーを認識する取り組みなどについて紹介する。

これらの研究によって，熱中症の原因となる深部体温の上昇をモニタリングすることができるようになり，熱中症の予防に貢献できると考えている。また，スポーツにおけるプレー認識の実現は，選手の技能向上に寄与するだけでなく，適切なタイミングでの交代や戦術の変更といった戦略面での活用も期待されている。さらに，観客に対して定量的なデータを示すことにより，エンターテイメント性を高め，競技の魅力向上にも利用可能である。

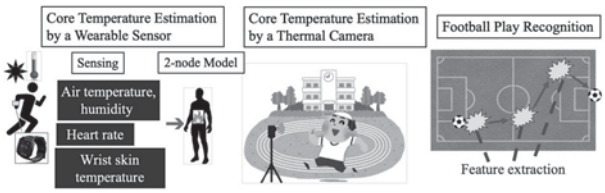


Fig.1：Core temperature estimation by a wearable sensor and a thermal camera, and football play recognition.

2-2-3

2-2-3　スポーツ観戦者から聴取した競技者への声援に関する研究 —各場面における日本語話者と外国語話者の声援の比較—

A Study on Cheering for Athletes by Sports Spectators
- A Comparison of Japanese and English Speakers' Cheering -

○久永将太（慶應大・総），浅野恵子（順天堂大・医）

◆「声援」の意味とは、競技・試合などで声（心の声も含む）による応援をし、選手やチームを元気づけることを指す

◆久永、他（2024）が日本語話者のスポーツ観戦者に対して行った声援に関するアンケート調査を外国語話者（英語話者）の観戦者に実施

◆外国語話者である 18-35 歳の 70 名（男性 35 名、女性 35 名）に実施

◆外国語話者も日本語話者と同様に「声を出して」、「小声で呟くように」、「声には出さず心の中で」と、主に３種類の声援のスタイル

◆【楽しい】・【嬉しい】・【ワクワク】・【驚いた】の場面において、両者共に非常に高い割合で声を出して声援

◆【悲しい】・【悔しい】・【がっかり】・【怒り】の場面では、外国語話者よりも日本語話者のスポーツ観戦者の方が声を出して声援（$p<0.01$）

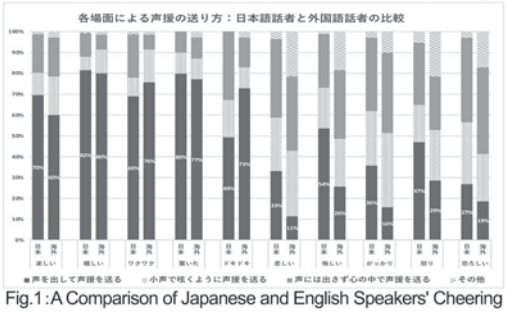


Fig.1：A Comparison of Japanese and English Speakers' Cheering

2-2-4

2-2-4　野球の木製バットの打球音の聴感評価 ～野球未経験者が捉える打球音の印象～

Evaluation of the sound of hitting a baseball wooden

☆三國陸冬、内藤光平(神奈川工科大)
三浦貴大(産総研)、上田麻理（神奈川工科大）

◆ 野球競技において，聴覚情報は重要な要素の一つである．本稿では，特に打球音に着目し，様々な観点から木製バットの打撃結果と聴感印象に与える影響の関係性についての研究結果の報告をする．

Fig.1 野球選手の打球フォーム

2-2-5

2-2-5　木製バットの打球音聴取における打球種の判断に関するデータ分析*

Data analysis on determination of hitting type in listening to hitting sound of wooden bat

☆秋山智哉，田中哲雄，上田麻理(神奈川工科大)

◆野球競技で発生する音は選手たちのパフォーマンスに影響を与えていることが予想される．

◆本研究では，野球の打球音から打球の種類を判別できるか否かを明らかにするため，打球種当てクイズアプリケーションを作成した．

◆作成したアプリケーションを使用して，野球競技経験者・未経験者を対象としたクイズの回答と打球音の特性を分析する．

Fig.1 本アプリケーションのクイズ画面

2-2-6

2-2-6　野球の非木製バットにおける打球音の聴感評価

Auditory Evaluation of Hitting Sound in Non-wood Baseball Bats

☆内藤光平(神奈川工科大)，三國陸冬(神奈川工科大)，
三浦貴大(産総研)，上田麻理(神奈川工科大)

◆ スポーツ競技における用具の性能向上が進み，日進月歩で新たな記録が生まれている．一方で，野球競技においては投手の受傷事故が増えてきたため，2022 年 2 月に競技者の安全性を考慮し，非木製バットの規定が改定された．その結果，打球音における高さや大きさに変化が生じたと訴える者が増え，競技者に様々な影響を与えている．しかし，この非木製バットの音響特性や競技者への影響については定量的には不明瞭である．

Fig 野球の競技イメージ

2-2-7

講演取消

2-2-8

2-2-8　FPS ゲームにおける聴覚情報利用の重要性に関する検討その 4 －Overwatch2 での聴覚情報ごとの活用状況－

A basic study on the importance of using auditory information in FPS games Part 4　–The use of auditory information in Overwatch2

☆矢内 遼太郎，上田 麻理(神奈川工科大)，三浦 貴大(産総研)

◆目的：FPS ゲーム Overwatch2 における聴覚情報の活用状況と，プレイヤーのスキルの関係性を明らかにする．

◆結果：ゲーム内のアビリティ音は，銃声よりも有意に注意される．BGM 音量は SE 音量より有意に小さく設定される．経験年数が長い者は，多くのプレイヤーが注意しない銃声に有意に注意しうる．

2-3-1

2-3-1　首都圏の住宅地域における音環境と保育施設に関する意識調査

Survey on Attitudes towards Acoustic Environment and Childcare Facilities in Residential Areas in the Tokyo Metropolitan Area

☆黒木幸穂(九州大･芸工)，片岡寛子(電通大)，高田正幸(九州大･芸工)

◆待機児童問題を解消する方法の一つに保育施設の新設があるが，新設が計画される地域の住民の反対運動が起きることがあり，この争点として保育施設から発生する音の問題がある。

◆福岡市で実施された先行研究では，保育施設新設への賛否は，保育施設との関係性や騒音感受性といった個人の特性が関与していることが示された。地域の環境や人口密度などの特性が異なる地域でも同様の傾向が見られるか検討するために，首都圏で同様の調査を実施した。

◆音環境の実測調査と調査票を用いた住民の意識調査を実施した。調査票には，先行研究と同様の質問にソーシャルキャピタル（住民の社会参加の状況や地域コミュニティとの繋がり）に関する質問を追加し，保育施設新設への賛否と関係する各種要因の検討を行った。

◆結果として，保育施設からの音の聴取状況と音環境の満足度には関連が見られなかった（Fig.1）。さらに，保育施設新設への賛否には回答者の年齢，居住年数，他者への信頼の程度といった個人の特性が関連することが分かった。

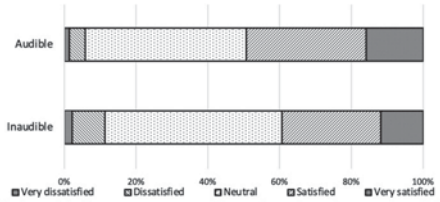


Fig.1 Relationship between sound listening situation and satisfaction with acoustic environment in sites with childcare facilities.

2-3-2

2-3-2　風鈴の音色に関する研究 鳴動パターンを統一した心理的評価

A Study of Impressions of Wind Chime Sounds
Psychological evaluation of unified ringing patterns

○土田義郎（金沢工大）

◆25 種類の風鈴について各種音響評価指標の分析を実施

◆主観的評価→自然な鳴動パターンの複合音を使用

◆音響評価指標と印象との関連を明らかにすることを目的

音響評価指標	評価性	明瞭性	繊細性	複雑性
基本周波数	−0. 25	0. 10	0. 46	0. 03
スペクトル重心	−0. 11	0. 15	0. 37	−0. 40
高調波比	0. 10	0. 06	0. 05	−0. 33
シャープネス	−0. 08	0. 16	0. 42	−0. 41
うなり	−0. 06	−0. 04	0. 44	0. 05
ラフネス	−0. 10	−0. 18	0. 06	−0. 09
変動強度	0. 09	−0. 24	−0. 05	0. 30
減衰時間	0. 09	0. 43	0. 11	−0. 45

◆4 因子によって説明可能

◆評価性：どの指標とも相関が低い
→音の好みに個人差が影響

風鈴ごとの因子の標準偏差平均：評価性 0.949、明瞭性 0.935、繊細性 0.932、複雑性 0.863

◆明瞭性：減衰時間との相関が最も高い
余韻が長い＝金属の風鈴の特徴　金属は明瞭性が高い

◆繊細性：基本周波数、シャープネス、うなりとやや相関が高い

◆複雑性：スペクトル重心、高調波比、シャープネス、減衰時間と負の相関、変動強度でやや正の相関があり

2-3-3

2-3-3　屋外での執務活動における音環境の寛容さに関する研究 —騒音レベルと音環境の許容度の関係—

A Study on Tolerance of Sound Environment in Outdoor Office Activities
—Relationship Between Noise Level and Tolerance of Sound Environment—

◎原健登（鹿島技研），村上由衣（茨城大・院），
星野嗣人，矢入幹記（鹿島技研），辻村壮平（茨城大・院）

◆近年，屋外空間の日常的で柔軟な活用が推進されるなか，屋外での執務活動に関わる音響的な設計指針や基準値は定められていない．

◆そこで，提示騒音レベルを統制した屋内，屋外，半屋外空間での騒音に対する許容度の差異を捉えることを目的とした印象評価実験を行った．

◆実験から得た場所ごとの音環境に対する許容度の回帰直線を比較し，被験者が音環境として許容できると評価する平均の騒音レベルが，屋外では屋内よりも 4.7 dB 大きいことが分かった．

Fig.1：Experimental locations for impression evaluation experiments.
(Left：Inside space, Center：Semi-Outdoor space, Right：Outdoor space)

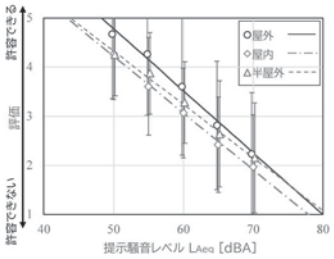


Fig.1：Evaluation of tolerance (sound environment) under each condition
And Regression line for the mean of the ratings.

2-3-4

2-3-4　屋外での執務活動における音環境の寛容さに関する研究

−執務活動の違いによる音環境の許容度の差異の要因分析−

A study on tolerance of sound environment in outdoor office activities
-A factorial analysis of differences in tolerance of sound environment in different office activities-

○村上由衣（茨城大学・院），原健登（鹿島技研），
星野嗣人，矢入幹記（鹿島技研），辻村壮平（茨城大学・院）

◆ 近年ワークスタイルが多様化し，屋外を執務空間として利用することが検討されていることから，屋内・屋外・半屋外で音環境条件を統制した印象評価実験を行い，執務活動（休憩，創造性テスト，会話）の違いに着目し，音環境の許容度の差異を調べた．

◆ 執務行為ごとに，各音圧レベルにおける音環境の許容度の平均値に対して単回帰分析を行った結果，全ての執務活動で屋外の許容度が高く，屋外で執務活動ごとに比較すると，屋外では許容できる音圧レベルは休憩時に最も低いことが示唆された．

◆ 創造性テストや会話時は，作業に注意が向けられるため，音圧レベルが高くても比較的許容されやすいと考えられる．

◆ 音環境が許容できる程度は，重回帰式より，休憩で聴覚・視覚要因の項目が，作業時は聴覚要因の項目の説明変数で表された．

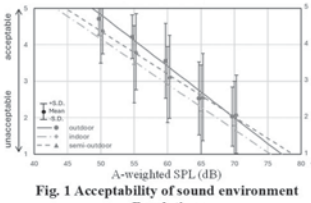


Fig. 1 Acceptability of sound environment
Break time

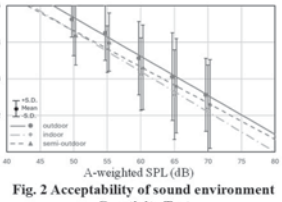


Fig. 2 Acceptability of sound environment
Creativity Test

2-3-5

2-3-5　障害物の聴知覚における背景音の影響

The Effect of Background Sound on Auditory Perception of the Object

☆勝倉朱音，亀川徹，丸井淳史（東京藝大）

◆人間が聴覚を用いて空間を認知する際の、周囲の背景音の影響を検証した。

◆実験参加者を取り囲む５チャンネルのスピーカから背景音を再生した。実験参加者は、背景音が再生される中で、前方のスピーカから再生される放射音を手がかりに、障害物の位置の判断を行った。

◆背景音がピンクノイズである場合の結果(Fig.1)と同様に、全ての背景音において、実験参加者から障害物までの距離が 80 cm から 160 cm にかけては、障害物の距離の遠近感がおおむね正しく知覚された。一方で、40 cm の場合は、障害物の位置が正しく知覚されなかった。

◆実験で用いた放射音と背景音の音響特徴量を検証した結果、放射音の音圧レベル、及び周波数特性が障害物の位置の知覚の手がかりになる可能性がみられた。また、背景音の周波数特性、スペクトラルフラックス、IACC の影響がみられた。

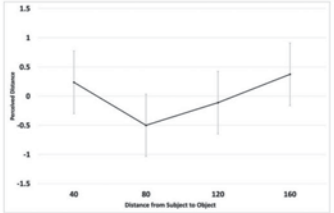


Fig.1：Interval scales of auditory perceived distance
(Background Sound: Pinknoise)

2-3-6

2-3-6　「音空間デザイン」の提案 –デザインプロセスの見直し及び本校での実践–

Proposal of “Sound Space Design”

- Review of the design process and practice in our college -

☆宮本岬, 石川あゆみ(岐阜高専)

◆本研究では、音空間デザインのプロセスを見直したうえで、本校にて音空間デザインの実践を行った。

◆デザインプロセスの見直しとして、事前調査の項目に既設又は新設の放送設備の調査を追加し、順序の変更を行った。

◆本校図書館及び図書館前テラスを対象地として音空間デザインの実践を行った。

◆Table2 に使用した音の一覧を示す。また、Fig.4 にスピーカーの設置場所やそこで再生する音、音を届けたい範囲を図面に示した音空間デザインのイメージを示す。

Table2 List of sounds

No.	テーマ	説明
M1	非日常	ピアノの音で作成した不思議な音
M2	鳥	様々な鳥の鳴き声
A1	旅 1	ドア開閉音や食事の音などの生活音
A2	旅 2	波音や汽笛の音など
A3	旅 3	雑踏音や足音など
A4	時間	ヒグラシの鳴き声や秒針が進む音など
A5	建築	鳥の鳴き声や木を切る音など
A6	冒険	歩く音や雨音、風の音など
A7	し	ドアをノックする音など
A8	書	筆記音やペンを出す音など
A9	森	小川のせせらぎ音や虫の鳴き声など

Fig.4 Image of Sound Space Design

2-3-7

2-3-7　バイオフィリックデザインにおける視覚と聴覚の複合影響を考慮したオフィスワーカーの評価モデルとその性差に関する検討

Investigation of the Evaluation Model of Office Workers Considering the Combined Effects of Visual and Auditory Stimuli in Biophilic Design and the Investigation of Gender Differences

☆遠藤環, 吉田香純(茨城大院), △横田健治, △佐土島哲次,
△坂井剛(パナソニック ホールディングス), 辻村壮平(茨城大院)

◆バイオフィリックデザインが施されたオフィス空間に適した自然音の選定ならびに性差を検討するため，聴感評価実験を実施した.

◆実験で用いた「好ましさ」,「働きやすさ」,「快適さ」の総合評価３項目について，性別と音源の種類を要因とした繰り返しありの二元配置分散分析を行った結果，性別と音源の種類において交互作用がみられた（Table 1）.

◆各音源に対して性別を要因とした *t* 検定を行った結果，鳥の鳴き声と雨音は総合評価３項目の全てにおいて有意差がみられ，鈴虫は「好ましさ」のみ有意であった．川のせせらぎ，暗騒音，波音では性別による有意な差はみられなかった．川のせせらぎと暗騒音は男女ともに SN 比の影響があることが示唆された.

Table. 1:Result of two-way analysis of variance

** : p<0.01

	Gender	Types of natural sounds	Gender × Type of natural sounds
Favor	n.s.	**	**
Ease of work	n.s.	**	**
Pleasantness	n.s.	**	**

2-4-1

2-4-1　音響分析を言語聴覚士の臨床で使うための壁

Obstacles to using “acoustic analysis” in clinical practice for speech therapistes

〇竹内京子(順天堂大), 青木直史(北大), 荒井隆行(上智大),
大金さや香(国際医療福祉大), △鈴木恵子, 秦若菜, △村上健(北里大),
世木秀明(千葉工大), 安啓一(筑波技大)

◆ 言語聴覚士の養成校では，音響学・聴覚心理学が必修で，学生が最も苦手な科目。この傾向は卒業後も続く。2021 年から現役言語聴覚士対象の講習会「ST のための音響学」を開催，「音響分析をいかに臨床で使うか」を掲げてきた。本発表は, 2024 年現在の報告である。

◆ 音響分析ができない理由は，まず音響学が苦手であること，養成校で音響分析を習わなかった，音響分析を実践する ST がいない，音響分析を教える音響学教師の不足，職場の音環境の悪さ，設備投資のための予算不足，講習会の機会・広報が足りないなどである。

◆ 具体的な解決策は，講習会の内容の充実，音響学教師・学生限定講習会の開催などである。そして，短期間で結果が出なくても，あせらず，時間をかけ，続け，やめないことである。

◆ 音響学の応用は「音響分析」だと言う言語聴覚士が多い。しかしながら，その音響分析も臨床で行うのには，様々な壁が存在する。学校で習って終わりではなく，それを社会で実際に応用することを目的とした職業教育ならではの問題である。

◆ 今後は，ただカリキュラムにあるから教えるのではなく，時には教室を出て，社会に目を向け，学生の卒業後の姿も頭に描きながら，よりよい授業を考えることが求められているだろう。

2-4-2

2-4-2　篩骨洞の開放の程度が鼻音に与える影響

Effects of degree of opening of the ethmoid sinus on nasal sounds.

☆福島侑希, 竹本浩典(千葉工大), 北村達也(甲南大),
△鴻信義, △武田鉄平, △小黒亮史, △大村和弘(慈恵医大)

◆鼻副鼻腔の形状は複雑で，音声の個人性の生成要因の一つである。その形状を手術などで変化させると，音声も変化する。前著では前後の篩骨洞のみを開放した影響を検討したが，実際の篩骨洞の開放手術では，前篩骨洞の開放と同時に上顎洞や前頭洞も開放される。

◆本研究では，４名の CT 画像から鼻副鼻腔形状の抽出と模擬手術を行った。抽出のみを行ったモデル（ORG）と，前篩骨洞のみ（AES），前後篩骨洞（ES），上顎洞（MS），前頭洞（FS）を開放したモデル（計９種），合計１０種類のモデルで音響シミュレーションを行い，実際の手術に則した開放が鼻音スペクトルに及ぼす影響を検討した。

◆前篩骨洞のみ（AES），前・後篩骨洞（ES），前頭洞（AES+FS）の開放による伝達関数では，第１，２ピーク（P1，P2）の間にディップ（▽）を生成したが，上顎洞の開放によるディップよりも相対的に小さく，音声への影響は小さいと考えられる。

◆上顎洞（MS）の開放による伝達関数では，４名の被験者に共通して１kHz 付近に非常に深いディップ（▼）が生成された。これは術後に患者が，鼻の中の響きが大きく変化したと感じる要因である可能性がある。

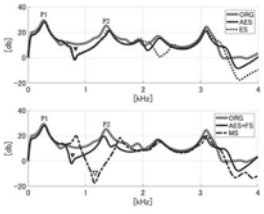


Fig.1:Comparison of transfer functions

2-4-3

2-4-3　無声音を含む発話を可能にする代用発声デバイス

Alaryngeal voicing devices that enable speech with voiceless sounds

☆小野寺悦希 (筑波大・院・シス情工)，若槻尚斗，海老原格 , △前田祐佳 (筑波大・シス情系)

著者らは，代用発声によって無声音を再現することを目的として，(Fig.1)のような代用発声デバイスを試作した。

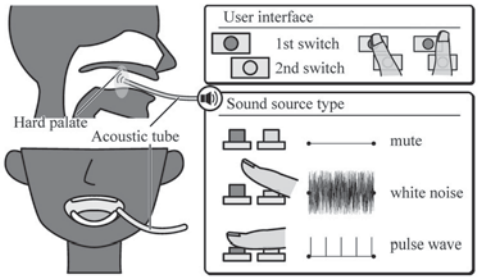


Fig.1 : Experimental system.

- **音響管をくわえ，音源信号を伝達する方式**
  無声音特有の高周波成分の伝達を可能にする
- **隣接したスイッチによる有声音・無声音の切り替え**
  簡単な操作によって/p/,/h/などの無声子音の再現を試みる

結果・考察

(Fig.2) に示すスペクトログラムを比較した。
無声子音の特徴である「瞬時的な無声音」を再現
フォルマントの保持を確認
聴取実験などを行い
性能の評価を行う予定

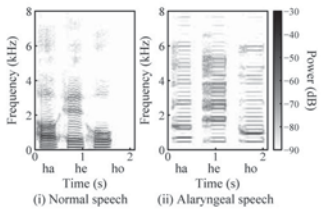


Fig.2 : Spectrograms of vocalized example sentence.

2-4-4

2-4-4　高齢者の認知機能レベルと BCI ひらがな単音発声の韻律情報との関係

☆金成健太朗，田中久弥(工学院大･工)

本研究では，認知機能低下レベル(L0，L1，L2)により群を区別し，単音発声時の韻律情報を用いて認知機能レベル群間でクラスカルウォリス検定, Steel-Dwass の多重比較と機械学習による分類を行い, 認知機能と単音発声時の韻律情報との関係について検討した. /a/を発声した際の第１フォルマント周波数の最小値と第４フォルマント周波数の最大値を用いて認知機能低下レベルの分類を行った. その結果，２分類(L0，L1)では 88%，３分類(L0，L1，L2)では 75%の分類精度を達成し，この単音発声時の韻律情報が認知機能との関係がある可能性が示唆された.

表１ クラスカルウォリス検定と群間比較

文字	韻律情報
/a/	第４フォルマント周波数の最大値 第１フォルマント周波数の最小値 MFCC の９次元
/mo/	MFCC の 11 次元
/ha/	MFCC の９次元
/re/	第３フォルマント周波数の平均値 MFCC の９次元

表２　k 近傍法による認知機能レベル分類率

文字	韻律情報	２分類	３分類
/a/	①第４フォルマント周波数の最大値	48%	46%
	②第１フォルマント周波数の最小値	64%	57%
	③MFCC の９次元	65%	54%
	①と②	88%	75%
	①と③	48%	47%
	②と③	64%	57%
	①と②と③	88%	75%
/mo/	MFCC の 11 次元	50%	44%
/ha/	MFCC の９次元	61%	56%
/re/	①第３フォルマント周波数の平均値	51%	54%
	②MFCC の９次元	51%	57%
	①と②	48%	50%

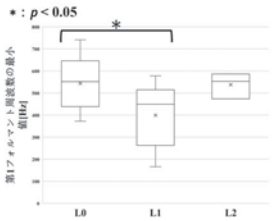


図１　/a/の L0 と L1 における群間比較

2-4-5

2-4-5　吸気発声における声門流の数値流体力学的検討

A study on numerical fluid analysis of glottal flow in inspiratory phonation

○加藤日花里（KDDI), 李庸學, 若宮幸平, 鏑木時彦（九州大）

- 吸気発声のメカニズムを明らかにするため、仮声帯を持つ喉頭モデルを用いて、呼気発声と吸気発声時の声門流の動態を仮声帯間隙・声帯間隙・声門下圧を変化させて数値解析した。
- 本研究では、声門流を非圧縮・非定常流体として扱い、数値シミュレーション手法にはLarge Eddy Simulation（LES）を用いた。
- 吸気発声においては、呼気発声で観察される声門ジェットに加えて、仮声帯から喉頭室に向かう帯状の流れが発生した。
- 仮声帯間隙・声帯間隙・声門下圧が声門流の動態に影響を与えることが明らかになった。
- 吸気発声において、コアンダ効果に起因する声門流の非対称性が、声帯音源に一定の効果を与える可能性が示唆された。

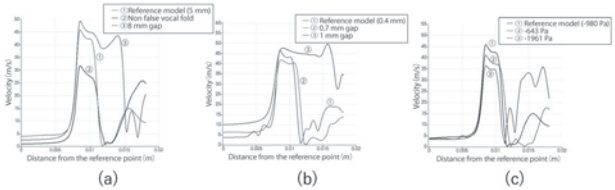


Fig.1：Velocity distribution under conditions of (a) false vocal fold gap, (b) vocal fold gap and (c) subglottal pressure in inspiratory phonation.

2-4-6

2-4-6　声帯−仮声帯発声における音源分布に関する検討

A study on the sound source distribution in vocal and false vocal fold phonation

○鏑木時彦（九州大）

- 有声音の音源は、声帯弾性体と呼気流の相互作用による声帯音源である。声帯上部にある仮声帯は、通常の話声では両側に開かれた状態にあるが、喉歌などの歌唱発声では声帯とともに振動し、音声のスペクトルにサブハーモニクスを生じる。我々の研究より、このような声帯−仮声帯発声では、声帯と仮声帯それぞれの振動周波数に加えて、側帯波に相当する周波数成分が音声のスペクトルに表れ、これが従来はサブハーモニクスとして解釈されていたことを明らかにした。
- 本研究では、声帯−仮声帯発声における音源機構を明らかにするため、声帯と仮声帯を含む喉頭内流路を通過する 2 次元の非定常・低圧縮性の流れの数値解析を格子ボルツマン法により行なった。喉頭内の各点における圧力の時間変動より、特定の周波数成分の振幅を求め、音源の空間分布を検討した。
- その結果、圧力変動の空間分布として、声帯の振動周波数に対応する成分は声帯下流、仮声帯の振動に対応する成分は仮声帯の上流と下流、側帯波に相当する周波数成分は仮声帯の下流と、それぞれに特徴的な分布が存在することがわかった。本研究は、JSPS 科研費 JP24K15010 の援助を受けた。

2-4-7

2-4-7　声帯結節を模擬した物理モデルを用いた病的な発声の特徴に関する研究

Studying the feature of pathological vocalization by using physical model of the vocal cord nodules

☆上田可奈子，△竹内理人，徳田功（立命館大）

◆人間の声帯を MRI で撮影して得られたデータに基づき，声帯結節を模擬した物理モデル（MRI 結節モデル）を作成した。

◆結節の大きさ及び硬さがそれぞれ４種類ずつ異なる物理モデルを用意した。

◆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，結節のある物理モデルと結節のない物理モデルを比較した。

◆声帯結節が大きく及び硬くなっても、最大圧力とオンセット圧に影響がないことが確認できた。

◆キモグラフより，声帯の背側では周期的に振動しているのに対し、腹側では複雑な振動が周期的に起こることがわかった。

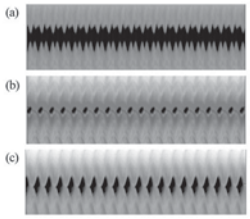


Fig.1 Kymograph of vocal fold (a) anterior, (b) center, and (c) posterior.

2-4-8

2-4-8　rtMRI を用いたオペラ歌唱における演奏技法や音高が異なる際の体内運動の分析

Analysis of internal body movements in different performance techniques and pitches in opera singing using rtMRI

◎戸田菜月，竹本浩典（千葉工大），高橋純（大阪芸大）

◆優れたオペラ歌手は声道形状や横隔膜を精密に制御することで，演奏技法や音高を変化させた場合でも，特有の響きや豊かな声量を維持することを可能にしていると考えられるが，詳細は明らかでない。

◆先行研究では演奏技法に合わせて肺圧を制御することや，音高の変化に合わせて声道形状と肺圧を制御することを明らかにしたが，演奏技法の変化に伴う声道形状と肺圧の制御については未検討である。

◆そこで本研究では，バリトン歌手１名の演奏技法と音高の変化に伴う声道形状と肺圧の制御を検討するとともに，その制御をこれまでに数名の歌手で得られている知見と比較検討した。

◆その結果，音高の変化に伴う制御はこれまでに得られた知見と一致し，また演奏技法の難易度が高くなるほど肺圧は高く，音高の上昇に伴う声道の制御は大きくなることが明らかとなった（Fig. 1）。

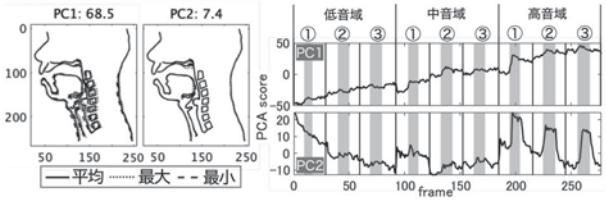


Fig.1：声道形状の主成分分析結果（左）とそのスコア（右）

（①：演奏技法の指示なし，②：アクセント，③：ポルタメント）

2-4-9

2-4-9　能登半島地震におけるアナウンサー避難指示音声の音響特徴

Acoustic characteristics of evacuation calls by TV announcers during the Noto peninsula earthquake disaster

○牧 勝弘，天野 成昭，小川 一美（愛知淑徳大）

◆2024 年１月１日に発生した能登半島地震直後の NHK の緊急報道では、津波に対する住民の迅速な避難を促すために、強い口調でアナウンサーによる避難の呼びかけ（「命を守る呼びかけ」）が行われた。

◆本研究では、津波警報、および大津波警報発表報道時における NHK アナウンサーの避難呼びかけ音声の音響特徴を調べた。

◆その結果、平時のニュース音声に比べて避難呼びかけ音声では、基本周波数が平均で約 100 Hz 以上高く、周波数スペクトルの１kHz から 3 kHz までのパワーが大きく、スペクトル包絡の動的変化が大きいことが明らかになった。

◆また、避難呼びかけ音声では、基本周波数の時間変化が小さい時に顕著化する基本周波数と音の強さ（200 Hz～300 Hz の帯域）の連動（図１矢印）が弱まることも明らかになった。

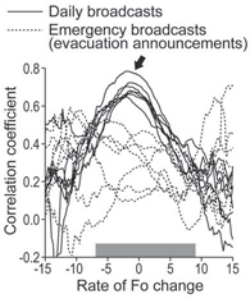


Fig.1 Correlation coefficient between the fundamental frequency and the RMS value of the amplitude envelope for various rates of change of the fundamental frequency. Regarding the difference between daily and emergency broadcasts, the areas where the Welch's t-test (p < 0.05) indicated significance are shown in grey bar.

2-4-10

2-4-10　人声のステアリングにおける放射特性の音素間平均の影響

Effect of averaging each phoneme's radiation characteristics on steering sound for human voice

○木下光太郎，杉本岳大（NHK）

◆AR / VR などの視聴時に**音源の放射特性**を再現することで，音源の向きに依存した音色の変化を表現する**ステアリング**を検討している。

◆前報では，音素ごとに放射特性が異なることを根拠として，音素ごとにステアリング処理した音声を再結合し，総合的なステアリング音声を生成していた。

◆ただし，会話などの自然な発話においても放射特性の音素依存性が聴感上，必要なのか未確認であった。

◆本稿では，放射特性を音素間平均したデータベースからステアリング音声を生成する手法を提案法とし，主観評価実験で従来法含めて聴き比べることで，放射特性の音素依存性を確認した。

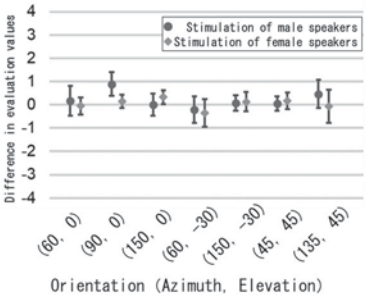


Fig.1 Difference in evaluation values between proposed and conventional methods

2-4-11

2-4-11　ポップアウトボイス生成時における声道の音響特性

Acoustic characteristics of the vocal tract during production of pop-out voice

☆相馬巧海，竹本浩典（千葉工大），北村達也（甲南大）

◆背景雑音などの妨害音がある環境下でも聞き取りやすい音声をポップアウトボイスと呼ぶ。本研究では，ポップアウトする声（PV）とポップアウトしない声（NV）で母音発声時の 3 次元 MRI データから声道断面積を抽出し，音響特性とともに比較検討した。

◆特にポップアウトする度合いが高かった 5 つの母音のスペクトルを比較したところ，PV では F1，F2 が上昇し（F01 を除く），F3 が上昇して F4 が下降する傾向が見られた（Table 1）。ポップアウトする度合いが大きい母音では PV では F3 が上昇し F4 が下降する傾向が見られたが，F1，F2 の上昇は前舌母音では見られなかった。

◆５母音の声道断面積関数を PV と NV で比較し（Fig.2），モーフィングした結果，咽頭，特に喉頭蓋付近を拡大させると F3 が上昇，F4 が下降することが明らかとなった。

◆F3 が上昇，F4 が下降すると，歌唱フォルマントと同様に，聴覚感度の高い 3.0 kHz 付近のパワーが高まり，これにより音声がポップアウトする可能性が示唆された。

	F1	F2	F3	F4
F01/a/	-5.26%	-3.59%	3.64%	-2.82%
F02/a/	18.14%	6.73%	8.81%	-9.29%
F02/o/	7.69%	7.81%	9.88%	-7.99%
F03/a/	10.51%	3.45%	4.61%	-7.77%
F04/o/	9.09%	10.24%	6.06%	-1.03%

Table.1 The change rate of F1-F4 in PV compared to NV

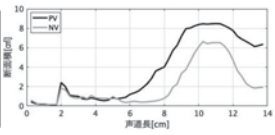


Fig.1 vocal tract area functions

2-4-12

2-4-12　発話明瞭度に差のある話者の舌運動の比較

Comparison of tongue movements between two speakers with different speech intelligibility

☆小松梨華，竹本浩典（千葉工大），能田由紀子，前川喜久雄（国語研）

◆発話明瞭度が高い話者 F01 と低い話者 M02 がイソップ童話の「北風と太陽」を朗読したときの舌運動を予備的に比較検討した。

◆舌上の正中面に前から T1，T2，T3 の 3 つの WAVE センサを約 1 cm の間隔で設置し，100 Hz サンプリングで位置を記録した。

◆Fig.1 は文中に出現する「・・した」の/ta/における 6 回分の各センサの軌跡である。発話明瞭度が高い F01 では，発話明瞭度の低い M02 に比べて発話ごとに各センサの軌跡ばらつきが大きかった。また，F01 の方が/ta/における基本周波数のばらつきも大きかった。

◆この結果は，発話明瞭度の高い話者では表現が多様であることを示唆する。逆に言えば，多様な表現が可能であることが発話明瞭度の高さを裏付けている可能性がある。

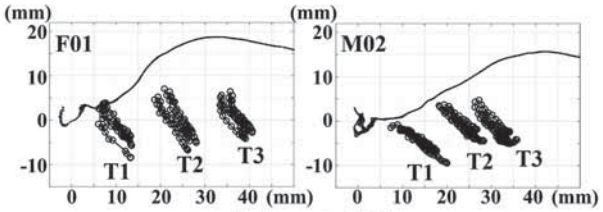


Fig.1:「した」の/ta/の舌運動

2-4-13

2-4-13　自己の録音音声の印象評価に個人特性が与える影響の分析

Impressions of One's Own Recorded Voice are Associated with Personal Traits

◎柳田耀，井島勇祐，俵直弘（NTT）

◆自己の録音音声が聞き手自身に与える影響についてポジティブな反応 [Peng+, 2020] とネガティブな反応 [Holzman+, 1966] の両方が示されている.

◆反応の違いには聞き手の個人特性（年代，性別，性格など）が影響することが示唆されており [Peng+, 2020]，前報 [Yanagida+, 2023] では，自己の録音音声の魅力と親近感に聞き手の個人特性（性格，習慣など）が影響することが示されている．しかし，魅力と親近感以外の印象に対する個人特性の影響は明らかになっていない.

◆本稿では，自己，他己音声の印象に関する大規模な主観評価実験を実施し，自己の録音音声の印象（自分らしさ，自信，聞き取りやすさ）と複数の個人特性との関係性を分析する.

◆分析の結果，自己の録音音声の聴取習慣など複数の個人特性が自己の音声の印象に影響していることが示された.

Fig.1:Top five effect sizes for similarity ratings

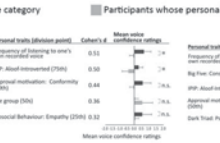


Fig.2:Top five effect sizes for confidence ratings

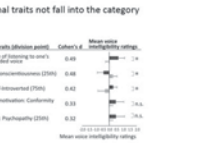


Fig.3:Top five effect sizes for intelligibility ratings

2-5-1

2-5-1　音質評価指標を用いた人工音の感覚的快さの検討

Study on sensory pleasantness of artificial sound with sound quality metrics

◎磯山拓都，△谷口亮太郎，上江洲安史，木谷俊介，鵜木祐史（北陸先端大）

◆【問題意識】Aures の感覚的快さのモデル（Aures のモデル）が人工音の感覚的快さを正確に説明できるのか不明である

◆【目的】Aures のモデルを用いて人工音の感覚的快さを正確に説明できるかを検討し，音質評価指標が感覚的快さにどのように影響するかを明らかにする

◆【方法】

① 音質評価指標の異なる音刺激の感覚的快さを評価

② 主観評価値と Aures のモデルによる推定値との相関分析

③ 各音質評価指標と感覚的快さの関係の調査

◆【結果】

① ラウドネスが小さい音は感覚的に快いと評価していた

② Fig.1(a)，Fig.1(b)より Aures のモデルは，再適合することで主観評価値を十分に説明できた

③ 感覚的快さの推定においてラウドネスとラフネスが重要であった

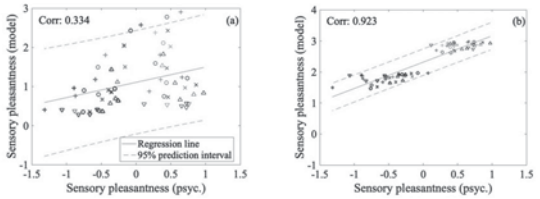


Fig.1:Estimation results of sensory pleasantness with Aures' model: (a) original model and (b) refitted model.

2-5-2

2-5-2　音声のポジティブな感性印象に関する質問紙の作成

Development of questionnaire about positive emotional impressions of voice

◎林大輔，△田端里絵子（日本たばこ産業）

◆声を聞いた時に「よい声だ」とポジティブな感性印象を抱きうることは，日常的な感覚としては自明なことのように思われる。

◆しかし，幅広く多面的な要素を含む声のポジティブな感性印象を適切に表現・評価できる質問紙は存在していない。

◆そこで本研究では，音声のポジティブな感性印象に関する質問紙を作成した。

◆まず事前調査において音声のポジティブな感性印象に関する表現語の収集と整理を行った。

◆続けて，事前調査で整理した 33 語の表現語を用いて，実際に音声を聞いた際の印象について調べる実験を行った。

◆得られたデータについて探索的因子分析を行い，基準にしたがって項目を削除した結果，以下の 4 因子 14 項目の質問紙が作成できた。

【あたたかさ因子】

落ち着く，安心感のある，心地よい，あたたかみのある，やわらかい

【明瞭さ因子】

ハキハキした，通る，さわやかな，元気な，透き通った

【かわいさ因子】

愛嬌のある，かわいい

【独特さ因子】

耳に残る，特徴的な

2-5-3

2-5-3　ロコミ投稿者の属性情報が消費者の購買意欲に及ぼす影響：文章と音声提示による比較

The impact of reviewer attributes on purchase intentions: A comparative study of text and speech word-of-mouth

◎長野瑞生，井島勇祐，廣谷定男，小林のぞみ（NTT）

◆ロコミ投稿者の属性開示が感情の媒介効果に与える影響

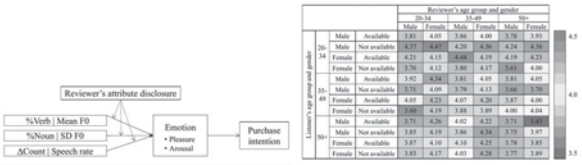
- インターネット販売におけるロコミや、ロコミと共に開示された投稿者の属性情報は他の購入者の購買意欲に影響する[Chan+, 2017]
- 先行研究ではロコミ文章に焦点を当てており、ロコミ音声は未検証
- “宣伝”音声の場合、音声の購買意欲への影響を感情が媒介[Nagano+, 2023]
- ロコミ音声の場合は音声そのものから投稿者の年齢や性別を推測出来る
- ロコミ文章と異なり投稿者属性の開示が購買意欲に影響しない可能性

◆ロコミ文章・ロコミ音声を用いた投稿者の属性開示の影響の検証

- 投稿者属性を男女各 8 名×3 年代（計 48 名）とする。ロコミ文章 12 文を ChatGPT で生成。ロコミ音声 576 刺激（48 話者×12 文）を音声合成で作成
- 参加者を投稿者属性の開示あり群/なし群に分け、ロコミ文章・ロコミ音声を聴取。感情と購買意欲を回答させ、属性開示の影響を調査

◆投稿者の属性開示の感情の媒介効果への影響を調整媒介分析で検証

- ロコミ音声では投稿者属性の開示によって感情の媒介効果が調整されない
- 投稿者属性開示と快感情の関係は聴取者が 35 歳以下の場合に傾向が異なる



Tab. 1：仮説モデル

Fig. 1：投稿者属性開示と快感情の関係

2-5-4

2-5-4　改良した MUSHRA 法に対する符号化音声を用いた検証

Validation using coded speech for an improved MUSHRA method

☆俣野 文義，小口 純矢，森勢 将雅(明治大)

◆ TTS の品質が向上したことで MOS による評価が困難に

◆ MUSHRA 法に条件を追加 → **リファレンス・アンカーを削除**

◆ MOS っぽく使える MUSHRA 法を提案（**Taut-MUSHRA 法**）

- **AST で先行公開中！（doi: 10.1250/ast.e24.34）**

◆ 本研究では Taut-MUSHRA 法の **MP3 圧縮に対する検出能力**を調査

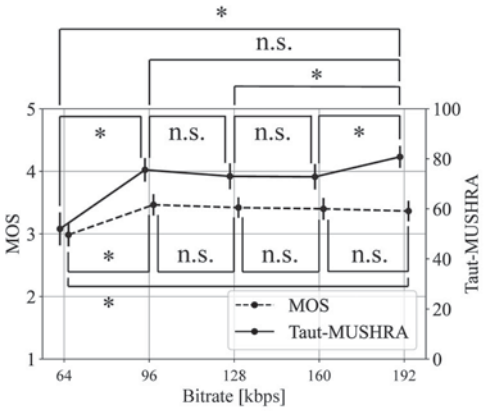


Fig.1: Results of subjective evaluation. Error bars are 95%CI.

2-5-5

2-5-5　発話と楽器演奏における遅延聴覚フィードバックの影響の比較

Comparison of the effects of delayed auditory feedback on speech and instrumental performance

○廣谷定男，持田岳美（NTT CS 研）

◆発話と楽器演奏はいずれも聴覚フィードバックを利用している。

◆発話は自らの声を聞くことで学習が行われる一方で、ピアノの訓練の際は指を見て、指の動きに注意を払って学習が行われるため、聴覚フィードバックの重要性は発話の方が楽器演奏よりも大きい可能性が考えられる。

◆しかし、発話と楽器演奏における聴覚フィードバックの影響を直接比較した研究はほとんどなく、この重要性が異なるかについては明らかになっていない。

◆本研究では、単音節「ぱ」の発話および単音「ド（C4）」の電子ピアノ演奏において遅延聴覚フィードバック（DAF）を行い、その影響を比較した。

◆実験の結果、発話は楽器演奏よりも聴覚フィードバックの影響を受けやすいことが分かった。

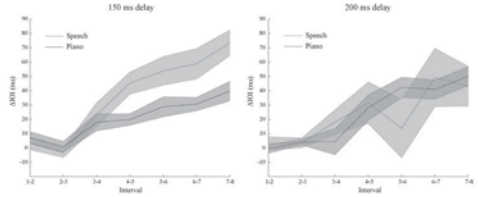


Table.1：Experimental results.

2-5-6

2-5-6　話者性を操作した遅延聴覚フィードバックが日本語の発話リズムに与える影響の検討

Study on the effects of delayed auditory feedback with manipulated speaker characteristics on Japanese speech rhythm

○上江洲安史，赤木 正人，鵜木 祐史（JAIST）

◆本研究は、フィードバック音声の話者性が遅延聴覚フィードバック（Delayed auditory feedback: DAF）下における発話リズムに及ぼす影響を、モーラ区間長に基づく発話リズム指標を用いて調べた。

◆日本語母語話者 12 名を対象に、fo・スペクトル伸長圧縮操作を与えた遅延量の異なる DAF 下での発話実験と、発話音声のアノテーションおよびモーラ区間長に基づく発話リズム指標による評価を行った。

◆遅延は全体のモーラ数と 100 ms 以上のモーラの増加をもたらした。また遅延量の増加は発話リズム指標の増加をもたらし、モーラ区間長の平均やばらつき、変動の大きさの増加への関与が示唆された。

◆モーラ区間長の平均やばらつきに対して、遅延量 0 ms では話者性の変化がこれらの増加に寄与し、遅延量 200 ms では遅延量が支配的であることが示唆された。またモーラ区間長の変動の大きさに対して、話者性の効果は遅延量で逆転することが示唆された。

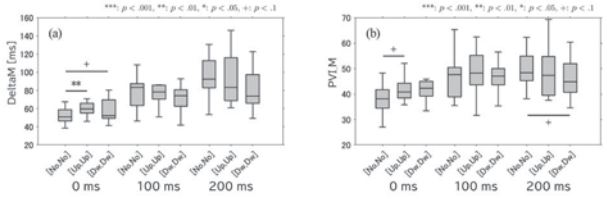


Fig.1: Speech rhythm index for each condition of delay and speaker characteristics: (a) DeltaM [ms] and (b) PVI_M.

2-5-7

2-5-7　Auditory neural activity as the basis for temporal processing

Shuji Mori (Kyushu University, Dept of Informatics, Faculty of Information Science and Electrical Engineering)
△ Hyunsoo Cho (CRI Middleware Co. LTD)
○ Willy Wong (Kyushu University, Dept of Informatics, Faculty of Information Science and Electrical Engineering)

It is clear that neural activity has a role to play in the processing of temporal information in the auditory system. What is not clear is the exact role.

We have been developing a model based on peripheral auditory activity which can give a unified account of gap detection, duration discrimination and time perception at short intervals. This model is based on the statistics of the neural response and can account for a number of experimental findings or results including: (1) The shape and qualitative characteristics of gap detection psychometric functions; (2) The role of critical bands in determining size of auditory gap; (3) How temporal discrimination varies with duration. Fig. 1 shows the prediction of the model against gap discrimination data obtained experimentally in our lab in a one parameter fit.

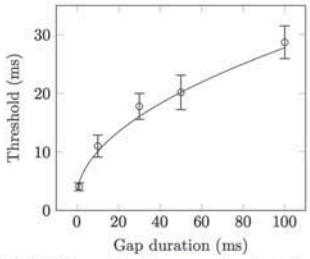


Fig.1：Data points show gap discrimination by 13 subjects with standard error in mean.

2-5-8

両耳分離聴時の選択的注意効果を ABR・MLR・LAEP 間で比較する

Comparison of selective attention effects on human ABR, MLR, and late AEP with a high temporal resolution during a dichotic listening task

○池田一成（東京学芸大）

◆聴覚誘発電位（AEP）の出現は潜時の早い順番に，聴性脳幹反応（ABR），中間潜時反応（MLR），後期 AEP（LAEP）となる。従来の有力な見解は，**選択的注意効果**が MLR もしくは LAEP の N1 波以降の成分で成立するとみなしている。一方，**聴覚系皮質遠心路**が脳幹部に接続することを前提に，選択的注意効果が ABR に成立することを示唆する知見も存在する。

◆本研究は**両耳分離聴**時の AEP を参加者 18 名から計測し，MLR と LAEP だけでなく ABR IV/V 波の段階から選択的注意効果の成立を支持する知見を得たので報告する。

◆両耳分離聴の手続きで，左耳へ低い周波数音（0.5 kHz 標準音と 0.6 kHz 標的音）を，右耳へ相対的に高い周波数音（1 kHz 標準音と 1.2 kHz 標的音）を提示した。標的音の確率は単耳で 1 または 10%にした。参加者の課題として，左耳標的音検出，右耳標的音検出，視覚刺激検出の 3 つを設けた。全 AEP を共通のサンプリング周波数（10 kHz）で記録した。波形の引き算から 3 つの課題差を求めた（聴覚注意 - 聴覚非注意，聴覚注意 - 視覚注意，聴覚非注意 - 視覚注意）。

◆分散分析において課題差の主効果は一連の 8 区間（ABR/MLR の IV/V，No/Po，Na，Pa，Nb，LAEP の Nd 前，初期 Nd，後期 Nd）に見られた。効果量（η_p^2）は LAEP の後期 Nd で最大（0.802）になり，MLR の Pa（0.437）と Na（0.410）で続く大きさになった。

2-5-9

2-5-9　聴覚による身体近傍空間の方位依存性の検討

Effects of approach direction of sound stimuli on the peripersonal space representation.

◎寺岡諒（室蘭工大），△黒田尚輝，△寺本渉（熊本大）

◆身体近傍空間（peripersonal space; PPS）は，身体のすぐ周囲の空間で，他者や周辺環境と相互作用するうえで重要な役割を果たしている。

◆本研究では，身体の前後左右から聴覚刺激を接近させ，身体周囲の身体近傍空間の様相を計測した。

◆実験の結果，身体近傍空間の様相が背後（180°）と前方（0°）で異なることを示した。

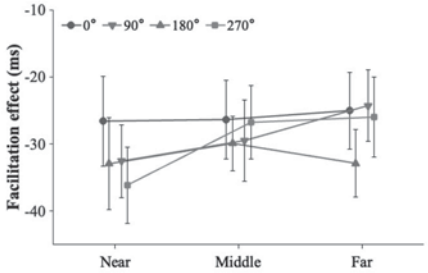


Fig. 1: Average RT difference (probe RT – baseline RT) across all listeners at each distance in each direction condition. Error bars denote the standard error of the mean.

2-6-1

2-6-1　系列変換型声質変換モデルにおける単調アライメント探索の改良

Improvement of Monotonic Alignment Search in Sequence-to-Sequence Voice Conversion Models

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

◆ 非自己回帰型の系列変換型声質変換モデルでは教師アライメントを作成する Alignment Module の精度が十分でないため、複雑な問題ではアライメントが取れず、十分な変換が出来ないという問題があった。

◆この原因の一つは、従来の Alignment Module は TTS タスク用に調整されたものをそのまま用いているからである。

◆そこで本稿では、Alignment Module の VC タスク用への調整、また Alignment Module の入力であるアライメント特徴量についても、より最適な特徴量があるかを調べる。

◆VC タスク用に調整し、アライメント特徴量にメルスペクトログラムを用いることで変換精度が向上することが分かった。

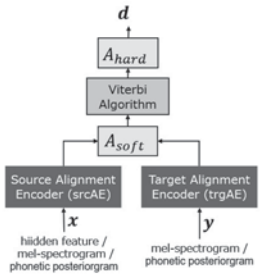


fig1: Alignment Module Architecture

2-6-2

2-6-2　Digital Audio Effect for Simulating Perceived Sound of One's Voice

「自分の声」を再現するデジタルオーディオエフェクト

〇Yoshiki Sakai, Julián Villegas（会津大学）

◆ Voice confrontation is a known phenomenon where people feel uncomfortable while hearing their recorded voice.

◆ We explored whether several generic Head-Related Impulse Responses, including those from a Head and Torso Simulator (HATS), a binaural microphone (CS-10EM), and a spherical head (SH) simulation could be used for improving this condition.

◆ A subjective evaluation suggests that generic HRTFs are insufficient for our purpose

◆ Independently of subjective ratings, these HRIRs are provided as a real-time digital audio plugin.

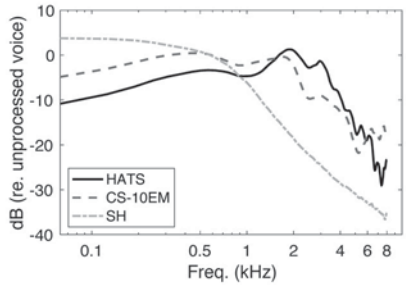


Fig.1 : Long-Term Average Spectrum of speech filtered with acquired HRIRs

2-6-3

2-6-3　プロ声優が発話した様々な発話スタイルの統計解析

Statistical analysis of various speech styles uttered by a professional voice actor.

☆蟹江世莉奈，俣野文義，小口純矢，森勢将雅（明治大）

◆近年の音声合成ではこれらの伝統的な感情ではなく，様々な発話スタイルによる合成が可能になる→**発話スタイルを決定づける特徴量の導出を目指し，その第一歩として基礎的な音響特徴量による発話スタイルを解析**

◆発話スタイル６種類に対して基本周波数、スペクトル重心、フォルマント周波数を算出

◆namidame の基本周波数がおよそ **140Hz 高い**

◆上記３つの音響特徴量での高精度な分類は困難

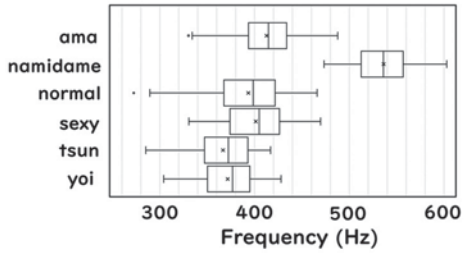


Fig.1 : Results of fundamental frequency analysis for each speech style. The results are calculated for all vowels.

2-6-4

2-6-4　時変複素 AR 音声分析に基づく GCI 推定の検討

A Study on GCI estimation based on TV-CAR speech analysis.

〇舟木慶一（琉球大）

音声信号は、声門の開閉により生成された周期性を持つ声帯音源が声道を通過する際、その形状により周波数特性が付与されて生成される。声帯音源の閉鎖点は GCI(Glottal Closure Instance) と呼ばれ、GCI の正確な推定は音声処理における大切な技術となっている。GCI 推定として T.Drugman らの方式がよく使われている。この方式は音声信号の区間平均の零交差点により GCI と GOI(Glottal Open Instance) の存在区間を推定し、区間内で LPC(Liner Predictive Coding:線形予測) 残差が最大値を取るサンプルを GCI、GOI と推定する方式である。音声ではなく LPC 残差を用いるのは、音声から声道特性を除去した声帯音源のみの特性を持つ信号を対象とするためであり、声道の特性がより適切に除去される方がより精度の高い推定が行うと考えられる。一方、我々は解析信号に対する時変複素 AR(Auto-Regressive: AR) 音声分析の研究を行っている。解析信号の性質により低周波数領域の推定精度が向上するため、声道特性がより効果的に除去できるので、時変複素 AR 分析により得られる複素 AR 残差を用いることにより GCI と GOI の推定精度向上が期待できる。本稿では、時変複素 AR 分析を用いた GCI、GOI 推定方式の提案を行う。

2-6-5

2-6-5　マルチコーパス学習に基づくドメインロバスト音声感情認識の検討

A study of domain-robust speech emotion recognition based on multi-corpus learning

☆林崎由，能勢隆，駒形晃太，伊藤彰則（東北大）

◆本稿では，実応用において汎用的に利用できるドメインロバストな音声感情認識の実現を目指し，クロスコーパスドメインにおいて，複数の異なるコーパスによるマルチコーパス学習によりドメイン汎化性能の向上を試みる．

◆音声感情認識では，感情の表出方法や言語などのドメインが学習時と評価時に異なる場合に，認識性能が低下することが課題である．

◆従来は学習と評価のコーパスが同じ within 条件での評価が多く，コーパスが異なるクロスコーパス条件での性能低下についての議論が十分とは言えない．

◆本稿では，コーパスの多様性や関係性について分析を行い，多様な学習データの追加により汎化性能が向上することを示す．

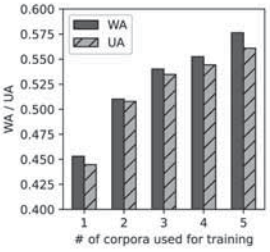


Fig. 1 Relationship between the number of corpora and cross-corpus emotion recognition performance

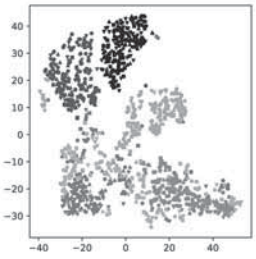


Fig. 2 WavLM-large features visualized by t-SNE

2-6-6

2-6-6　感情認識ベンチマーク構築に向けた人間・機械の感情認識能力の比較分析

A comparative analysis of human and machine emotion recognition capabilities for building an emotion recognition benchmark

☆駒形晃太，能勢隆，林崎由，伊藤彰則（東北大）

◆本研究では，感情認識の実社会応用を広げるための感情認識ベンチマーク構築に向け，人間と機械の感情認識能力の比較・分析を行った．

◆アノテータの過半数を得た感情を正解とする「聞き手ラベル」と，発話者の意図を正解とする「話者ラベル」を別々に算出した．

◆自己教師あり学習特徴量を利用したモデルとの比較の結果，学習・評価が同一の within 条件でも，ほとんどの場合に人間の感情認識能力が未だ上回る結果となった．

◆特に少数の感情クラスが存在する IEMOCAP では人間の感情認識性能が認識モデルを大きく上回っているため，実環境においては人間の方が高性能だと考えられる．

◆今後はベンチマーク構築のため，多様なコーパスに対して基準となる人間による感情認識率をアノテータにより算出し，クロスコーパス条件でのベンチマークとなるデータセットを構築する．

Table 3　IEMOCAP(9 クラス)、RAVDESS(8 クラス)、CREMA-D(6 クラス) における人間と認識モデル (within 評価) の結果．ただし IEMOCAP は認識モデルの評価では disgust の除いた 8 クラス分類

		IEMOCAP		RAVDESS				CREMA-D			
		UA	WA	UA		WA		UA		WA	
		聞き手		聞き手	話者	聞き手	話者	聞き手	話者	聞き手	話者
人間	A	-	-	.760	.572	.769	.566	.677	.407	.703	.400
	V	-	-	.797	.688	.815	.690	.750	.577	.762	.574
	A+V	.721	.756	.837	.766	.848	.772	.764	.634	.775	.631
モデル	A	.348	.505	.545	.498	.551	.515	.611	.680	.654	.680
	V	.222	.302	.661	.575	.646	.587	.648	.605	.667	.605
	A+V	.318	.492	.650	.684	.678	.687	.742	.734	.788	.734

2-6-7

2-6-7　CNN/LSTM/Transformer に基づく単発咳識別手法の性能比較

Performance Comparison of Single Cough Recognition Methods Based on CNN/LSTM/Transformer

○村岡拓実，△池田賢二（セコム）

◆身体状態を把握する手段の 1 つとして，咳やいびきなど非言語音声の発声状態の変化を捉えるアプローチがある。本研究では，このうち疾患の重症化の目安として利用される咳に着目し，その発生回数を捉えることを目指している。

◆咳は「笑い」や「咳払い」といった他の非言語音声との区別が難しいが，深層学習の出現で従来技術よりも性能が高まってきている。単発咳の回数を捉えるためには，凡そ 400 ミリ秒とも言われている短時間データでの識別率の向上が肝となる。

◆咳や笑いなど６種類の非言語音声を収録した VOCALSOUND を利用し，単発咳の回数把握の目的に照らして充分短い時間長データでの「単発咳」と「その他非言語音声」の識別性能比較を行った。

◆深層学習の代表的な複数手法（CNN/LSTM/Transformer）を比較し，LSTM が単発咳識別に有効であることが確認できた。

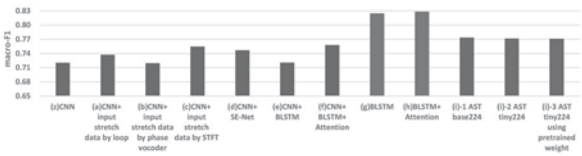


Fig.1 : performance comparison results of single cough recognition methods based on CNN/LSTM/Transformer

2-6-8

2-6-8　スペクトログラムの自己注意重みを特徴量とする吃音検出モデル

Stuttering detection based on self-attention weights of spectrogram as features

☆宮原紘造、加藤恒夫、田村晃裕（同志社大院）

◆吃音の症状は，音素や単語の反復，音の延長，発話途中の長いポーズ，フィラーの挿入として現れる．特定の音素に現れるなどの音響的特徴よりも，時間的な構造に特徴がある．

◆従来の吃音検出モデルは音声認識用の特徴量を用いていたが、吃音音声データが大量にない現状において，訓練データの不足に直面する．

◆本研究では，音響的特徴にかかわらず吃音のイレギュラーな時間構造を抽出するため，スペクトログラムの自己注意重み特徴量を提案する．

◆約 28,000 の発声からなる英語吃音音声データセット sep-28k を用いた評価で，提案の特徴量を用いた吃音検出モデルは，振幅スペクトルを特徴量に用いるモデルよりも優れた検出精度を示した．

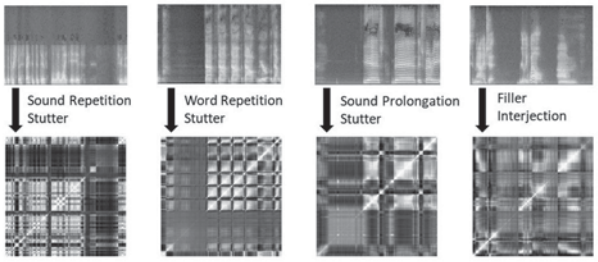


Fig. 1 : Spectrograms of stuttered speech (top row) and their corresponding self-attention weight features (bottom row)

2-7-1

2-7-1　トランスオーラル技術を用いたステレオ音源の受聴方式に関する一検討

A Study on a Listening Method for Stereo Sound Source Using Transaural Reproduction System

○藤井優美，村山好孝（シーイヤー），渡邉祐子（東京電機大・デザエ）

◆受聴者の耳元に左右のステレオ信号を再現する"ヘッドホン音場"を、トランスオーラル技術を用いた再生方式について、スピーカ受聴とヘッドホン受聴との聞こえ方の違いを明らかにする実験を行った。その結果、提案法は振幅パンニングのゲイン係数が最大の時、他の受聴方式と比較して知覚する音像が外側に広がることが明らかになった。

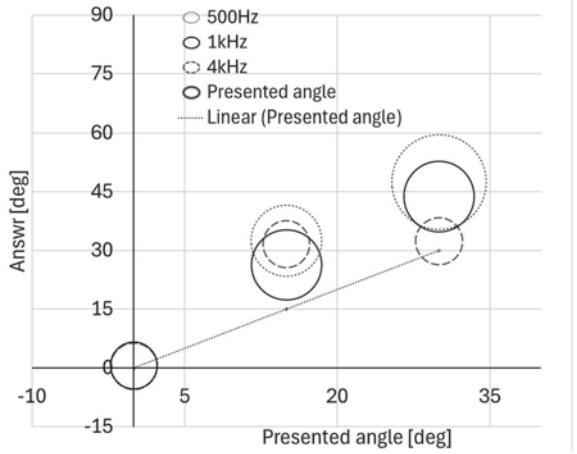


Fig. 6: Angle at whitch stereo sound images is localized in the proposed method.

2-7-2

2-7-2　PNP HRTF モデルのノッチ・ピークレベルの最適化

- 2 種類のヘッドホンを用いた音像定位精度の検証 -

Optimization of notch and peak levels of PNP HRTF model
- Verification of sound image localization accuracy using two kinds of headphones -

◎中村風香（千葉工大・院），加古達也，渡邊悠希，千葉大将，
野口賢一（NTT），飯田一博（千葉工大・先進工）

◆PNP モデルで生成した個人化 HRTF の音像定位精度の向上を目的として，ノッチ・ピークレベルを最適化した．また，その有効性を確認するために音像定位実験を行い，以下のことを示した．

◆従来モデルと提案モデルの定位精度を比較すると，水平面では両者ともに概ね目標方向付近に定位した．一方，正中面では正面方向において提案モデルに定位精度の改善がみられた(Fig.1)．

◆２種類の FEC ヘッドホンの間に顕著な定位精度の差はみられず，平均定位誤差は水平面で 18°程度，正中面で 27°程度であった(Fig. 2)．

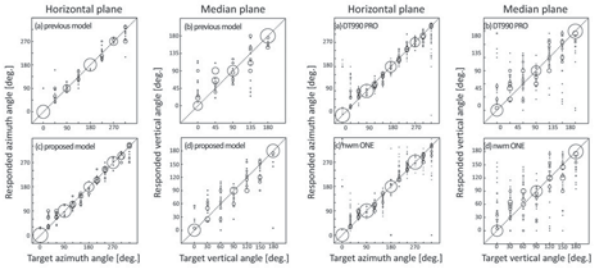


Fig. 1 Results of localization tests in the horizontal and the median planes using previous model ((a) and (b)) and proposed model ((c) and (d)).

Fig. 2 Results of localization tests in the horizontal and the median planes through DT990 PRO ((a) and (b)) and nwm ONE ((c) and (d)).

2-7-3

2-7-3　近似した HRTF と高速 1bit 信号の畳み込みの提案

Convolution of approximated HRTF and high speed single-bit signal

☆櫻井意織，△五味佑太，△野口文也，坂口智弘，及川靖広（早大理工）

◆高速 1bit 信号に対して加算・乗算のような演算処理を行う際は，一度高速 1bit 信号をマルチビット信号に変換し、マルチビットでの処理をすることが多い。

◆処理の簡略化に向け，高速 1bit 信号に対して演算クロックの高速化をすることにより、高速 1bit 信号のまま直接的にミキシングする手法が提案されている。

◆本稿では，その手法を応用して高速 1bit 信号と近似した HRTF の畳み込み処理を行い，高速 1bit 信号のバイノーラル再生を実現した。Fig.1 は畳み込み処理の概要図を示す。

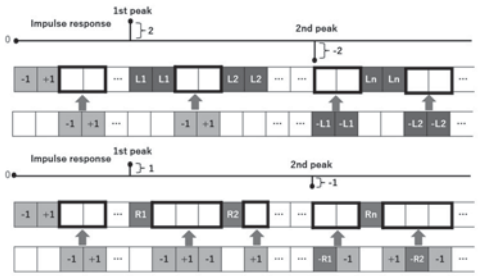


Fig.1：高速1bit信号の畳み込み演算処理の概要図

2-7-4

2-7-4　既存音声卓の機器制御プロトコルを用いたオブジェクトベース音響のライブ制作システムの開発

Development of an object-based audio live-production system using a control protocol for existing audio consoles

◎久保 弘樹，岩崎 泰士，大出 訓史（NHK）

◆オブジェクトベース音響（Object-based audio: OBA）に対応した番組制作設備の開発を進めている．これまでに開発した，OBA に対応していない既存のデジタル音声卓と音響メタデータ送出装置とを組み合わせて OBA のライブ制作を可能とする OBA 簡易ミキシングシステムは，フェーダやパンナの操作値を取得するためのインタフェースに課題があり，変換装置を介して接続する必要があった．

◆そこで音響メタデータ送出装置を，昨今盛んなリモート制作において用いられる音声卓等の機器制御用プロトコル，Ember+に対応させることにより，Ember+に対応している音声卓でさえあれば変換装置を必要とせず，音響メタデータ送出装置から任意の音声卓の操作値を XML 形式で直接取得できる，OBA ライブ制作システムを開発した．

◆Ember＋では必要な XML の要素のパスを指定することで必要最小限のデータを転送することが可能で，XML 形式としては比較的軽量な処理を実現できる．処理遅延の評価のため，音声卓のパンナを操作した値が音響メタデータ上の音声オブジェクトに反映されるまでの遅延量を測定した結果，従来システムと同程度の遅延量（17～21 ms）となり，実用に供することが分かった．

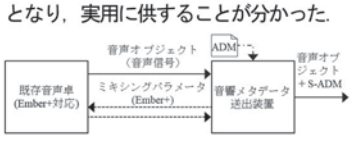


Fig.1：OBA live-production system

```
<Root Type="ElementCollection">
  <Parameter Path="1.1.2.3.4">
    <Contents>
      <Value Type="REAL">
        Real(-3.000)
      </Value>
    </Contents>
  </Parameter>
</Root>
```

Fig.2：Sample of Ember+ message

2-7-5

2-7-5　音源オブジェクトの放射特性取得における マイクロホンの指向性の影響

Effect of microphone directivity in acquiring the radiation characteristics of a sound source object.

◎富塚ゆみ子，木下光太郎，松井健太郎，中山靖茂（NHK）

◆任意視点映像に対応する音場を表現するため，マイクロホンアレイを用いて観測した音圧から音源オブジェクトの放射特性を推定・取得することを検討している。

◆音圧の観測に単一指向性マイクロホンを使用した際に，マイクロホンの指向性が放射特性の推定結果に及ぼす影響について検証した。

◆Table 1 に示す条件で放射特性を推定した結果，単一指向性マイクロホンの持つ周波数ごとの指向性が，推定結果の誤差に繋がることを確認した（Fig. 1）。

Table 1：Conditions for estimating radiation characteristics.

音源オブジェクトの放射特性		音源オブジェクトの位置 [m]	
Source A	全指向	Position 1	$(x, y, z) = (0, 0, 0)$
Source B	単一指向	Position 2	$(x, y, z) = (1, 1, 0)$
		Position 3	$(x, y, z) = (2, 2, 0)$

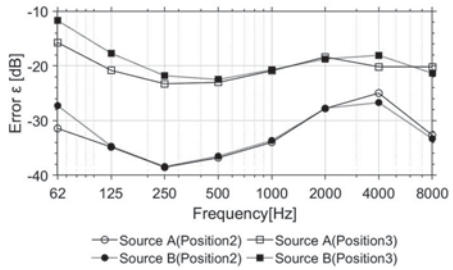


Fig. 1：Errors in estimated radiation characteristics. (Each error for Position 1 is $-\infty$)

2-7-6

2-7-6　3D オーディオにおける中層と上層の 相関の知覚について

On the perception of middle and top layer correlations in 3D audio

○亀川徹，丸井淳史（東京芸大）

◆3D オーディオにおける上下方向の相関と聴取印象との関係について調査した。

◆水平方向における広がり感が左右の信号の相関係数と関連があるのに対して，左右が無相関のノイズを上下方向の相関係数を変えて再生した場合，上下方向の広がり感と相関係数との関係は見られなかった。

◆一方で上下方向の相関係数が低くなることで，中央の聴取位置と印象が変わらない聴取範囲が広がることが示された。

◆150～200Hz 以上の周波数帯域において上下の相関係数が低ければ，上下が無相関の場合と同じ程度の聴取範囲が得られることが示唆された。

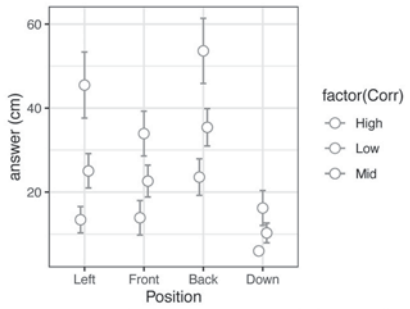


Fig.1：Differences in perceived distance and vertical correlation values with the central position for each direction of movement. Circles indicate mean values, and error bars indicate 95% confidence intervals.

2-7-7

2-7-7　近距離場音響ホログラフィ法の 境界音場制御の原理への適用による 収録システムの小型化 ーバーチャル BoSC マイクロホンによる 音場再現精度の改善ー

Downsizing of the BoSC microphone array by applying NAH to the boundary surface control principle -Improvement of sound field reproduction accuracy by virtual BoSC microphone-

○伊勢史郎（東京電機大・システムデザインエ）

◆研究の背景

- 境界音場制御（BoSC）の原理に基づく音場再現において再現音場のエリアが BoSC マイクのサイズよりも大きい。
- BoSC マイクの小型化が可能
- 理論的根拠として近接場音響ホログラフィ法（NAH）を適用

◆研究内容

- 上記を数値計算により定量的に確認
- NAH を用いてバーチャル BoSC マイクを構成

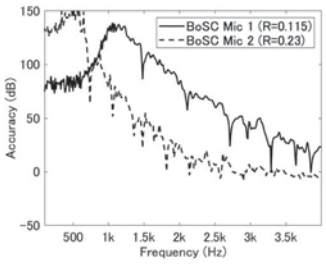


Fig. 2 Accuracy of SFR within Area1

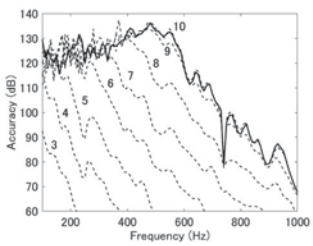


Fig. 5 Accuracy of SFR using virtual BoSC mic within Area2

2-7-8

2-7-8　球面調和関数を用いたバイノーラル 3 次元音空間収音再生法

Binaural sound-space sensing and reproduction method using spherical harmonics functions

○坂本修一（東北大 通研／院情科研）

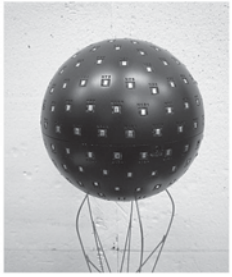
◆バイノーラル音空間収音再生法をターゲットに，アンビソニックス，高次アンビソニックス（higher-order Ambisonics：HOA）などの球面調和関数を用いた音空間表現の基本的な流れを紹介

◆収音信号を HOA エンコードし，展開係数を取得

- 剛球アレイの場合，「アレイがない時の原音場」と「アレイがあることによって生じる散乱音場」で収音信号をモデル化

◆バーチャルスピーカアレイを用いて HOA デコードし，各スピーカの駆動信号に頭部伝達関数を畳み込むことでバイノーラル信号を生成

◆システム構築時は，収録音空間を表現する展開係数の次数と，再現システムにて使用可能な展開係数の次数との対応関係を考慮



球状マイクロホンアレイの例（252ch 球状マイクロホンアレイ）

2-7-9

2-7-9　（招待講演）音響規格における高次アンビソニックスの応用例

Applications of higher-order ambisonics in audio-related standards

○杉本岳大（NHK）

◆没入感の高い 360° 映像や仮想空間を活用したメディアが身近になるのに伴い，用いる音響技術にも同様に高い没入感を提供できる能力が望まれてきている。その期待に応えられる音響技術の一つと目されているのが，高次アンビソニックス（higher-order ambisonics：HOA）であり，様々な音響規格に応用されている。

◆ITU-R 勧告では，勧告 ITU-R BS.2051 で音響コンテンツを表現する音響方式の一つとしてシーンベース音響方式が取り上げられており、それを代表する音響技術として HOA が登場する。

◆MPEG 規格では，MPEG-H 3D audio（3DA）と MPEG-I Immersive audio で HOA が活用されている。

◆MPEG-H 3DA では，放送システムなどの実装用の LC（low complexity）プロファイルのレベル 4 において，4 次の HOA まで取り扱える。

◆MPEG-I Immersive audio では，internal HOA（in-HOA）と external HOA（ex-HOA）の二つの音場の表現方法がある。

◆in-HOA は，予め範囲を規定した空間 V に対して，V の境界線の外側に位置する発音源が生成する，V の内部音場の表現である。

◆ex-HOA は，予め範囲を規定した空間 V に対して，V の境界線の内側に位置する発音源が生成する，V の外部音場の表現である。

2-7-10

2-7-10　高次アンビソニックス技術のヤマハでの活用事例紹介

Application of higher-order ambisonics technology at Yamaha Corporation.

○奥村啓、トレビーニョ ホルヘ (ヤマハ)

高次アンビソニックス（Higher Order Ambisonics，以下 HOA）技術は、音場の球面調和関数展開表現に基づく立体音響技術であり、音場の物理的な特徴量を再現することで臨場感の高い音場を創り出すことができる。ヤマハでは、HOA 技術などの立体音響技術を用いることで、音楽視聴や楽器演奏の新しい体験価値の発見・創出を目指している。本発表では、HOA 技術の活用事例として、立体音響コンテンツの制作と、楽器の空間印象評価に用いた事例を紹介する。

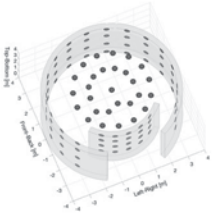


Fig. 1: スーパーサラウンドシアター：106 個のスピーカ (紫) と 220 度のスクリーン (青) により高臨場感コンテンツを体験できる。

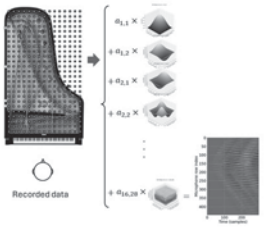


Fig. 2: 空間印象評価実験に使用した仮想振動板の音源。総モード数を設定することで異なる複雑さの音源を用いた感性実験を実施した。

2-7-11

2-7-11　3GPP IVAS 標準規格の概要

Overview of 3GPP IVAS standard

○守谷 健弘，原田 登，杉浦 亮介，鎌本 優（NTT）

◆3GPP の新たな音声音響符号化 IVAS (EVS extension of Immersive Voice and Audio Services)規格が完成。技術・品質・応用分野を紹介。

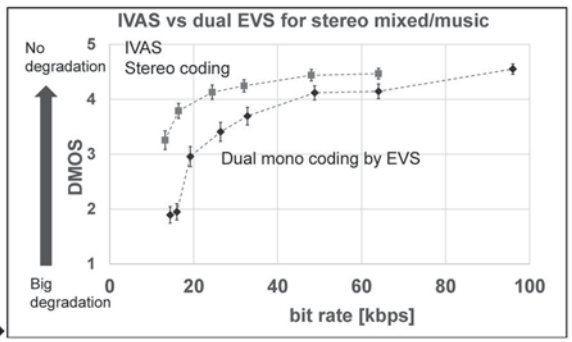
◆双方向通信が可能な世界初の立体音響符号化の国際標準規格。

◆立体音響符号化、マルチストリーム処理、レンダリング、EVS 互換

- ステレオ、バイノーラル、シーンベース（HOA）、オブジェクトベース、チャンネルベース、MASA、それらの組み合わせに対応
- フレーム長：20 ms，レンダリングも含む原理遅延：37 ms，サンプリング：48 kHz，ステレオから 7.4.1 チャネル ビットレート：13.2 から 768 kbps

◆ステレオ、多チャネル符号化は EVS によるモノラルの組合せより高品質

◆高臨場通話、対戦ゲーム、多地点会議など双方向立体音響通信用途



◆

Fig.1：Example of subjective listening test results comparing IVAS stereo coding and dual mono EVS coding for mixed/music signals

2-8-1

2-8-1　強力集束超音波照射下でのイメージングプローブ近傍の反射音場の光学的評価

Evaluation of reflected acoustic field of high-intensity focused ultrasound near ultrasound imaging probe by optical measurement method

☆小野佑真(東北大院・医工学研)、森翔平、吉澤晋(東北大院・工学研)

◆本研究では強力集束超音波の反射波の焦点が超音波イメージングプローブ近傍に発生した際の音場を評価することを目的としている。

◆反射音場の焦点をイメージングプローブ表面から超音波伝搬方向に 2 mm, 0 mm, -2 mm の位置に設定し、光学的手法の 1 つである差分シャドウグラフ法により、超音波照射時の画像から音場の圧力分布画像を作成し評価した。

◆50 ns 間隔で画像を取得し、超音波伝搬の動画の作成を行った。

◆イメージングプローブへの入射波と反射波およびそれらの合成による表面近傍での定在波成分の伝搬を確認した。

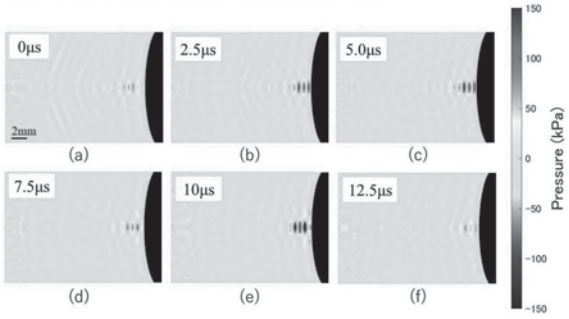


Fig.1：Pressure distribution of the reflected sound field near the imaging probe

2-8-2

バースト波支援動的造影超音波法による液体の粘性評価

Evaluation of liquid viscosity by burst-wave-aided contrast-enhanced active Doppler ultrasnography

○吉田憲司（千葉大・CFME），大村眞朗（富山大・エ），
平田慎之介，山口匡（千葉大・CFME）

◆体液の粘性を非接触に評価する方法として，音響放射力で移動する造影剤の移動速度から周囲液体の粘性係数を評価する方法を提案

◆ファントム実験により，造影の移動速度は粘性係数の増加に対して非線形的に減少することを確認（Fig.1(a)）

◆移動速度と粘性係数の関係をべき乗則（$v_z = A_i \mu^{B}$）と仮定して，係数AおよびBを評価（Fig.1(a)）

◆数値計算により移動速度と粘性係数の関係性を検証し，実験結果と比べて大きな差異がないことを確認（妥当性の確認）（Fig.1(b)）

◆べき乗則を用いることで造影剤の移動速度から粘性係数の評価が可能

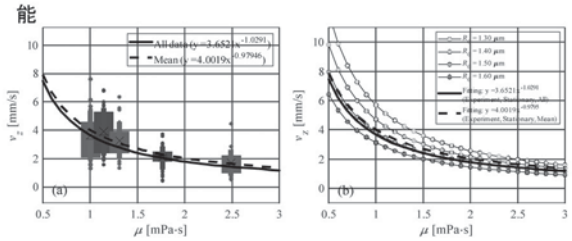


Fig.1：Relationship between translational velocity of ultrasound contrast agents and viscosity coefficient of surrounding liquid. (a) experiment and (b) numerical calculation. Bold solid and dashed lines are fitting curves obtained from the all experimental data and mean values, respectively. In (b), R_0 means initial radius of ultrasound contrast agent.

2-8-3

2-8-3　平面振動子とマトリックスアレイプローブを対向させた透過音響特性計測システム

Measurement system of acoustic transmission characteristics with opposed planar transducer and matrix array probe

○橋本 瑞紀，吉田 憲司，山口 匡，平田 慎之介（千葉大）

◆生体組織を伝搬する弾性波の速度は，その組織性状を推定できる有用な物理量であり，海綿骨内の縦波音速による骨密度計測や肝臓内のせん断波速度による肝硬度計測などに応用されている．

◆超音波の縦波音速などの透過音響特性を高空間分解能で計測・評価することができるシステムを構築し，生体組織性状診断に応用することを目的とする．平面振動子とマトリックスアレイプローブで構成される計測システムを試作し，その性能評価として，真鍮板の透過音響特性を計測した．

◆厚さの異なる真鍮板を計測し，縦波伝搬速度の算出や概形の把握が可能であることを確認した．

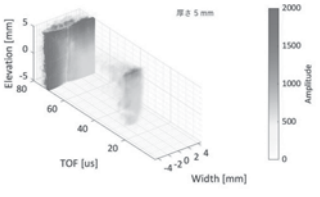


Fig.1：Amplitude distribution of ultrasonic pulse transmitted through brass plate

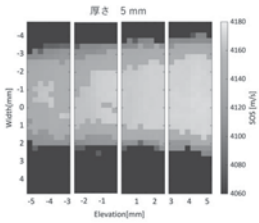


Fig.2：Longitudinal wave velocity of ultrasonic pulse transmitted through brass plate

2-8-4

2-8-4　ラット脂肪肝におけるエコー振幅統計量のステアリング角度依存性

Steering angle dependence of echo-envelope statistics in fatty rat liver

◎奈良千尋（千葉大），森翔平（東北大），吉田憲司，山口匡，平田慎之介（千葉大）

◆超音波ビームのステアリングを行うことで，同一部位を描画した異なる複数の超音波画像を作成し，その振幅統計量をコンパウンドすることで，統計的なゆらぎを低減できる．ラット肝臓を対象とした脂肪定量における提案手法の検証を行った．

◆通常肝モデルラットと，1~4 週間高カロリー食を与えた脂肪肝モデルラットの摘出した肝臓に対し，リニアアレイプローブを用いて±20°の範囲を 5°間隔でステアリングを行い計測した．

◆ファントム実験とは異なり，0°付近では 3 次モーメントが高く，角度が大きくなると低下する傾向がみられた（Fig. 1）．このため，今回は振幅統計量コンパウンドを行わず，この原因を追究した．

◆通常肝から脂肪沈着すると一度モーメントは低下し，高カロリー食の週数が増加するごとに大きくなるが，脂肪化が進むと再び小さくなった（Fig. 2）．

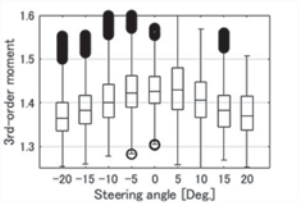


Fig.1：Box-and-whisker plot of 3rd-order moments of a normal-liver model rat.

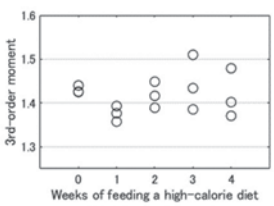


Fig.2：median of 3rd-order moments.

2-8-5

2-8-5　機械的バースト加振 SWE におけるせん断波伝搬による2次元変位の評価

Evaluation of 2-D displacement by shear wave propagation in mechanical burst excitation SWE

☆齋藤慎一郎，△小林翔，吉田憲司，山口匡，平田慎之介（千葉大）

◆せん断波伝搬速度 (SWS) の周波数特性評価手法の確立を目的とした，機械的バースト加振 (MBE) SWE を提案している．

◆高粘性高分子ファントムにおける SWS 分布 (Fig. 1) に縞模様状の揺らぎが生じる原因を解明するため，せん断波伝搬による 2 次元変位について検討を行った．

◆ブロックマッチングにより 2 次元変位を計測した結果 (Fig. 2)，①と③に比べ，②と④は上下方向の変位が小さいことが分かった．

◆②，④は位相差として計測できる上下方向変位が充分ではなく，信号対雑音比が低いため SWS に揺らぎが生じると考えられる．

◆上下方向の振動振幅 (Fig. 3) が 0.2 μm 以上となる画素で算出された SWS のみを選択した場合，揺らぎが低減することを確認した．

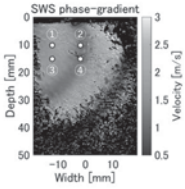


Fig. 1 SWS distribution in the gel phantom

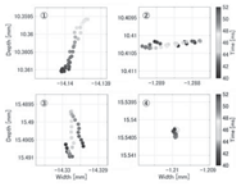


Fig. 2 2-D displacement by shear wave propagation

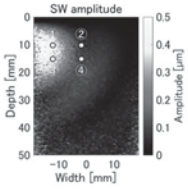


Fig. 3 Vertical shear wave vibration amplitude

2-8-6

2-8-6　超音波エコーの散乱特性を識別するＨスキャンによる肝線維化・脂肪化の定量評価

Quantitative evaluation of liver fibrosis and lipidification by H-scan that identifies scattering characteristics of ultrasonic echoes

〇小林亮太（千葉大），△Tai Dar-In，△Tsui Po-Hsiang（Chang Gung Univ.），吉田憲司，山口匡，平田慎之介（千葉大）

◆散乱体の表面が平面と考えられる場合は送信パルスに対応したエコーが、パルスの波長より十分薄い膜の場合はパルスの一階微分に，パルスの波長より十分小さな散乱体の場合はパルスの二階微分に対応した波形で近似できるエコーが得られる。

◆送信パルス・その一階微分・その二階微分を参照信号として、超音波画像の各走査線の RF 信号と最も相関値の高いものを色情報として超音波画像上に重畳させるＨスキャンを用い、びまん性肝疾患における肝線維化・脂肪化の定量評価に応用する手法の基礎的な検討を行った。

◆B モード画像とＨスキャン画像の拡大図では、テクスチャが対応していることが確認できる（Fig. 1）。

◆各色に対応するエコー振幅の振幅確率密度分布は、レイリー分布に類似しており（Fig. 2），これらの形状パラメータと各色の割合を複合的に解析することで，肝線維化・脂肪化の定量的に評価できる可能性がある.

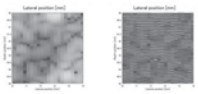


Fig.1 B-mode and H-scan images in cases with fibrosis stage F4 and 0% fat content

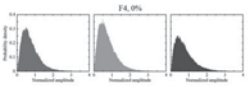


Fig.2 Amplitude probability density distribution of each RGB color in the cases in Fig. 1

2-8-7

2-8-7　超音波受信のためのカーボン・トランスデューサとその特性

Carbon-based transducer for receiving ultrasound and its characteristics

〇中村健太郎（東工大）

◆かつて電話機などに使われたカーボン・マイクロフォンの原理を超音波受信に利用することを検討した。

◆アレイ化が可能な小形な構造として，フレキシブル基板の上に厚さ 0.2 mm 程度のカーボン・トランスデューサを製作した。

◆1 素子の幅は使用した基板パターンの幅である 1.27 mm である。

◆カーボン粉は 4B の鉛筆の芯を 2000 番のサンドペーパーで削って作り，スペーサ（厚さ 0.1 mm の PET フィルム）に開けた穴に入れた。

◆上部はアルミフォイル（厚さ約 12 μm）とスコッチテープ（厚さ約 60 μm）で覆った。

◆電気抵抗は 0.5～数 kΩ であった。

◆28 kHz 空中超音波定在波（140～160 dB）により動作を確認した。定在波音場の反射面に配置した圧電センサとほぼ同位相の出力波形が得られた。

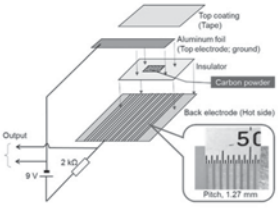


Fig.1 : Configuration of the carbon transducer.

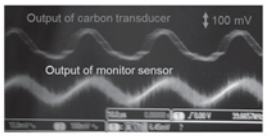


Fig. 2 : Received signal waveforms.

2-8-8

2-8-8　空中超音波による非接触型吸熱アクチュエータの基礎的検討

A basic investigation of a non-contact heat absorption actuator using airborne ultrasound.

〇松村蒼一郎，△牧野泰才，△篠田裕之（東大）

◆研究目的: 形状記憶合金など熱を駆動源とするデバイスに対し、空中超音波のエネルギーによって熱を非接触供給。

◆手法: 40 kHz で共振する共鳴型吸音デバイスによって超音波を熱へ変換し、加熱対象へ熱を供給。

◆実験: 超音波フェーズドアレイによって超音波を銅板と接触させたデバイスに収束させ、音響パワーと銅板の温度上昇からデバイスのエネルギー変換効率を推定。

◆結果: 音響パワーから熱へのエネルギー変換効率は 33%。

◆今後の展望: 吸音率の改善と実用的なアプリケーションの開発。

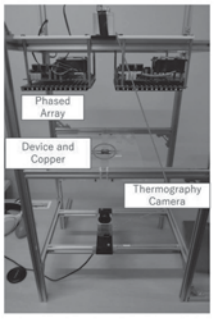


Fig.1 : Experiment Setup.

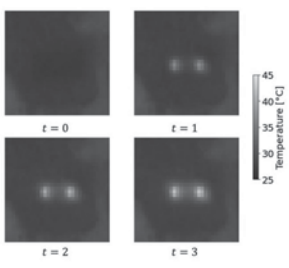


Fig.2 : Temperature rise of device.

2-8-9

2-8-9　接触インピーダンス型触覚センサの設計指針の検討

The study of design guidelines on contact impedance change type vibratory tactile sensor

〇工藤すばる（石巻専修大・理工）

本研究は，振動型触覚センサの設計の体系化を計るために，特に硬い対象物を測定する場合の設計指針について検討したものである。Fig.1 は本研究で使用した縦振動子触覚センサの一例である。Fig.1 に示すような縦振動子を試験片 (Steel (SPCC)ヤング率 206GPa) に接触させ，その時の共振周波数の変化量を測定した。Fig.2 は，異なる寸法の縦振動子を用いて作製した触覚センサの荷重 W=30gf における周波数変化率をまとめたものである。周波数変化率 $\Delta f/f_0$ は，振動子の等価スチフネス s_0 に反比例し，フィッテングにより $\Delta f/f_0 \propto s_0^{-1.37}$ と求められた。

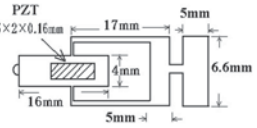


Fig.1. Construction of vibratory tactile sensor (thickness2mm, f_0=158.8kHz).

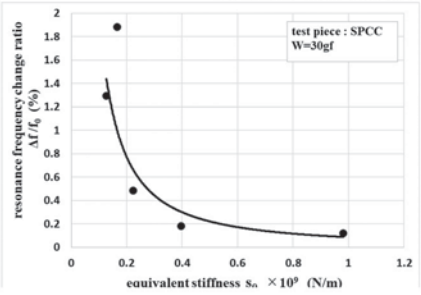


Fig.2. Measured characteristics of resonance frequency change of tactile sensor.

2-9-1

2-9-1　残響付加システムのモデルベースド制御に関する研究－室内音響指標の再現性－

Study on model-based control of reverberation enhancement system
- Reproducibility of room acoustic parameters

☆河野光貴，佐久間哲哉（東大・工）

◆筆者らは、電気音響による残響付加システムの自動調整に向けて、制御用スピーカとマイク間のインパルス応答計測に基づくモデルベースド制御の検討を進めている。

◆YAMAHA 製 Active Field Control Enhance (AFC)を使用し、今回は初期反射音制御部に限定して AFC 内の信号処理を計算機上で模擬した。

◆AFC 作動時に発生するフィードバックループを反映した評価用スピーカ・マイク間の伝達関数を、逆行列を用いた周波数領域の計算式で表現することでシミュレーションモデルを構築した。

◆AFC 作動時に測定した評価用 RIR とシミュレーションにより再現した RIR をもとに室内音響指標の比較検証を行い、再現性を確認した。

◆EDT については測定値より再現値が長くなる傾向が、ST_{Early} については AFC 内で設定するスピーカゲインや FIR フィルタ(条件 A,B)によらず良い対応が確認された。

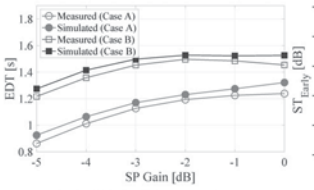


Fig.1：EDT with changing the loudspeaker gain

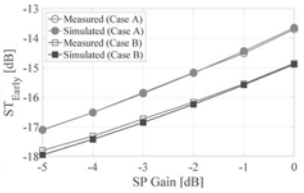


Fig.2：ST_{Early} with changing the loudspeaker gain

2-9-2

2-9-2　音場支援システムを用いた目標音場の創生手法

― FIR フィルタ設計による残響特性の制御 ―

Creation method of target sound field using acoustic enhancement system: control of reverberation characteristics through FIR filter design

◎大木 大夢，橋本 悌，渡辺 隆行，宮崎 秀生(ヤマハ)

◆音場支援システムはマイクとスピーカ間の音響フィードバックを用いて空間の音響エネルギ密度を向上させ，音響条件を制御するシステムである。

◆システムに FIR フィルタを適用することで残響特性の可変幅が向上することが知られている。本研究では，任意の残響特性を創生するための FIR フィルタの設計を検討した。

◆フィルタ設計にはホワイトノイズを用いた。ホワイトノイズをオクターブバンドごとに分割しそれぞれ処理することで，自然な減衰をもつ FIR フィルタを設計することができる。

◆本発表では FIR フィルタの設計手法を提案し，実測結果をもとに提案手法の有効性を考察する。

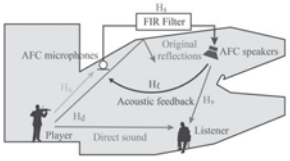
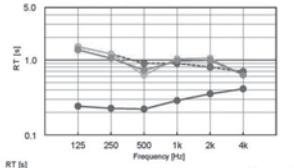


Fig.1: Basic model of AFC system



RT [s]

Frequency [Hz]	125	250	500	1k	2k	4k		250-2kHz Average
Original sound field	0.24	0.23	0.22	0.29	0.35	0.41	●	0.27
Target value	1.50	1.20	0.90	0.90	0.80	0.70	●	0.95
Numerical analysis	1.51	1.20	0.63	1.04	1.05	0.63	○	0.98
AFC ON measured value	1.36	1.05	0.73	0.96	0.99	0.63	●	0.93

Fig.2: Frequency characteristics of reverberation time

2-9-3

2-9-3　小学校の普通教室における音環境の調査 –教室の音環境改善に向けた予備的検討–

Investigation of classroom acoustics in ordinary classrooms at a primary school in Japan: A preliminary study towards improving classroom acoustics

○エバンズ直子(TOA(株)/大阪大学，△Wang Yutan(明治大学)，
△清野健(大阪大学)，上野佳奈子(明治大)

◆小学校の普通教室における音環境の実態を明らかにするため、ICT 機器やデジタル教材を導入している小学校 3 年生の普通教室 3 教室にて、授業中の音環境と教室の音響性能の計測を行った。

◆1960 年代後半に建築された 3 階建て RC 造校舎の典型的な片廊下型教室で行われた、国語・外国語・音楽の授業中に等価騒音レベルを、児童下校後に室内騒音、残響時間、および STIPA を測定した。

◆その結果、教室内は室内騒音、室内残響の両面で推奨条件を満たしていないこと、授業活動に伴い発生する音により耳への悪影響が懸念される騒音レベル(Fig. 1)となっていることが確認された。

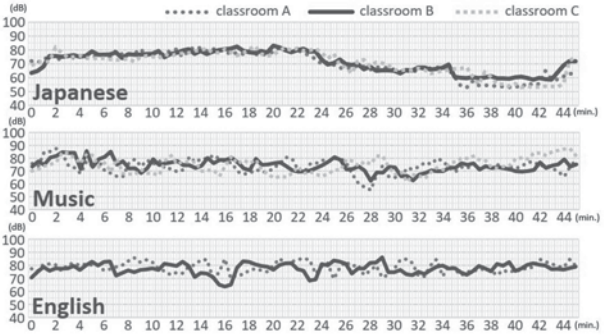


Fig.1：Equivalent continuous A-weighted sound pressure Level ($L_{Aeq,30s}$) during lessons in three classrooms at a primary school in Japan.

2-9-4

2-9-4　発達支援を行う保育空間における吸音カーテンの効果

Effects of sound-absorbing curtains in childcare spaces that support children's development

☆栗山秀一郎，上野佳奈子(明治大)，野口紗生(浜松学院大・ADC)，
松本知子(浜松市根洗学園)

◆児童発達支援センターにおいて，吸音性能をもつ 2 種類のカーテンを設置した結果を報告する。リトミック等の集団遊びや行事、入園前の親子教室が行われるプレイルームを対象室とした。

◆Tab.1 に示す 8 条件で残響時間の簡易測定を行った。全て取り外した条件 A では 1.22 秒であるのに対して，2 種類のカーテンを両方閉めた条件 G では 0.72 秒と短くなった。

◆職員 24 名を対象に，吸音カーテン設置前後における対象室の印象についてアンケート調査を行った。結果として、ほとんどの評価項目で有意差が見られ、室全体の吸音性能を高める手段として有効であること、音の聞こえ方の変化に加えて、光の透過性のあるカーテンと組み合わせることで視環境的な効果もあることが示唆された。

Tab.1 Reverberation time measurement conditions /results (500 and 1k Hz)
(―:removal, ×:closed, ○:opened)

Conditions	Blackout curtain (before)	Sound absorbing curtain (after)	Sound absorbing lace (after)	Reverberation time [s]	Average Sound absorption coefficient
A	―	―	―	1.22	0.12
B	×	―	―	1.15	0.13
C	○	―	―	0.86	0.17
D	―	×	×	0.95	0.16
E	―	×	○	0.86	0.18
F	―	○	×	0.73	0.21
F'	―	○	―	0.74	0.20
G	―	○	○	0.72	0.21

2-9-5

2-9-5　保育室におけるカームダウンのための小空間の設置実験 —室内および小空間内の吸音条件と利用度との関係—

An experiment on the installation of small calming spaces in childcare rooms - Frequency of use with internal sound absorption

○川井敬二，津留芽実（熊本大院）

◆幼稚園・保育園では多数の幼児が一緒に活動し、空間が喧騒になることがある。その喧騒が苦手な子が「避難」して落ち着くための小空間を、実際の保育室に設置する現場実験を実施した。

◆実験では５歳児１クラスを対象に、吸音仕上げのある保育室とない保育室で各約２ヶ月間、内部の吸音なし／あり（吸音材4mm厚）／あり（吸音材25mm厚）の３種類の小空間を設置した（Fig. 1）。小空間内には荷重によりスイッチが入るマットセンサーを敷き、小型コンピュータとWifiを使って自動観察するシステムを構築し、観察した。

◆結果として、吸音仕上げのある保育室ではほとんど使用されていなかった小空間が、吸音のない保育室に移動後に使用時間が長くなり、内部吸音のある小空間の方が使用時間が長い、という結果が見られた。なお内部吸音の効果は、1~2 kHz帯域で、吸音なしと比べて4 mm厚では1 dB程度、25 mm厚では4 dB程度であった。

◆知見はまだ不十分であり、さらなる検証が望まれる。

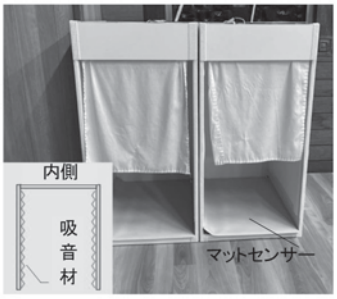


Fig.1：Small calming space

2-9-6

2-9-6　小規模オフィスにおけるスピーチプライバシーの測定事例(2) 対策例の効果

A case study on measurement of speech privacy in a small office: (2) Examples of countermeasures and their effects

○原行浩（OTO 技研），佐藤逸人（神戸大院・工学研）

◆2021 年１月にスピーチプライバシーの評価規準と設計指針-音声による情報漏洩防止-（日本建築学会環境基準 AIJES-S0003-2021）が刊行された。

◆本稿では，小規模オフィスに一般的な対策例を施した場合について，上記規準に準拠した測定及び評価を行った。

◆音源近傍を囲むコの字型ブースを設置し，さらにブース内吸音とブース天井設置の対策を行うことで，上記規準記載の情報漏洩のクラス 3rd を達成できる場合があることを示した。

Table 1 SPL of speech at each measurement point (LS2) (unit: dB)

	No countermeasure	With booth (One side open) +Ceiling +Absorption (inside) +Absorption (wall)
P1	55.6	46.3
P2	48.8	40.2
P3	45.4	36.9
P4	43.2	35.6

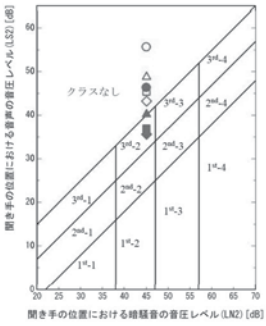


Fig. 1 Rating chart on AIJES-S0003-2021

No countermeasure： P1○, P2△, P3□, P4◇
Booth (one side open): P1●, P2▲, P3■, P4◆

2-9-7

2-9-7　マルチスピーカ拡声時の SNRstat による STI 予測と室内音響設計

STI prediction using SNRstat for multi-speaker public address system and room acoustic design

○増田　潔（大成建設技術センター）

◆設計時に統計室内音響理論で予測可能な SNRstat を開発し，それを基に STI に対する最適残響時間曲線式を導出した。

◆SNRstat はマルチスピーカ拡声についても，一つ一つのスピーカの指向特性やディレイ，パワーレベル，室内の平均吸音率等を考慮して任意の聴取点における STI 予測が可能である。

◆実空間において複数の聴取点にマルチスピーカ拡声時の STI 予測を行った結果は，実測値と良い対応を示した。

◆STI 実測値の平均値と STI に対する最適残響時間曲線式の対応には当初乖離があったが，暗騒音やスピーカのカバーエリアを考慮すれば良い対応を示すことが分かった。

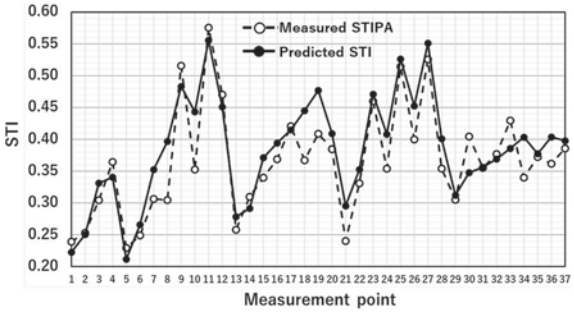


Fig.1：Measured and Predicted STI for multi-speaker PA system.

2-9-8

2-9-8　多言語同時拡声による案内放送の了解度(4) 目的音声の到来方向に対する注意の影響

Intelligibility of multi languages simultaneous broadcast (4) Influence of attention to the direction of arrival of the target voice

☆大井丈実，佐藤逸人（神戸大院・工学研），森本政之（神戸大）

◆本研究では，多言語同時拡声システムの性能向上を目的として，目的音声が再生されるスピーカを事前に視覚サインで示し，聴取者の注意をその到来方向に向けさせることで単語了解度が上昇するかについて，バーチャルリアリティ（VR）技術を用いた実験より検討した。

◆日本語，英語，韓国語を同時に再生し，聴き取れた日本語の単語を VR 空間内に表示されるキーボードで回答させた。目的音声の再生位置，空間分離の有無，視覚サインの有無を実験パラメータとした。

◆視覚サインの提示により，目的音声と妨害音声の両者が同方向から到来する場合に，単語了解度は統計的に有意ではないがやや上昇した。また，視覚サインの提示により，目的音声の到来方向による単語了解度のばらつきを抑制できる可能性を示した（Fig.1）。

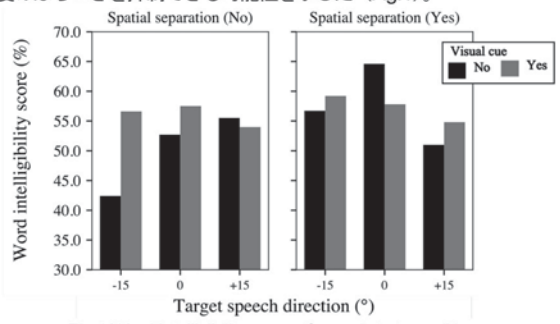


Fig.1 Word Intelligibility scores for each test condition.

2-9-9

2-9-9　風及び地表面が屋外拡声音の伝搬特性に及ぼす影響

Effects of wind and ground surface on the propagation characteristics of outdoor acoustic mass-notification systems

☆中村達哉，佐藤逸人，（神戸大院･工学研）　栗栖清浩，松石遼太（TOA）

◆防災用屋外拡声システムからの拡声音の空気吸収以外の要因による超過減衰（ΔL”）について，長期観測の結果を風と地表面が及ぼす影響に着目して分析した。

◆ベクトル風速（スピーカとマイクロホンを結ぶ線に平行な風速の成分）とΔL”の関係は測定サイトによって異なり（Fig. 1），特に追い風が強くなった場合の影響に違いが見られた。地表面の傾斜，周囲の土地利用状況，スピーカの指向性など様々な要因の影響が示唆された。

◆また，無風条件におけるΔL”の周波数特性も測定サイトによって異なり，平坦となる場合と周波数が高くなるほどΔL”が大きくなる場合が確認された（Fig. 2）。

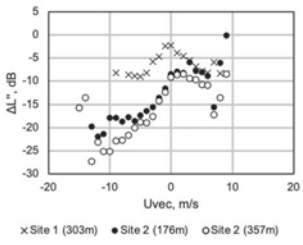


Fig. 1　ΔL” as a function of Uvec (1/3 Oct. band, cf = 2kHz).

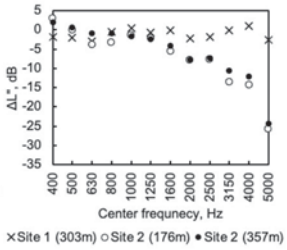


Fig. 2　ΔL” in no wind condition (1/3 Oct. band).

2-9-10

2-9-10　災害時の避難情報取得における言語バリアフリーの実態調査

A survey of language barrier-free in obtaining evacuation information during disasters

○森本侑希，佐藤逸人（神戸大院･工学研）

◆災害発生時などの混乱した状況では，外国人被災者は情報を得る手段が相対的に少なくなり，情報弱者となる。そこで本研究では，情報取得の手段が限られる移動時に着目し，在日外国人を対象として避難情報の取得に関する実態についてアンケート調査を実施した。

◆日本語，英語，韓国語，中国語の質問文を用いて Web アンケート形式で調査を行い，171 名からの回答を得た。

◆旅客施設では，日本語能力が低い場合でも日本語の音声案内はある程度聞き取れていることが示された。その一方で，回答者の第一言語の音声案内でも「ほとんど聞き取れる」と回答した割合は６割弱にとどまり，音響的要因や放送内容の要因が聞き取りの妨害原因となっていることが示された。

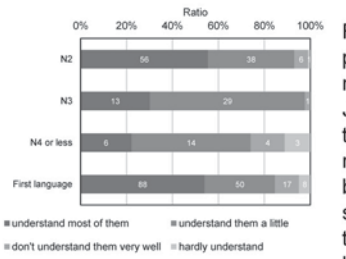


Fig. 1 Respondents' impression on Japanese announcements broadcast at Japanese passenger facilities aggregated by Japanese language ability. The bottom bars show the same impression but for those broadcast in their first language.

2-9-11

2-9-11　コンサートホールにおける遅れ時間の長い反射音の音響効果に関する基礎的検討 －その１　音に包まれた感じと異なる音響効果の検討－

Basic study on acoustic effects of long-delayed reflected sound in a concert hall –Part 1 Investigation of subjective effects that differ from listener envelopment–

◎河合茜音（日大･院），羽入敏樹，鈴木諒一，星和磨（日大･短大）

◆本報では，遅れ時間の長い反射音が，ASW や LEV とは異なる音響効果を有するか否か検証するため，聴感実験を実施した（実験１）。また，遅れ時間の長い反射音による音響効果が，エコーディスターバンスとならず好ましく感じることがあるのか否か検討した（実験２）。

◆遅れ時間の長い反射音によって何らかの音響効果が生じ，それが ASW と LEV とは異なる可能性が示唆された（実験１）。遅れ時間の大きい反射音を付加することで，エコーディスターバンスを避けながらコンサートホールの聴衆に好ましさを感じさせる可能性を示した。（実験２）

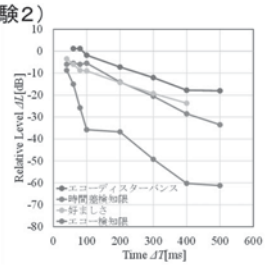


図１　遅れ時間の長い反射音による各種聴覚イベントの検知限

2-9-12

2-9-12　コンサートホールにおける遅れ時間の長い反射音の音響効果に関する基礎的検討 －その２　中期反射音の提案とその音響効果－

Basic study on acoustic effects of long-delayed reflected sound in a concert hall – Proposal of the mid-term reflected sound and its acoustic effects

○羽入敏樹（日大･短大），河合茜音（日大･院），鈴木諒一，星和磨（日大･短大）

◆本報では，遅れ時間の長い反射音による時間差知覚による聴感印象について検討するため心理実験を実施した。

◆遅れ時間の長い反射音は必ずしもロングパスエコーのようなエコーディスターバンスとならず，ASW，LEV とは異なる「音に立体感があり，余韻に動きを感じ，響きに透明感を感じる印象（印象 A）」をリスナーに供給する可能性があることがわかった。この「印象 A」は，新たに定義した 80 ms から 300 ms もしくは 500 ms 程度の時間領域である「中期反射音」の時間構造によってもたらされる。

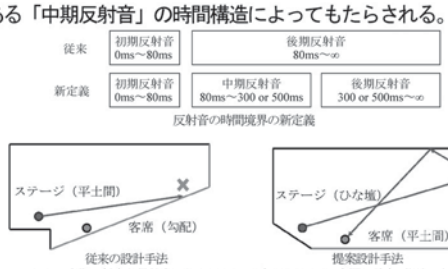


図７　中期反射音を供給するためのシューボックスホール音響設計案（縦断面）

◆今後，「印象 A」をもたらす「中期反射音」を制御する新たな視点によるコンサートホールの音響設計手法の可能性を検討していくつもりである。

2-9-13

2-9-13 パノラマ写真により VR 表示された建築空間の視覚情報に調和する残響時間の提案

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

☆千葉妙，石川あゆみ（岐阜高専）

◆ 我々の研究室では、人が建築空間の視覚刺激（見た目）から予想した響きを“予想残響時間（以下 ERT）”と定義し、これを被験者実験で測定してきた。

◆ 先行研究では、パノラマ写真により VR 表示された建築空間の視覚刺激に調和する残響時間を求めるモデルとして、実験における対象空間の容積とその ERT の散布図が提案された。

◆ 本報では、新たな対象空間に対して実験を行い、その結果を先行研究で提案されたモデルに追加して分析を行った。追加データも含めたモデルを Fig.1 に示す。なお、T_{60} は各空間の実際の残響時間を示す。

◆ 対象空間を増やした結果、同じ容積の空間では T_{60} よりも ERT の方が短く、容積が大きくなるにつれ T_{60} と ERT の差が開いていくという傾向が見られた。

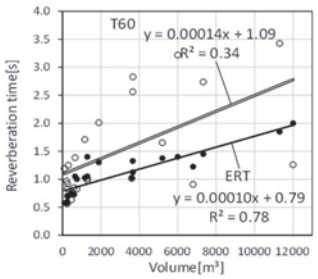


Fig.1 : Proposed Model (this and previous research)
(○ : T_{60}, ● : ERT)

2-9-14

2-9-14 パノラマ写真により VR 表示された建築空間の視覚印象に調和する残響時間の提案

－ 被験者の実験への慣れが視覚印象と予想残響時間に与える影響 －

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

☆岩佐茉音歌，石川あゆみ（岐阜高専）

◆パノラマ写真により VR 表示された建築空間に対する、ERT（予想残響時間）の同定実験と視覚印象の主観評価実験によって得られるデータの、信頼性を高めることが目的である。

◆過去に同実験を受けたことのある経験者と、受けたことのない未経験者の実験結果を比較することで、被験者の実験への慣れが、実験結果に与える影響について検証した。

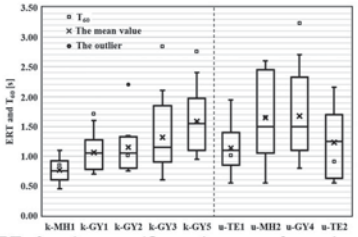


Fig.1 ERT of each space (Group A, group of experienced subjects)

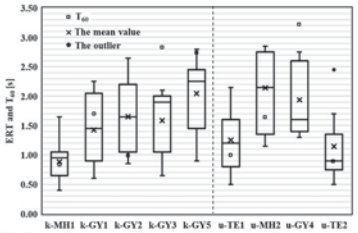


Fig.2 ERT of each space (Group B, group of unexperienced subjects)

2-9-15

2-9-15 音線法による室内音場の簡易バイノーラル可聴化に関する研究 －VR 提示による聴感実験－

Study on a simple binaural auralization system for room acoustics using a ray tracing simulation - Psychoacoustical experiments using VR presentation

☆大林 紅音（東大・工），劉 金雨（神奈川大・建築），佐久間 哲哉（東大・工）

◆音源系・受音系の指向性を反映した音線法による室内音響解析に基づき、ヘッドホンを用いた簡易バイノーラル可聴化手法を提案した。

◆会議室内の音声伝送を想定して、話者・受聴者位置の水平 4 方向の指向性の再現性を検証し、提案手法の妥当性を確認した。

◆可聴化システムの聴感実験では、実空間と HMD による VR 空間において実音と再現音の比較評価を行い、再現音の違和感を検証した。

◆遠方の座席位置では再現音の音量が小さく感じられ、座席方向が音源に対して横向きでは音像がやや後方かつ明瞭になる傾向が見られたが、全体的に再現音の印象に対する VR 映像提示の影響は小さかった。

Fig.　実験時の視点イメージ（左：実視野、右：仮想視野）

2-9-16

2-9-16 縮尺模型実験に基づくオーケストラ演奏の 3 次元可聴化 その 1 立体音場再生のための要素技術

3-D Auralization of Orchestral Performance Based on Scale Model Experiments, Part 1: Elemental Technologies for 3D Sound Field Reproduction

○鈴木和憲，日高孝之，小柳慎一郎（竹中技研）

◆コンサートホールの音響設計の実施・詳細設計段階で，縮尺模型実験が広く行われている．解析結果の可聴化は，ソロや小アンサンブルなどの比較的単純な構成の音源に限られていた．その主な理由は，音源，空気吸収補正，収音方法などの技術的制約や課題が残されていたためと考えられる．本報告では可聴化を目的としたレーザー誘起音響パルスの適性について検討を行った．

◆レーザー誘起音響パルスの瞬時音圧レベルは，エネルギー400mJ で近傍では 150 dB を超えることから非線形性を有し，オーバーオールのピーク値については，逆二乗則により大きく減衰することが報告されていた．

◆本報告では 1/3 オクターブバンド音圧レベルの距離減衰特性を測定及び空気吸収減衰補正を行い，概ね音源から 1m 以上離れると，逆二乗則に沿った結果が得られることを確認した．

◆時間・周波数特性を測定し，直接音成分の逆フィルタリングにより，模型実験の室内インパルス応答の測定に十分な時間分解能，周波数特性を有することを確認した．

◆レーザー音響パルスは，信号処理を施すことで可聴化用の模型用音源として十分な測定精度を確保できることを確認した．

2-9-17

2-9-17　縮尺模型実験に基づくオーケストラ演奏の３次元可聴化 その２　オーケストラ演奏の再生事例

3-D Auralization of Orchestral Performance Based on Scale Model Experiments, Part 2: Example of Reproduction of Orchestral Performance

鈴木和憲，○日高孝之，小柳慎一郎（竹中技研）

◆縮尺模型内の音場を立体的に収音するには多チャンネルマイクロホンでの測定が必要であるが，超音波まで使用できるアンビソニックスマイクロホンが無いため，受音点を中心とした半径 r の仮想的な等価球面上に 1 個の小型マイクロホンを移動させて室内音響インパルス応答（RIR）を測定した．

◆オープンな球形アレイマイクロホンと同様，低音域の SN 比向上のためには球半径が大きい方が望ましく，空間的なエイリアシングを避けるにはマイクロホンの間隔を小さくする必要があることから，１つの等価球半径で測定できる周波数帯域は限られる．このため，球半径を変えて RIR を測定し，適切に組み合わせることで，可聴域全体をカバーする４次のアンビソニックス信号を合成した．

◆1/10 縮尺模型ホールを用いてオーケストラを模擬するため，ステージ上の演奏者位置からレーザー誘起音響パルスを放射させて，受音位置で測定した RIR から 4 次のアンビソニックス信号を合成し，無響室録音を畳み込んで音場で再生した．再生音は音の立ち上がりが良く定位も良好であることを確認した．

◆縮尺模型実験によるオーケストラ演奏の可聴化技術について検討し，所期の品質を有する合成音を達成した．さらに高い音質を得るための新たな課題が明らかになった．

2-9-18

2-9-18　インパルス応答の空気吸収・分散補正への解析的アプローチ

Analytical Approach to the Air Absorption and Dispersion Correction of Impulse Responses

○小柳慎一郎，山田祐生，鈴木和憲，日高孝之(竹中技研)

◆温湿度および模型実験スケールの変更に伴うインパルス応答の空気吸収および分散の変化を補正する提案済み手法に対し，新たに解析的アプローチを提案する．

◆ISO9613-1 に従う空気の吸収係数とそのヒルベルト変換から得られる伝搬定数を周波数の解析関数とみなすことで，これまで数値的な評価が必要であった発散する振動積分を解析的に実行できた．これにより時間的または周波数的に近似する必要のあった空気吸収補正を厳密に実施できる手法が得られた．高周波・長距離伝搬の補正は従来法と同程度に発散するが，音速の変化に起因する位相変化も補正可能であり可聴化に利用するインパルス応答へ適用できる．

◆これまで音波の伝播距離と吸収係数の差の積として計算されていたが補正量は実験系 h と補正系 r の伝播定数が等しい条件 $\gamma_h(\nu) = \gamma_r(\omega)$ の４つの解の一つとして得られることが明らかとなった．

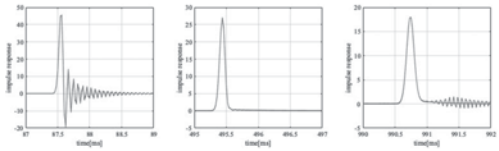


Fig.1：Results of converting the calculated impulse response of a 1/10 scale model obtained at 25°C, 75% RH to the full-scale impulse response at 20°C, 50% RH using this method(blue lines). The red lines represents the exact solutions obtained through direct calculation.

2-9-19

2-9-19　室内音場の狭帯域分析について －帯域外モードの影響－

Narrow-band analysis of sound fields in rooms - Effect of out-of-band modes-

○山田祐生(竹中技研)

◆室内音場の過渡（成長・減衰）過程では，入力信号の断続にともない，室の音響モードの関与により，正弦波で励振した場合にはその周波数以外，帯域信号で励振した場合には励振帯域外の成分も観測される．

◆これら帯域外モードの影響を定量的に把握する方法について検討を行った．過渡過程のエネルギースペクトルについて定式化を行い，それより周波数統計量，帯域エネルギー比を求めて評価する方法を示すとともに，実際の室内音場の分析に適用を試みた．

◆その結果，定常状態から減衰過程に移る際の周波数シフト，帯域の拡がりにより，音響特性を特徴づけられる可能性が示唆された．

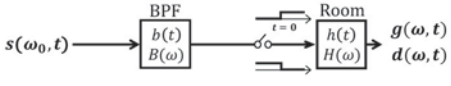


Fig.1：Acoustic measurement system.

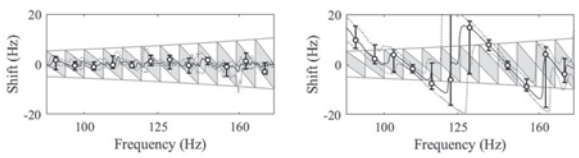


Fig.2：Median frequency and effective band [f_p, $f_{1\text{-}p}$] (p=1/4) of decay process in a chamber music hall (left) and a meeting room (right).

2-9-20

2-9-20　超低周波音源を用いた外部建具のがたつき閾値計測

Measurement of threshold value on rattling of facade elements by using low frequency sound source

○杉江聡，鈴木肇，岩永景一郎，土肥哲也(小林理研)

◆窓やドア等の外部建具のがたつきを抑制することを目的とし，純音によるがたつき閾値（がたつきが始まる又は止まる音圧レベル）の計測方法を提案した。

◆音源として，１ m× １ mのアルミハニカム板を空圧サーボアクチュエータを用いて振動させて低周波音を発生させる装置と18インチのサブウーファを8台配列したものを用いた。

◆がたつきが生じている場合に障子上の振動加速度に高調波が生じることを利用して，がたつき閾値を判定した。

◆ある純音を放射し，音圧を徐々に上昇させた条件又は下降させた条件において，障子上の振動加速度レベル(250～1500 Hz）が急激に増加又は低下した時刻の屋外の音圧レベルをがたつき閾値とした。

◆Fig.1に，クレセントによる施錠の有無でのがたつき閾値（音圧下降条件）の比較を示す。共振による落ち込みが，施錠の有無でオクターブ以上シフトすることがわかる。

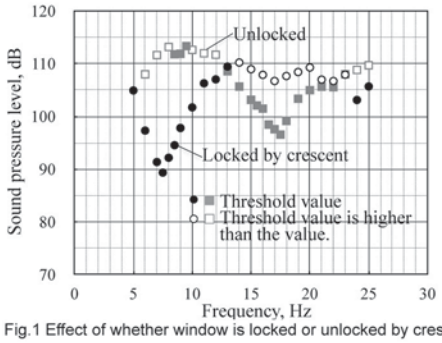


Fig.1 Effect of whether window is locked or unlocked by crescent.

2-10-1

講演取消

2-10-2

2-10-2 航空機騒音に対する他の騒音の重畳判定：CNNの学習データに含まれる機種の違いに伴う精度の差

Superimposing judgment of other noises on aircraft noise: Accuracy differences due to aircraft type variations included in CNN training data

☆小泊楓奈，森淳一，土屋健伸（神奈川大），森長誠（大同大），山元一平（防衛基盤整備協会），野澤竜祐，花香和之（成田国際空港振興協会），川瀬康彰（成田国際空港振興協会/小林理研）

◆航空機騒音に係る現場測定における、有人による各単発騒音の機種や妨害音重畳等の識別を省力化するために、音響情報を用いたCNNによる航空機騒音識別を行っている

◆この種の現場測定で高い精度の機械学習識別を実現するためには、当該飛行場における学習データを用いることが望ましいものの、新規で現場測定を行う飛行場等では、他の飛行場のデータによる識別モデルの応用が期待される

◆しかし、その場合、飛行場ごとに運用される航空機の構成は異なるため、識別精度が低下する可能性がある

◆本報では、複数の飛行場で得られた学習データによって各々の識別モデルを生成し、各モデルによって他の飛行場の騒音データを識別した

◆その結果、識別精度は10～20%程度低下した

Table 1 Results of 10-fold cross validation

True Label \ Predicted Label	air	gray	others	N
air	741	51	8	800
gray	70	626	104	800
others	10	85	705	800
N	821	762	817	2400

Acc : 86.3%

Table 2 Results of another model

True Label \ Predicted Label	air	gray	others	N
air	540	203	57	800
gray	92	330	378	800
others	40	55	705	800
N	672	588	1140	2400

Acc : 65.6%

2-10-3

2-10-3 欧州連合における環境騒音指令の発令の経緯に関する研究　－欧州連合及び関連諸機関による文書の分析－

Study on the history of the issuance of the Environmental Noise Directive in the European Union: analysis of the documents from the European Union and related organizations

○大嶋拓也（新潟大），平栗靖浩（近畿大），原田和典（岡山県立大）

- 欧州連合（EU）の指令2002/49/EC（END）発令に至る経緯として，1996年の政策提案書であるFuture Noise Policy（緑書）が知られる。
- しかしながら，それ以外の経緯はあまり知られていない。
- 本報では，経済協力開発機構（OECD），世界保健機関及びEUの文書を収集し，ENDの法案に至る経緯を分析する。
- OECDでは，1980年代には1.3億人が，65 dB以上の等価騒音レベル（L_{Aeq}）に晒される一方，データ不足による不確かさ，他の環境問題より騒音政策の優先度が低いことが指摘される。また行動計画として，騒音源への賦課金などの経済的手段，及び教育・情報開示など非経済的手段が提言されている。
- WHOには，騒音曝露による睡眠妨害等の健康影響，及びL_{Aeq}の指針値が示されている。
- 両機関の問題意識は緑書に継承された一方，ENDの法案には行動計画の内容は継承されていない。
- EU法の基本原理である補完性原則に基づき，加盟国自身による行動計画作成を求める法案となっている。

2-10-4

2-10-4 沿線観測点におけるパンタグラフ空力音の音源寄与度推定手法

Methods for estimation of source contribution of aerodynamic pantograph noise at a railway side measurement

○山崎展博，△光用剛（鉄道総研），中山雅人（阪産大），西浦敬信（立命館大），△若林雄介（JR東日本）

高速で走行する新幹線のパンタグラフから発生する空力音の低減対策を効率的に行うためには，沿線観測点に対する影響が大きい部位から優先的に対策を施すことが重要である。

本研究では，パンタグラフの部位毎の音源強さを風洞試験で測定したうえで，沿線観測点におけるパンタグラフ空力音の音源別寄与度を推定する手法を提案した。風洞試験によりマイクロホンアレイで取得した実機のパンタグラフ周りの音圧レベル分布図に対してデコンボリューション法を適用し，パンタグラフ構成部位の音源強さを推定した(Fig. 1)。次に，3次元音響数値計算により得られた地上 25m 地点における沿線観測点に伝播する際の音圧レベル変化を補正することで，パンタグラフ空力音の大きさを推定した。今回の提案手法による推定結果が，現地試験結果と一致することを確認したうえで，パンタグラフ構成部位毎の音源別寄与度を明らかにした(Fig .2)。

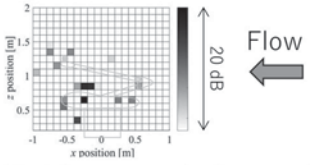


Fig. 1 Source strength of pantograph noise obtained in a wind tunnel test (1 kHz band ,270 km/h).

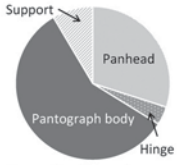


Fig. 2 Contribution of pantograph noise at a railway side observation point (1 kHz band).

2-10-5

2-10-5　縮尺模型実験による新幹線向け防音壁の形状検討

Evaluation for noise barrier shape and noise reduction effect for Shinkansen

○佐藤大悟（鉄道総研），阿部幸夫，山本健次郎，中村祐太（日鉄建材）

◆新幹線高架橋の老朽化対策として，様々な形状や構造の防音壁の取替が予定されている。取替前後で同等以上の騒音低減を実現するためには，防音壁形状と騒音低減効果の関係性を評価しておくことが必要である。そこで，防音壁形状等の違いが騒音低減効果へ与える影響を評価するため，無響室内にて縮尺模型を用いた音響試験を行った。

◆吸音材の有無は、防音壁形状の違いと比較して車両下部騒音の低減効果へ与える影響が大きい。そのため，例えば逆Ｌ型（吸音材なし）を直立型（吸音材あり）へ取替えるなど，保守が容易なシンプルな構造への変更が見込める。

◆防音壁の上部形状の違いによる低減効果の差異は，レールレベルより下方では 1dB 程度であった。そのため，実際の形状検討では騒音低減効果だけではなく強度や施工性等の両立を重視することが望ましい。

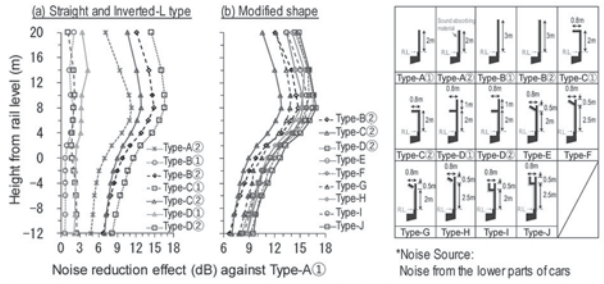


Fig.1：Noise reduction effect against Type-A① at 25m away from center of track

2-10-6

2-10-6　二重壁間の地表面が道路交通騒音伝搬に及ぼす影響　—3 次元波動解析による実用計算モデルの検討—

Effects of ground surface between double walls on road traffic noise propagation: examination of a practical calculation model using 3-D wave-based numerical analysis.

☆小林遥斗，△信田聖人（神奈川大・工），
安田洋介（神奈川大・建築），福島昭則（ニューズ環境設計）

◆道路交通騒音予測モデルである ASJ RTN-Model では，平行に設置された二重の遮音壁を二回回折する伝搬計算方法が示されているが，そこでは遮音壁間に地表面が存在する場合の影響が考慮されていない．

◆本報では，二重壁間の地表面が道路交通騒音の伝搬に及ぼす影響について，3 次元波動数値解析を通して把握すると共に，ASJ RTN-Model に基づいて地表面の影響を簡便に考慮するための実用計算モデルについて提案・検証した．

◆数値解析の結果，道路交通騒音を前提とした場合，二重壁間の地表面の影響は小さく，本検討においては 2 dB 以下に止まった．

◆本提案手法の適用により，ASJ RTN-Model をそのまま適用した場合と比べて波動数値解析結果との差が小さくなることを確認した．

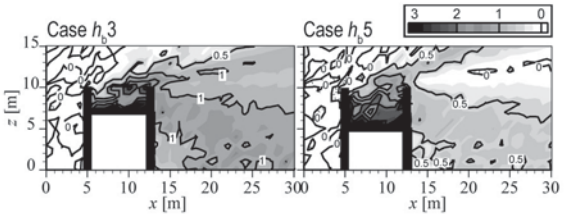


Fig. 1：Difference in L_A from a case without ground surface between walls.

2-10-7

2-10-7　高架構造の自動車専用道路における高架構造物音と自動車走行騒音の相関

The relationship between structure-borne noise and running vehicle noise on viaduct highway.

○横見瀬大地，池谷公一，糸畑 浩（高速道路総合技術研究所），
太田達也，一木智之，福島昭則（ニューズ環境設計）

◆ 高架構造の自動車専用道路を高速走行する大型車類からの高架構造物音と自動車走行騒音について，実測調査結果により両者の関係を検討した。

◆ 高架構造物音のパワーレベル $L_{WA,str}$ と自動車走行騒音のパワーレベル L_{WA} は，$L_{WA,str} = L_{WA} + a$ の関係にあった（Fig.A）。

◆ 高架構造物音も自動車走行騒音も，パワーレベルは大型車類の軸数に比例して増加する傾向が見られた。

◆ 大型車類走行時の高架構造物音や自動車走行騒音のスペクトルには，タイヤのトレッドパターンに起因すると考えられる共通の卓越成分がみられた。

◆ これらのことから，$L_{WA,str}$ と L_{WA} には，橋種と舗装をパラメータとした一定の関係があるものと推測される。

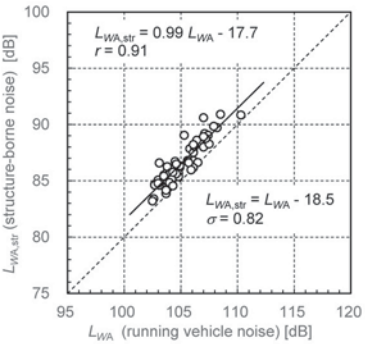


Fig.A. Relationship between $L_{WA,str}$ and L_{WA} of heavy vehicles on viaduct highway.

2-10-8

2-10-8 高い位置にある音源による建物群の挿入損失に関する縮尺模型実験 –連続移動音源の利用による実験効率化–

Scale model experiment on insertion loss of buildings in the case of high-position sound sources —Improvement of experiment efficiency by using a continuously moving sound source—

○坂本 慎一（東大生研），永田 茉莉咲，王 麒源（東大大学院），
矢野 博夫（東大生研），穴井 謙（福岡大）

◆建物群による挿入損失に関して，藤本らの考え方を踏襲し，高い音源位置の場合への拡張を目指し，縮尺模型実験を行った。

◆連続移動音源を利用することにより，実験の効率化を図り，データ取得を行った。

◆道路交通騒音および新幹線鉄道騒音の周波数特性の補正を施してそれぞれの騒音の A 特性オーバーオールレベルを求めるとともに，周波数別の挿入損失データ（中心周波数 125 Hz～2 kHz の 1/1 オクターブバンドレベル）としても整理した。

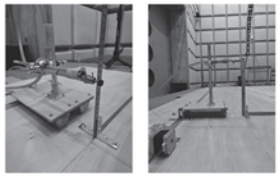


Fig. 1 Setting situation of a jet noise source mounted on an electric moving stage

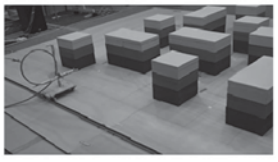


Fig.2 Setting situation of a jet noise source (3 m high) and buildings (10 m high)

2-10-9

2-10-9　橋梁版二層式排水性舗装及びＰＲＭＳ工法の騒音低減効果

Noise reduction effects of double layer porous asphalt pavement and PRMS construction method for noise control on viaducts

○萩尾淳一、名島眞樹子、赤木浩一（福北公社）、吉田元臣、海老澤尚、国生昌美（オリコン）、山内勝也（九州大･芸工）、岡田恭明（名城大･理工）

◆ ASJ RTN-Model でパワーレベル式が示されている舗装は、密粒舗装、排水性舗装、高機能舗装Ⅱ型である。福岡北九州高速道路公社は上記以外の低騒音舗装による騒音対策を検討するため、供用中の本線上に密粒舗装及び橋梁版二層式排水性舗装（橋梁に採用可能な総厚とした舗装）、ＰＲＭＳ工法等の試験舗装を舗設し、ASJ RTN-Model の予測結果を補正するような密粒舗装に対する騒音低減効果について調査した。

◆ 調査は一般車及び 60 km/h で定常走行する試験車（大型車２台、小型車２台）のＡ特性音圧レベル及び走行速度を測定した。パワーレベルの算定には最大騒音レベル法を用いた。

◆ 騒音低減効果は、密粒舗装との差分により算出し、橋梁版二層式排水性舗装で約４dB、ＰＲＭＳ工法で約３dB であった。今後３年程度は経年変化等の確認と併せて継続調査を実施する予定である。

Table.1 : **Noise reduction effects of low noise pavements (maximum A-weighted sound pressure levels)**

車種分類	騒音低減効果 [dB]	
	橋梁版二層式	ＰＲＭＳ工法
小型車類	-4.3	-2.6
大型車類	-4.1	-3.0

騒音低減効果：低騒音舗装のL_{eq}−密粒舗装のL_{eq}

2-10-10

2-10-10　2023 年に採集したアブラゼミの主鳴音のピーク周波数

Peak frequency of courtship song made by a large brown cicada in 2023

◎御園玲央，武藤憲司（芝浦工大）

◆本研究は，アブラゼミ１匹が発する各周期の主鳴音で，最も音圧レベルが大きい「ジ」音と「リ」音を抽出して周波数特性を分析した．

◆我々のこれまでの研究で，28 kHz まで自由音場とみなせる範囲で，ミンミンゼミの高潮音「ミ」音と「ン」音の周波数特性を正確に測定してきた．本研究では 28 kHz まで自由音場とみなせる範囲で，１匹のアブラゼミの主鳴音の周波数特性を測定した．

◆アブラゼミ１匹の主鳴音「ジ」音と「リ」音のピーク周波数を統計分析した．「ジ」音のピーク周波数の平均値は 5,440Hz，標準偏差は 274 Hz であった．「リ」音のピーク周波数の平均値は 14,793 Hz，標準偏差は 406 Hz であった．

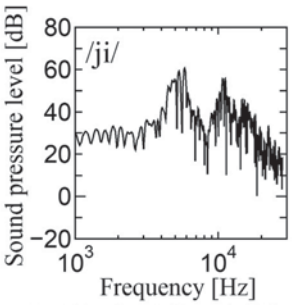


Fig. 1 Spectrum in the part of /ji/ made by a large brown cicada ($r_2 = 1$ m).

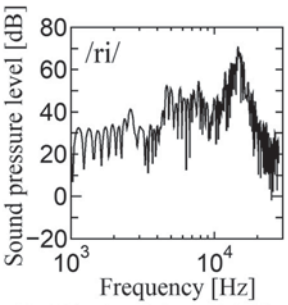


Fig. 2 Spectrum in the part of /ri/ made by a large brown cicada ($r_2 = 1$ m).

2-10-11

2-10-11　身の回りのVHF音 -VHF音の発生源調査その2-*

VHF sounds around us – A survey on VHF sound sources, Part 2-

☆青島怜生，大石まなか（神奈川工科大），廣江正明（小林理研），上田麻理（神奈川工科大）

◆先行研究では生活用品や家電製品や公共交通機関や太陽光発電施設等，身の回りの VHF 音発生源を調査してきた．

◆本研究では，身の回りの VHF 音の音源と周波数特性を明らかにする為，家電製品や交通機関などを対象に音カメラを用いて音源探査を実施し，PCM レコーダーを用いて VHF 音の計測、FFT 解析を行った．

◆ IH 調理器から VHF 音の発生が確認でき，20 kHz～25 kHz 付近と 40 kHz～45kHz 付近の 2 つのピーク周波数は約 80 dB であった．

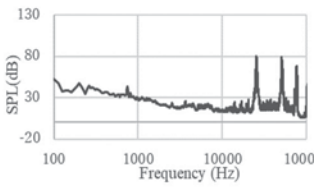


Fig. 1 Frequency characteristics of VHF sound before boiling observed on an IH cooking heater.

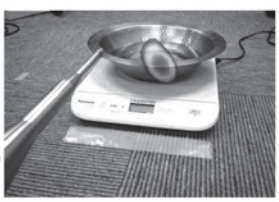


Fig. 2 Measured result by acoustic camera for a IH cooking heater.

2-10-12

2-10-12　衝撃性異音等検知に基づく住宅内の複合センサによる高年層の行動推定

Activity recognition of elderly based on impact sound detection using data collected by composite sensors mounted at home

◎岡本怜奈，△根岸美紗（IC），有光哲彦（フィート）

◆高年層の健康の見守りのため，心理的負担の少ない支援システムが必要とされるが，住宅内での生活行動の推定が課題であった。

◆本研究では，赤外線，音，加速度を含む複数種類のセンサデータの補完により，健康に関する食事，清潔，運動などの行動を推定する。

◆住宅内の台所，洗面所，和室，トイレ，および玄関扉に設置した装置から収集した赤外線センサと衝撃振動の発生数を時刻歴で統合し，サポートベクタマシンを用いた行動の二値分類モデルを構築する。

◆また，設置場所にて記録された衝撃騒音を朝・昼・夜の時間帯で分割し(Fig. 1)，MFCC などの音響特徴量を抽出する。

◆さらに，衝撃騒音からの特徴量を元に k-means 法によるクラスタリングを行い，各クラスタの騒音区分や特徴的な周波数帯域などの傾向を観察し，設置場所ごとの騒音と行動との結びつきを示唆した(Fig. 2)。

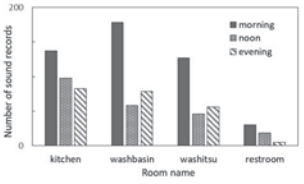


Fig. 1 Number of sounds recorded at each room

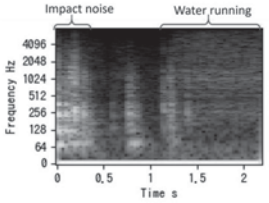


Fig. 2 Spectrogram of impact noise and water running recorded at a washbasin

2-10-13

2-10-13 時間領域差分法を用いた 3 次元大規模数値シミュレーションによる大気中で発生する微気圧変動（インフラサウンド，超低周波音）の可視化とその応用

Visualization and application of pressure fluctuations generated in atmosphere by 3-D large-scale numerical simulation using time-domain difference method

○大久保寛，齊△藤 義騎，徳田雅巳（都立大）

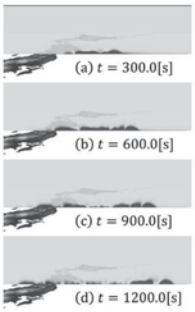
◆時間領域における差分法は音響分野において，線形音場の数値解析法として古くより利用されており，初学者から専門家まで幅広いシェアを有している．

◆大気中で発生する微気圧変動は火山噴火や巨大津波，隕石の大気圏突入などの自然現象や巨大爆発などの人工的な現象に伴って発生することが知られており，防災や監視などの用途での利用が期待され，一部実用化されている．

◆これらの大気中の微気圧変動は，頻繁に発生しないため，コンピュータによる数値シミュレーションは非常に重要な技術となる．

◆下図に，日本列島を含む計算領域の可視化結果を示す．図より海面上を伝搬する気圧変動の様子がわかる．

◆時間領域差分法では各平面の値を計算することで伝搬の時間発展を比較的容易に可視化することができ，伝搬の様相から物理現象を考察することが可能となる．



側面方向からの図示による伝搬様相

2-11-1

要素選択を用いた次元削減による楽器音分類の低演算量化

Reduction of Computational Costs for Musical Instruments Classification Using Element Selection

☆加藤瑠（都立大），植野夏樹（熊本大・都立大），小野順貴（都立大），
松田遼，近藤多伸（ヤマハ）

◆本研究では単一楽器演奏の楽器分類問題の低演算量化に取り組んだ．入力音の時間周波数表現に要素選択による次元削減を行った．

◆入力音の時間周波数表現に対し，乗算を必要としない軽量な低次元化手法として，要素選択による次元削減を行った．

◆ 選択要素 (周波数ビン) は元のデータとの平均再構成誤差を最小化するように最適化した．

◆CNN 分類に適用した場合、提案法は他の低次元化方法 (広いストライドの畳み込み、膨張率の適用、ランダム要素選択)よりも高い精度を示した．

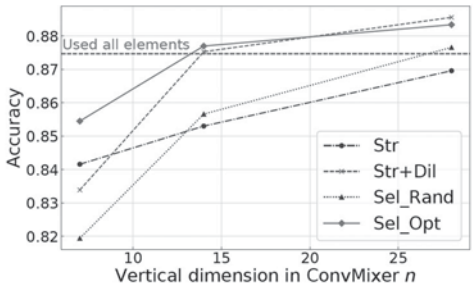


Fig.1 : Relationship between computational complexity of each model and classification performance.

2-11-2

2-11-2 マイクロホンアレイと環境音による室内状態変化検知の試み

An attempt to detect indoor abnormal states using environmental sounds and a microphone array.

○池野修一，△池田賢二（セコム）

◆本稿では，「環境音を用いた聴覚による障害物知覚」から着想を得て，空調音等の環境音の室内伝搬特性変化に着目した人物侵入等の室内状態変化検知手法に関する実験的考察を行った。

◆監視空間に設置したマイクロホンアレイの観測音より算出した空間スペクトルを画像化し，空間スペクトル上の微小な変化を捉えるため，製品の外観検査等に用いられる異常検知手法として高い性能が報告されている PatchCore の適用を試みた。

◆支配的な環境音が空調音(V)，サーバーマシン動作音(S)の場合について，正常状態(無人)/異常状態(人物存在/開口部開放)の識別性能を AUC 値で検証した。

◆本研究で用いた半径 100mm 円形アレイの場合，2kHz 以上の帯域での音圧レベル低下が少ない(S)を用いた時，正常/異常の二値分類において AUC 値 0.8 以上となることを確認した。また，補助音源を用いた検証実験より，2kHz 以上の帯域の有効性を確認した。

Table 1: AUC Score

	Freq. Band (kHz)			
	1 - 2	2 - 4	3 - 6	4 - 8
(V)	0.54	0.54	0.41	0.58
(S)	0.65	0.93	0.82	0.84

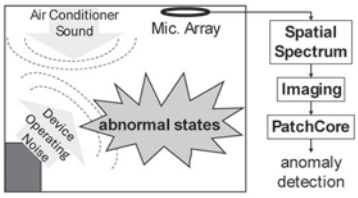


Fig.1: Overview of the proposed method

2-11-3

2-11-3 音響イベント物体の視聴覚教師あり検出

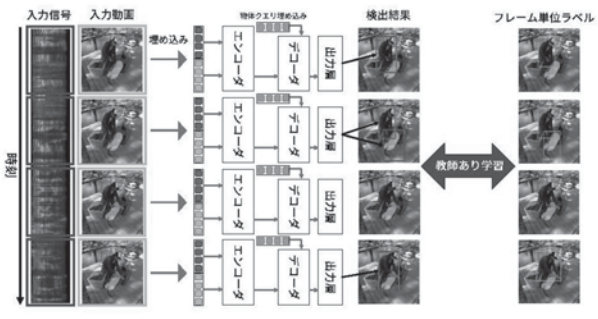
Audio-Visual Supervised Detection of Sound Event Objects

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

◆**目的：視界上の音源物体 (音響イベント物体)の位置・発生時刻の推定**
→ 周囲の物体の位置や状態を把握する自律型ロボットで重要

◆**従来手法：① 自己教師あり学習 ② クリップ単位の教師あり学習**
- ➢ ①：音と画像の共起関係を活用した対称学習
  → 学習した特徴マップから音源の位置を取得可能
- ➢ ②：クリップ単位に付与された音源物体の矩形を活用
  → ①より正確な検出が可能だが，発生時刻までは学習しない

◆**提案手法：時刻ごとに付与された位置の情報に基づき教師あり学習**
- ➢ 視聴覚データセットに音響イベント物体の位置・発生時刻を付与
- ➢ 既存の物体検出器を音響イベント物体検出向けに拡張



◆**実験結果：音響情報を物体検出に統合する際の課題を明らかにした**

2-11-4

2-11-4　音源信号の数式ドリブン合成に基づく音響イベント検出の事前学習

Formula-Driven Supervised Pre-training for Sound Event Detection

☆柴田優斗（産総研，慶應大），△田中啓太郎（早稲田大，産総研），
坂東宜昭（産総研），井本桂右（産総研，同志社大），
△片岡裕雄（産総研），△青木義満（慶應大）

◆**目的：数式ドリブンに音響信号データセット（Formula-SED）を作成し，実データを用いずに音響イベント検出モデル事前学習の実現**

◆**従来法：弱ラベル教師あり学習・自己教師あり学習**

- クリップ単位イベントラベルやマスク自己符号化器を活用
- 時間分解能が不足・個人情報保護やデータの権利問題

◆**提案法：パラメトリック生成モデルとガウス過程によるデータ生成**

- 調和振動子と加法合成器を用いてパラメトリックに信号生成
- ガウス過程を用いた合成パラメータサンプリングとラベル生成

◆**実験：DCASE 2023，Task4 のデータセットを用いて性能評価**

- CRNN と AST-GRU モデルに対して学習の高速化
- CRNN で PSDS1, 2，Event-F1, Intersection F1 において精度向上

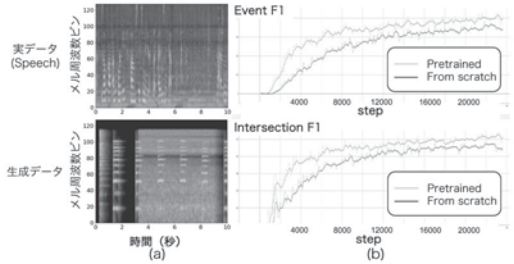


図1．生成音響データ（a）と事前学習による下流タスク精度向上（b）

2-11-5

2-11-5　音響イベント検出のための隠れセミマルコフモデルに基づくイベント単位損失

An event-wise loss function
based on a hidden semi-Markov model for sound event detection

☆吉永朋矢（早稲田大院/産総研），坂東宜昭（産総研），
△田中啓太郎（早稲田大院），井本桂右（同志社大/産総研），
△大西正輝（産総研），△森島繁生（早稲田大）

◆**目的：短時間の挿入・削除誤りの少ない音響イベント検出の実現**

◆**従来法：深層ニューラルネットワーク (DNN) + 中央値フィルタ**

- DNN の短時間フレーム単位の推定結果の誤りを後処理で訂正
- 既存の後処理手法は微分不可能なため，DNN との同時学習が困難

◆**提案法：DNN のフレーム単位の特徴量から音響イベントへの変換層**

- イベント系列を隠れセミマルコフモデル (HSMM) により表現
- **微分可能**な操作のみの前向き後ろ向きアルゴリズムを活用 (図 1)
- イベント単位の損失を設計し，DNN と HSMM を同時学習

◆**実験：URBAN-SED の合成音と DESED の実録音で性能評価**

- 両データセットで中央値フィルタを用いた従来法より高い性能

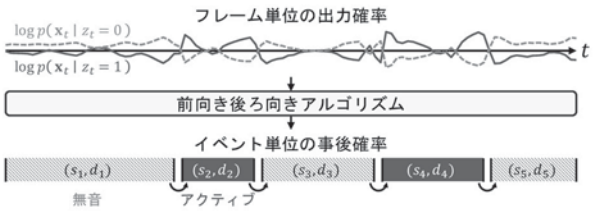


図1：HSMMに基づくフレーム単位の特徴量から音響イベントへの変換層

2-P-1

2-P-1　熱音響発電システムの実用化に向けた基礎検討
―ムービングコイル型発電機の試作―

Basic study for practical application of thermoacoustic power generation system- Prototype of a moving coil type generator -

☆中澤俊貴，坂本眞一，△平山智士（滋賀県立大）

◆熱音響発電システムの実用化に向けてムービングコイル型発電機を試作した．（Fig. 1）

◆直管型熱音響システムの管内にムービングコイル型発電機を設置し検討を行った．（Fig. 4）

◆試作したムービングコイル型発電機により発電を確認できた．得られた出力電力が，入力電力に対して単調増加することを確認した．

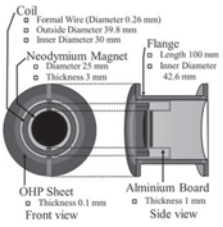


Fig. 1 Moving coil type generator.

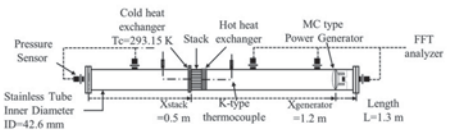


Fig. 4 Thermoacoustic moving coil type generator.

2-P-2

2-P-2　熱音響装置における導波管形状が自励振動に与える影響

Effect of Waveguide Shape on Self-Excited Vibrations in a Thermoacoustic Apparatus

○曽和信宏（京工繊大院），田中洋介（京工繊大）

◆熱音響発電は廃熱利用や経済的である点より将来性のある発電方法の一つである。しかし、技術課題として熱音響装置の設置自由度が考えられる。

◆この設置自由度の向上の方法として空間充填曲線に注目した。空間充填曲線は空間を最大限活用することのできる性質があり，工場や発電所の限られた空間で使用できる。

◆本研究では空間充填曲線を用いた小型化と異なる形状の空間充填曲線で導波管を構成しその形状による熱音響自励振動の影響を調査した。Fig.1 に製作した導波管の一つと結果を示す。

◆その結果，装置を小型化でき，導波管の形状によって熱音響自励振動の開始時間などに影響があることがわかった。

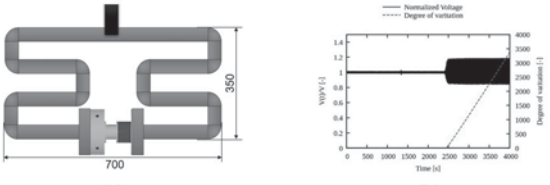


Fig. 2 (a) Thermoacoustic device with a Hilbert structure (b) Self-excited vibration

2-P-3

2-P-3　断熱シートを用いた二槽型カロリメトリ水槽の評価

Investigation of two-layer calorimetric water vessel with heat-insulating sheet

◎上原長佑（島根大学医学部付属病院），内田武吉（産総研）

◆超音波パワーは生体の熱的安全性に関連した重要な指標である。
◆そのため，産総研では測定の妥当性評価に必要不可欠な計量標準の整備を進めている。
◆高出力領域の超音波パワー標準にはカロリメトリ法による整備を進めている。
◆カロリメトリ法は超音波振動子の自己発熱が影響し，測定の不確かさが天秤法より大きいことが課題であった。
◆超音波振動子の自己発熱の影響低減のために二槽型水槽を試作した。
◆結果として，二槽型水槽の測定値が一次標準である天秤法の不確かさ範囲内で一致することを確認した（Fig. 1）。
◆二槽型水槽を用いることにより，測定の不確かさが従来法より小さく，且つ安定した超音波パワー標準が実現できる。

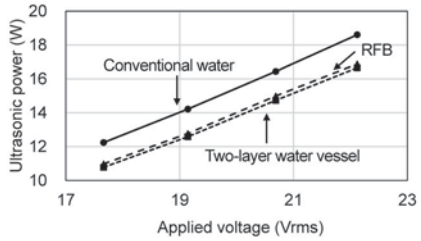


Fig.1 Comparison of ultrasonic power with two-layer water vessel, conventional water vessel and RFB method.

2-P-21

2-P-21　圧電膜／シリコンゴム構造を用いた球殻上レーザ誘起 Lamb 波の測定と評価

Measurement and evaluation of laser-induced Lamb waves on a spherical shell using a piezoelectric film/silicon rubber structure.

☆西田蒼，會澤康治（金沢工大院）

◆我々は高強度な弾性表面波を角質層に誘起させれば、皮下に与える影響が小さい低侵襲な TDD が可能になると考えた。
◆角質層への高強度な弾性表面波誘起の振動源として水浸した黒色樹脂球殻にナノ秒パルスレーザを照射して発生させた Lamb 波の利用を考えている。
◆水浸した球殻を用いる理由は、レーザプラズマ閉じ込め効果によるレーザ誘起 Lamb 波の増強や、Lamb 波の重畳による高強度化が期待できるからである。
◆本研究では黒色樹脂球殻に発生させたレーザ誘起 Lamb 波を圧電膜／シリコンゴム構造のセンサを用いて測定と評価を行った。
◆パルスレーザ光（波長 532 nm、パルス幅 10 ns）は集光レンズで黒色樹脂球殻表面に単発照射し、レーザ照射点の反対側に設置した圧電膜／シリコンゴム構造のセンサで球殻に生じる弾性波を検出した。
◆測定した時間波形のスカログラムより、レーザ誘起 Lamb 波だと思われる成分を確認した。
◆レーザ誘起 Lamb 波が表れた周波数領域をフィルタリングし、そこから逆ウェーブレット変換して得られた時間波形から信号の強度を評価した。
◆水浸した黒色樹脂球殻に発生させたレーザ誘起 Lamb 波は A0 モードであると判断した。

2-P-23

2-P-23　コウモリをモデルとする FM 超音波による浴室事故防止システムの試作と評価

Development and evaluation of a prototype of FM ultrasound sensor system modeled after bats for the prevention of fatal accidents in bathrooms

○藪謙一郎，伊福部達（東大）

◆浴室での死者数は年間で 1 万件以上発生していると推計され、プライバシーが高度に維持される見守りセンサが求められている。
◆我々は、周囲の物体の動きや状況を非接触で比較的広範囲に把握できる手段として、FM コウモリをモデルとした超音波センサの開発を進めている。
◆本稿では、これらの検出力を向上させるため、高出力タイプの送受信機を開発するとともに、浴室におけるヒトの検出を目的としたアルゴリズムの改良を行った。その開発と評価について報告する。

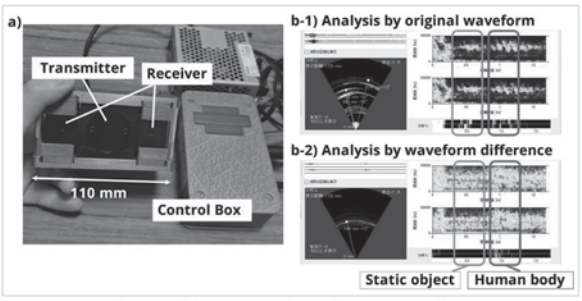


Fig.1：(a) Developed ultrasonic sensor system,
(b)Detection of static objects and moving human body

2-P-24

2-P-24　低周波超音波導入法のための小形キャビティ共振型トランスデューサ

A compact cavity resonant ultrasound transducer for low-frequency sonophoresis

☆山本真也（東工大・工）、杉田直広、△進士忠彦（東工大・未来研）

超音波照射によって皮膚の薬物浸透性を向上させる超音波導入法では、皮膚表面への低侵襲な超音波キャビテーションの生成が重要となる。本研究では、皮膚表面上に効率よくキャビテーションを生成し、100 kHz 以下での低周波駆動可能な小形超音波トランスデューサを提案し、経皮薬物導入を可能とするウェアラブルデバイスの開発を目指す。トランスデューサの発熱低減と高出力化のために、図 1 に示す小形キャビティ共振型トランスデューサ構造を提案した。トランスデューサの直径は 28 mm、厚さ 10 mm であり、共振周波数は 78 kHz である。図 2 に示すように、共振モードにおいて、半径方向に振動する開口端の直径と音波波長をマッチングさせ、半径方向の定在波を形成する。キャビティ共振によりアドミッタンスが低下し、トランスデューサに流れる電流は低下すると予想される。水中におけるトランスデューサの駆動において、電流が低下することを確認した。また、ウレタンゲルファントム表面でのキャビテーション生成を確認した。

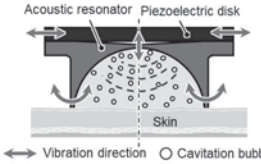


Fig. 1 Schematic illustration of the proposed transducer.

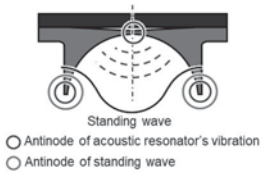


Fig. 2 Schematic illustration of cavity resonance.

2-P-25

2-P-25　マルチモード型超音波トランスデューサによるキャビテーション生成

Cavitation generation by multimode ultrasonic transducer

☆山﨑悠介，山本真也，杉田直広，△進士忠彦（東工大）

◆超音波導入法は，超音波によって発生するキャビテーション気泡を皮膚表面に作用させ，皮膚の薬物浸透性を向上させる技術である。従来の超音波トランスデューサは，超音波キャビテーションが音波照射面に対して不均一に発生し，薬物の輸送領域が局所的かつ不均一に形成される課題がある。

◆本研究では，音波照射面に超音波キャビテーションを広範囲かつ均一に発生させることを目的として，圧電素子の電極を多極化したマルチモード型の超音波トランスデューサを開発している。図1にトランスデューサの概略を示す。

◆トランスデューサの駆動方法としてマルチモード駆動を提案する。マルチモード駆動は，全電極に同位相の信号を入力する同相モードと，圧電素子の電極を４分割し，隣り合う電極に位相が $\pi/2$ ずつ異なる信号を入力する進行波モードを同時に励振する。図２に同相，進行波駆動での圧電素子の変形の様子を示す。

◆試作機を用いて同相と進行波，マルチモード駆動での超音波キャビテーションの発生の様子を観察した。同相駆動に比べ，マルチモード駆動ではキャビテーション発生範囲が拡大した。

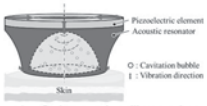


Fig. 1 Schematic illustration of the proposed ultrasonic transducer.

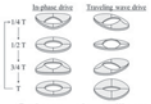


Fig. 2 Schematic illustration of piezoelectric element deformation.

2-P-26

2-P-26　２層分極反転 Al 極性 SiAlN/N 極性 SiAlN 薄膜の形成と２次モード BAW 共振子への応用

Fabrication of two-layer polarization inverted Al-polarity SiAlN/N-polarity SiAlN films and application to 2nd mode BAW resonators

☆福永慶，鈴木雅視，垣尾省司（山梨大学）

◆本研究では成膜プロセスを簡略化した分極反転構造薄膜 BAW 共振子を作製するため，SiAlN 薄膜の極性をガス比によって制御し，SiAl 合金ターゲットのみを用いた基板付き２層分極反転構造 SiAlN 薄膜 BAW 共振子を連続成膜で作製し，評価した．

◆２層分極反転 SiAlN 薄膜では２層目 SiAlN 薄膜が１層目の N 極性を一部引き継ぐことが判明した．膜厚比が理想的な 1：1 となる２層分極反転 SiAlN 薄膜を形成するために引継ぎ膜厚を考慮して成膜時間を変更し成膜を行い，変換損失の測定と Mason の等価回路による理論曲線を比較した．その結果，膜厚比が理想的な 1：1 に近づいていることが判明した．加えて，HBAR 共振子において狙いとした２次モード共振を達成していた．

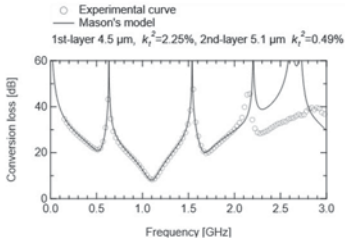


Fig.1：Conversion loss of two-layer polarization inverted SiAlN film BAW resonator (1st-layer deposition time≈2 h, 2nd-layer deposition time≈5 h)

2-P-27

2-P-27 Musical Saw の音階制御機構に基づく PLLA 曲げセンサ

PLLA Bending sensor based on the tone control mechanism of the musical saw

△戸部太陽，☆澁谷和典，△宇都大河，山本健，△宝田隼（関西大），
△森健一，△安藤正道（村田製作所），△田實佳郎（関西大）

◆ Musical Saw は弓でノコギリを擦って発音し，刃の曲率を変えることによってピッチを制御する体鳴楽器である。

◆ 発音には，flapping mode と呼ばれる翼が羽ばたくような振動モードが有効であることが分かっている。

◆ 本研究では，有限要素法を用いたシミュレーションにより flapping mode における音階制御特性を明らかにすることを目的とした。さらに，発音機構と PLLA 圧電フィルムの圧電特性を組み合わせることで，Musical Saw の音階制御機構を用いた角度検知センサの開発を検討した。

◆ 本モデルの曲げを徐々に大きくすると，flapping mode が現れる共振周波数は徐々に高くなるという結果が得られた。

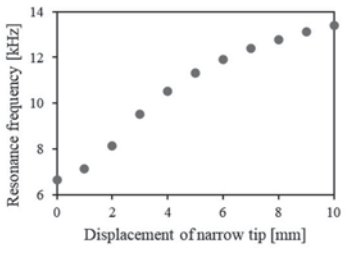


Fig. 1 Dependence of resonance frequency on bending displacement.

2-Q-1

聴覚機能の改善を目指した訓練アプリの開発および使用感調査 ―認知機能に関する検討―

Development and usability evaluation of a mobile application for auditory training −Training of cognitive functions−

○藤平晴奈（九州大学），△岡本潔和（サイバーエージェント），岩本沙也（九州大学），
△西元克茂，△竹松知紀，△飛松葉子，△石橋るみ，△田島康宇（九州リオン），
△越智亮太，森周司（九州大学）

◆これまで音に関する課題を繰り返し行う聴覚訓練によって、音声に対する聞き取りが向上することが報告されている。

◆そのため聴覚訓練は、加齢による聴取能力低下の予防や、補聴器や人工内耳装用の初期段階における聴取練習での活用が期待されている。

◆聴覚訓練の内容としては、ワーキングメモリや、注意力、短期記憶といった聴覚を使った認知課題がいくつか報告されている。

◆我々はこうした認知課題を含み、一般の方でも聴覚訓練を簡単に行うことができる聴覚訓練アプリを開発した。そして開発したアプリに対する使用感調査を若年者と高齢者を対象に行った。

◆使用感調査ではシステムや製品の使いやすさを評価する心理尺度として広く使われている System Usability Scale を用いた。その結果、すべての実験参加者がアプリに対して普通以上の使いやすさと評価していることが示された。

◆一方でアンケート調査からは、訓練の実施時間やボタンの配置、課題の音声の読み上げや画面遷移までのタイミングに関して、いくつか問題点が挙げられた。

◆今後は挙げられた問題点を改善した上で、アプリによる訓練効果を検証する予定である。

2-Q-2

2-Q-2　聴覚機能の改善を目指した訓練アプリの開発および使用感調査 ―聴覚処理能力に関する検討―

Development and usability evaluation of a mobile application for auditory training - Training of sensory functions -

☆岩本沙也(九州大)，△西元克茂，△竹松知紀，△飛松葉子，△石橋るみ，△田島康宇(九州リオン)，△越智亮太，藤平晴奈，森周司(九州大)

◆聴覚処理能力は音を用いた課題を繰り返し行う聴覚訓練にて向上することが、いくつかの研究で報告されている。

◆認知能力の改善の面が強い、覚訓練プログラムは存在するが、聴覚処理能力に特化した訓練プログラムは存在しない。

◆そこで、音声知覚に重要な振幅包絡と時間微細構造の処理能力に関する、無音検出、振幅変調検出、周波数変調検出、周波数弁別の４つを訓練課題として選出し、自宅で簡単に訓練を行えるアプリを開発した。そして、若年者と高齢者を対象としてアプリの使用感調査を行った。

◆使用感調査では、システムや製品の使いやすさを評価する心理尺度として広く使われている System Usability Scale や独自に作成したアンケートを用いた。結果、実験参加者全員が普通以上に使いやすいと評価した。

◆アンケート調査では、訓練課題や、訓練画面のわかりづらい点等の指摘や、評価をわかりやすくしてほしいといった要望が挙げられた。

◆今後は挙げられた改善点を反映し、アプリを用いた訓練による音声知覚の向上を検証する予定である。

2-Q-3

2-Q-3　若年者と中高年者の音韻修復機能の比較 ～単語親密度の影響～

Comparison of phonemic restoration between young and middle-aged adults ~Influence of word familiarity~

☆湯淺　麻生，大塚 翔，中川 誠司(千葉大)

◆加齢に伴い多くの知覚・認知機能の低下が見られる一方で，音韻修復機能は高齢者と若年者で差がないことが報告されている．それゆえ，音韻修復は，加齢に伴う他の知覚・認知機能の低下を補っている可能性がある．

◆一方，若年者を対象とした研究では，音韻修復機能を評価する課題に別の課題を課すことで，音韻修復機能が強く働くことが示されている．前報では，親密度が高い単語を使用し，若年者と中高年者における音韻修復機能について，二重課題に対する感受性も含めて比較した．本研究では，親密度が異なる単語に対する音韻修復の特性を比較することで，語彙の想起，および，音響情報に基づく音韻修復に対する加齢の影響を評価した．

◆親密度 1, 4 の 4 モーラの単語の 2 文字目の子音音素にノイズを置換，または，重畳した．被験者はいずれかの音声をランダムに聴取し，置換・重畳いずれかを解答した．音韻修復が生じるとノイズ置換音声を重畳と解答すると推察されることから，ノイズ置換音声に対する誤答率が高いほど音韻修復が生じていると解釈される．

◆親密度１では両群に差は見られず，親密度４では音声をノイズで置換したときに中高年者群の誤答率は若年者群より高くなった (Fig. 1)．加齢により，語彙の想起に基づく音韻修復が起きやすくなることを示唆している．

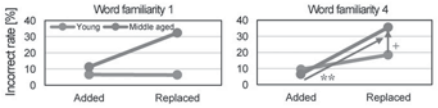


Fig. 1 Age cohort x added/ replaced interaction. $^{+}p<.10$, $^{**}p<.01$(simple main effect)

2-Q-4

2-Q-4　健聴者における加齢および純音聴力レベルに基づく雑音下文聴取閾値の推定

Estimation of thresholds for hearing sentences in noise based on ageing and pure tone threshold levels in normal-hearing

◎岡龍也，中市健志(リオン)，△岡本康秀(済生会中央/慶應大)

◆　本研究では健聴者における加齢および純音聴力レベルと雑音下の音声聴取能の関係を相関分析，重回帰分析にて評価した．

◆　雑音下音声聴取能と 0.125～8.0 kHz の範囲の純音聴力レベルとの関係は 8.0 kHz の聴力レベルに最も強く表れ，また平均聴力レベル 4 分法 B（(0.5+1.0+2.0+4.0 kHz) /4）の計算に 8.0 kHz の純音聴力レベルを含めることで雑音下文聴取閾値との相関係数は高くなる．（$r = 0.45 \Rightarrow r = 0.54$）

◆　重回帰分析の結果，年齢，純音聴力レベルの 0.125，2.0，8.0 kHz より雑音下音声聴取能を概ね推定できることが示唆された．

Regression coefficients	VIF	Noise Front model partial regression coef	p-value
0.125 kHz	3.894	0.0249	
0.25 kHz	5.354	-	
0.5 kHz	4.139	-	
1.0 kHz	1.823	-	
2.0 kHz	1.892	0.0289	
4.0 kHz	1.475	-	
8.0 kHz	1.560	0.0334	***
Age	1.720	0.0462	***
Intercept	-	-7.7166	***
Adjusted R-squared		0.405	
√R2		0.646	
p-value		***	
Residuals (Median)		0.029	

*** p<.001

Table.1 : Multiple regression analysis

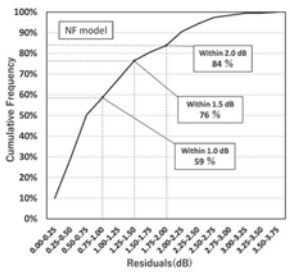


Fig. 1 : Cumulative Frequency curve

2-Q-5

2-Q-5　オンライン聴覚実験における再生環境の違いが両耳ピッチ知覚に及ぼす影響

Effects of sound devices on binaural pitch perception in online auditory experiments

☆赤木志帆(筑波大)，山岸慎平(NTT CS 研)，古川茂人(静岡社会健康医学大院)，寺澤洋子(筑波大)

◆研究目的：

・両耳ピッチ課題を用い音響機器や個人による聴覚特性の違いを測る

◆研究内容：

(1) Wired / Wireless / Speaker 条件で両耳ピッチ課題を実施

(2) (1)の Wired / Wireless 条件を同じ音響機器を用いて実施

(3) 両耳ピッチ課題の結果と Wireless 機器情報の比較

◆実験結果：

(1) 平均正答数は Wired > Wireless > Speaker の順序で高かった

(2) 平均正答数は Wired > Wireless の順序で高かった

(3) 両耳ピッチ課題の結果と Wireless 機器情報に相関関係はなし

◆まとめ：

Wireless 機器を実験から除外する必要がない可能性を示唆する

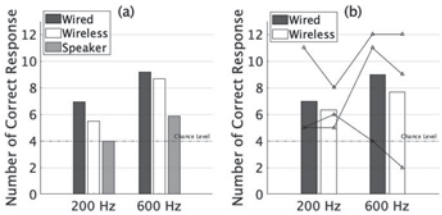


Fig.1 : Mean number of correct responses for three sound-device conditions (a) and for three participants who used the same sound device(b)

2-Q-6

2-Q-6　日本語における音象徴音韻マッピング

Sound Symbolism Phonological Mapping in Japanese

☆今井 裕大[1], △渡邊 智美[1], △村井 翔太[1,2], 小林 耕太[1]
1 同志社大学院 生命医科学研究科, 2 東京大学 国際高等研究所ニューロインテリジェンス国際研究機構 (WPI-IRCN)

◆音と意味の関係が非恣意的で強い関連性をもつ現象を音象徴と呼ぶ。これまでの音象徴研究から音声特性が形や大きさなどの意味と結びつくことを示されてきたが、刺激の恣意性（表示法、刺激選択など）が問題であった。

◆ 本研究では、刺激・提示方法を統制し音韻を網羅的に調査した。

◆ 結果、先行研究と一貫した結果が得られ大きさの評定では母音が、鋭さと硬さの評定では子音が大きな影響を及ぼすことが示唆できた。

◆また世代間比較（若年者層 vs 高齢者層）から年齢に有意な主効果はなく音象徴の効果は年齢依存的なものではなく年齢普遍的な現象であることが示唆できた。

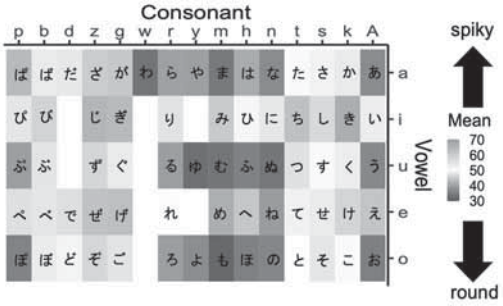


Fig.1 : Sound Symbolism Japanese syllable Table in Shape condition.

2-Q-7

2-Q-7　ノイズキャンセリングヘッドホンを利用した音質評価法の検討

Verification of sound quality evaluation method using noise-canceling headphone.

☆森脇想, 小口純矢, △辻野雄大, 森勢将雅(明治大)

◆音質評価の課題:暗騒音の少ない環境で行われなければならずコストや聴取者負担がかかる→**ノイズキャンセリングヘッドホンを利用した評価を提案**

◆Opus による情報圧縮をした音声を,防音室で通常のヘッドホンを装着した場合と,通常の室内でノイズキャンセリングヘッドホンを装着した場合でそれぞれ MUSHRA 法で評価 → 結果を比較

◆Two One-Sided Test (TOST) 法に基づく等価性の検定により**有意に同等**

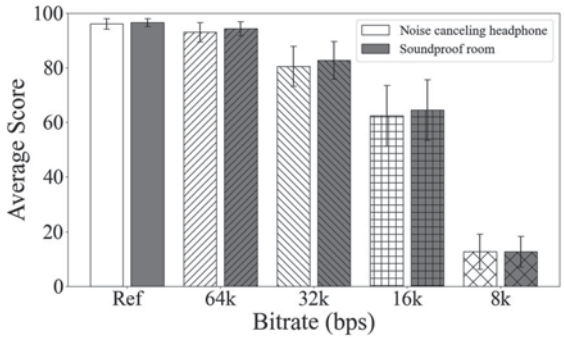


Fig.1 : Each histogram represents the mean value of the scores. Error bars are 95% confidence interval.

2-Q-8

2-Q-8　雪と音源の高さが音声の聞き取りやすさに与える影響

Effect of snow cover and sound source height on speech intelligibility

☆佐々木渉吾, 近藤和弘(山形大院・理工学研)

◆積雪・降雪・密度・温度・湿度・風速を定量的に定義して計測。

◆積雪は地表に積もっている雪の高さ、
降雪は１時間当たり雪が降った量を計測、
雪質は、雪密度を測定することで定量的に定義。

◆風速は、信号を録音している間の平均風速を用いて測定

◆音源の高さは地面または積雪面からの高さとして定義

◆積雪・降雪時に屋外でインパルス応答測定を実施、積雪・降雪時の伝達特性を取得。

◆取得した伝達特性を評価音声に畳み込み、積雪・降雪時を模擬した音声で二者択一型日本語版 DRT を用いて音声了解度評価。

◆無雪時は音源の高さによる音声了解度への一貫した影響なし。
→降雪時は高さと了解度に正相関。

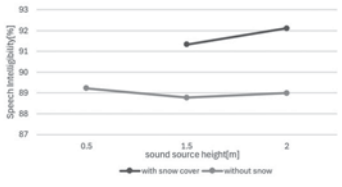


Fig.1:　音源の高さによる音声了解度の比較

2-Q-9

2-Q-9　高齢者を対象とした音声了解度実験と客観評価指標 GESI を用いた予測

Speech intelligibility experiments with elderly subjects and prediction using the objective evaluation index GESI

☆宮崎芙紀(和歌山大院), △馬野颯太(和歌山大), 山本絢子(和歌山大院), 森本隆司(リオン), 入野俊夫(和歌山大)

◆高齢難聴者の音声了解度の予測を目指し開発されている客観評価指標 GESI の高度化のため、高齢者を対象に音声了解度実験を行い、精度良く予測できるかを検討した。

◆結果
GESI に位相補正や
TMTF 特性の導入
↓
ある程度
予測可能性は増す。

高齢者 Sbj9 の予測で得られた係数で聴力レベルが同等な高齢者 Sbj10 を予測
↓
過小評価する。
般化能力に関しては改善の余地がある。

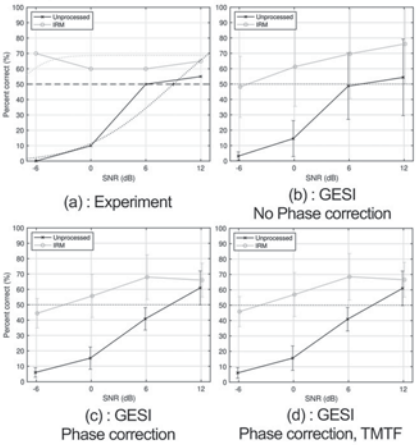


Fig.1 : Results of speech intelligibility experiment and GESI predictions.
Subject : Sbj9.　black : Unpro, grey : IRM.

◆さらに現在、高齢者 Sbj9 の特性を反映させた模擬難聴音声を用いて、健聴者対象に了解度実験を行なっており、途中経過も紹介する。

2-Q-10

2-Q-10　音声基盤モデルと聴覚モデルを用いた難聴者のための音声了解度予測の検討

A study on speech intelligibility prediction using speech foundation models and auditory models for hearing-impaired listeners

○山本克彦，郡山智樹（サイバーエージェント）

◆先行研究：音声基盤モデルベースの音声了解度予測モデル (SFMs)
- ➢ 音声了解度の予測コンペ (CPC2) で一位の予測精度
- ➢ 聴取者の特徴量はオージオグラムのみ
  → 問題点：難聴者の聴覚特性を考慮した特徴量抽出が不十分

◆提案手法：SFMs に聴覚モデル特徴量を導入
- ➢ 利点：音声信号とオージオグラムによって入出力特性が変化
  - ✧ 動的圧縮型ガンマチャープフィルタバンク (v234; 模擬難聴)
  - ✧ 変調フィルタバンク

◆実験：CPC2 のデータセットを用いた評価 (15 試行)
- ➢ 一部のテストデータで予測誤差 (RMSE) が有意に改善 ($p < .05$)
- ➢ 聴覚モデル特徴量を加えることで予測精度が向上することを示唆

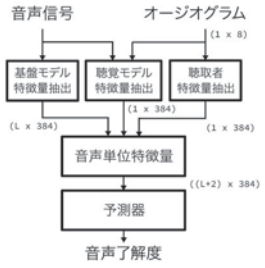


Figure 1: Overall structure of our proposed model.

2-Q-11

2-Q-11　自己録音音声に対する不快感情は普遍的に生じるのか？:オンライン実験による検討

Do unpleasant feelings towards one's own recorded voice arise universally? : An online experimental investigation

☆冨岡晟多，△長村秀一，小林耕太（同志社大院・生命医科学）

◆多くの人が自己録音音声に否定的な反応を示すことが知られており、この不快感が自己評価を低下させる可能性がある。

◆一方で、録音された自分の声を魅力的に感じるといった矛盾する事例も報告されており、これらの現象をうまく説明できるメカニズムは未だ不明である。

◆本研究では、対面実験とオンライン実験を通じて、評価軸におけるポジティブ感情とネガティブ感情を独立した評価手法を用いて、自己録音音声の感情評価を定量化した。

◆その結果、対面・オンラインともに、ポジティブ感情よりもネガティブ感情のほうが強く（Fig. 1）、実際に録音音声を聞いた後よりも普段の録音音声に対する感情のほうが高まった。これより、ポジティブ・ネガティブ感情を分離して評価することで、自分の録音音声に対してネガティブな感情が優勢に生じる可能性が示唆された。

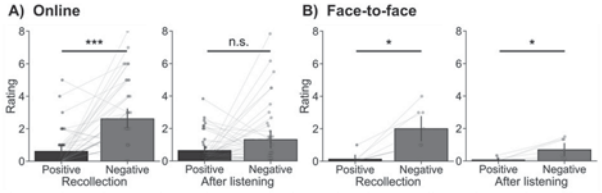


Fig.1 Negative and Positive evaluations of self-recorded voice, (A) Online (B) Face-to-face experiment. (Wilcoxon signed-rank test, ***$p < .001$, *$p < .05$).

2-Q-12

2-Q-12　骨導伝達特性の個人性を考慮した自己聴取音模擬の一検討

Investigation on mimicking speakers' own perceived voice by subjective evaluation considering bone-conduction transmission characteristics

△河野太一，☆久保田樹，小澤賢司，鳥谷輝樹（山梨大）

◆**背景**：発話者の自己聴取音知覚における骨導伝達過程の解明が期待されている．Toya ら（2022）は，自己聴取音の音色の主観評価に基づき，中耳・内耳への振動伝達（経路 B）と外耳道内放射（経路 C）の寄与を推定した．
→ 参加者間で平均した骨導伝達特性に注目しており，骨導伝達特性の個人性は無視された．

◆**目的**：骨導伝達特性の個人性を考慮して自己聴取音声を模擬することで，骨導経路 B・C の知覚的寄与を推定すること．

◆**方法**：同期収録した気導音と側頭部振動，外耳道内放射音の間の相対スペクトルに基づき，経路 B・C の骨導模擬フィルタを設計．
→ 経路 B・C の模擬骨導音間の混合比 $L_{B/C}$，気導・骨導音間の混合比 $L_{(B+C)/A}$ を主観評価により決定．

◆**結果**：$L_{B/C}$ は −9 ～ −1 dB（中央値 −4 dB）に分布し（Fig. 1），$L_{(B+C)/A}$ は −3 ～ 0 dB に分布．

◆**結論**：自己聴取音知覚に対して，経路 C は経路 B よりも約 4 dB 寄与が大きい可能性が示唆された．

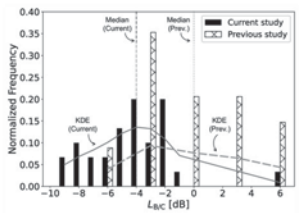


Figure 1: Histogram for the mixing ratio of B-filtered speech to C-filtered speech for mimicking speakers' own voice.

2-Q-13

2-Q-13　無声音声の明瞭度と異聴傾向の検討

Intelligibility and Phoneme Confusion of Unvoiced Japanese Speech

○今野英明，△林 桃歌（北海道教育大）

◆ささやき声は声帯振動を伴わずに発声される音声であり，その明瞭さは通常の音声よりも劣るとされる。

◆本研究では，通常音声 (ordinary)，通常音声の音源を雑音に変更して再合成した疑似ささやき声 (unvoiced)，ささやき声 (whispered) の明瞭性と異聴傾向を調べた。

◆対象音声は 6 名の話者による 100 単音節，および親密度が 1.0 - 2.5 の 200 単語。被験者は男女各 12 名。

◆ささやき声の明瞭度と了解度は通常音声に比べ約 20%低下

◆ささやき声では有声子音が無声子音に，疑似ささやき声では無声子音が有声子音に聞き取られる。

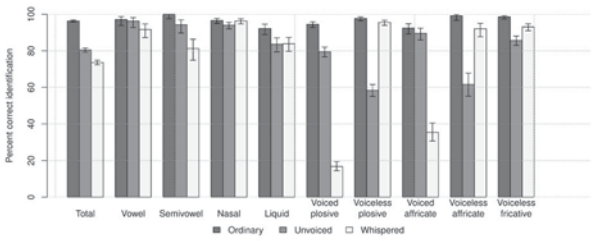


Fig. 1: Percent correct identification with 95% confidence intervals of Japanese 100 monosyllables grouped by the manner of articulation of the consonant part.

2-Q-14

2-Q-14　音声想起時の脳活動識別のためのデータ拡張手法の検討

Data augmentation for imagined speech classification.

◎矢野　肇, 高島遼一, 滝口哲也(神戸大), 中川誠司(千葉大)

◆音声想起時の脳磁図（MEG）および脳波（EEG）データセットに対し，時間・周波数・空間領域のデータ拡張手法を適用し，想起音声の分類に有効なデータ拡張手法を調査した.

◆MEG データでは加算平均と channel dropout が，EEG データでは Fourier transform surrogate が有効であった（Fig. 1）.

◆音声想起時の脳信号の位相を保つデータ拡張手法が想起音声の分類に有効である可能性が示唆された.

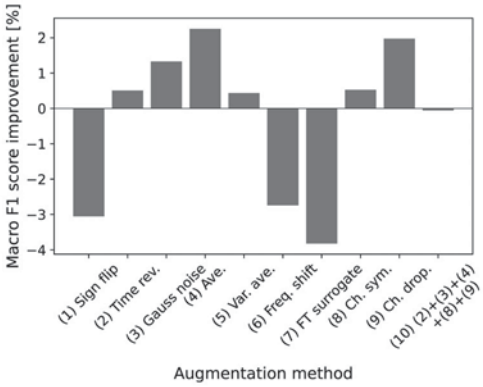


Fig. 1 Amount of the macro F1 score improvement of imagined speech classification for the MEG dataset.

2-Q-15

2-Q-15　協和性弁別能の生理・心理評価

Physiological and psychological assessments of discrimination capability of consonance

☆遠矢珠美, 大塚　翔, 中川誠司(千葉大学)

◆和音の聴取に伴い生じる協和感は構成音の周波数比に応じて変化することが知られているが，そのメカニズムには未解明な点が残る.

◆一方，音楽聴取能力の変化が脳機能に有意な変化を生じることが報告されている．和音の協和感についても弁別能と脳幹反応振幅の間の相関が報告されているが，協和感と皮質活動との関係は未解明である.

◆本研究では和音の協和性の弁別能を一対比較法およびミスマッチ・ネガティビティの計測によって評価し，両者の関係を論じた.

◆MMN 計測では，協和感の違いによる MMN 成分のみを観察するため，Virtala et al. (2010) の方法に準じた刺激呈示を行った.

◆一対比較法で得られた尺度値の差分と MMN 振幅の間には弱い正相関が認められた（$p<0.05$，$r=0.44$）（Fig. 1）.

◆本研究の結果は，一対比較法で得られた協和度の距離を MMN で評価できる可能性を示唆している.

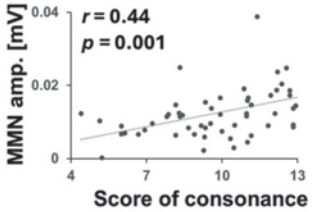


Fig.1： A correlation between MMN amplitude and score of consonance

2-Q-16

2-Q-16　楽曲の呈示タイミングが睡眠の質に及ぼす影響の検討

Effect of timing of music presentation on sleep quality

☆入交陽斗, 大塚　翔, 中川誠司(千葉大)

◆ 音楽聴取による睡眠の質の向上がいくつかの先行研究で報告されているが，その機序には依然として不明な点が多い．特に，楽曲の呈示時間やタイミングに統一された基準が存在しない．そこで，楽曲の呈示タイミングを変化させ，睡眠の質の比較を行った.

◆ 以下の 4 条件で睡眠時脳波および眼電図の計測を行った．(1) 無音 (None)，(2) 実験開始から終了まで呈示 (Keep)，(3) 実験開始から 30 分間のみ呈示 (Before)，(4) 実験開始 30 分後から終了まで呈示 (After)．また，覚醒後の被験者に対し，OSA 睡眠調査票簡易 MA 版 (OSA-MA) を用いて睡眠内省の評価を行った.

◆ Keep と比較して，After で総睡眠時間の有意な増加が確認された．また，実験開始直後に After で α 波のパワースペクトル密度の有意な減少が確認された.

◆ 今回の実験結果から，入眠前の楽曲の聴取は，その後の睡眠の持続を妨げる可能性が示唆される．しかしながら，被験者の心理状況や楽曲に対する好ましさが影響した可能性も否定できない．慎重な検討が必要である.

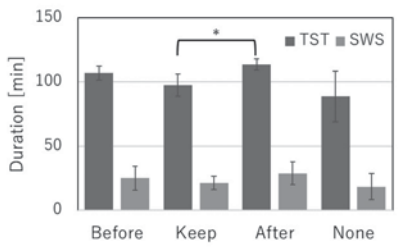


Fig.1：Total sleep duration and slow wave sleep duration for each condition.

2-Q-17

2-Q-17　オリーブ蝸牛束反射と雑音下での聴取成績の関係 –耳音響放射の遅延時間計測に基づく検討

Relationship between medial olivocochlear reflex and listening performance in noise - an examination based on latency measurements of otoacoustic emission, by DOWAKI, Yuri, OTSUKA, Sho and NAKAGAWA, Seiji (Chiba University).

☆堂脇悠里, 大塚翔, 中川誠司(千葉大)

◆ 脳幹から外有毛細胞へ投射するオリーブ蝸牛束には，音響刺激に反応し，外有毛細胞 (OHC) の増幅を抑制する働きがある（オリーブ蝸牛束反射：MOCR）．MOCR は，OHC のゲインを調節することで，雑音下での聴取を改善する役割（アンチマスキング効果）を果たす ことが電気刺激実験で示されている一方で，ヒトにおける有効性については議論が分かれていた．著者らはこれまでに，聴覚的注意に伴う MOCR の増強が，アンチマスキング機能に与える影響を検討してきた.

◆ 前報では，音響刺激による耳音響放射 (OAE) の抑制量を指標として，MOCR の効果を評価したが，本研究では，OAE の遅延時間の減少量を評価指標として導入した．OAE の抑制は MOCR による外有毛細胞のゲインの減少を反映し，遅延時間の減少は基底膜振動の周波数チューニングの鈍化を反映している．この両側面を評価することで，MOCR と雑音下での聴取の関係について総合的な理解を試みた.

◆ 聴取困難度実行中の OAE については，雑音下での信号検出閾との相関に，OAE 抑制量は有意な負の相関があり，OAE 遅延時間の減少量は有意傾向の負の相関があった (Fig. 1)．一方で，非課題実行時の OAE については，両指標とも，聴取困難度実行中に比べて相関係数が小さかった.

◆ 遅延時間の減少量よりも抑制量の方が，検出閾との相関が高いことから，MOCR による周波数チューニングの鈍化と比較して，ゲイン減少の効果の方が，雑音下での聴取への寄与が大きいことが推察される．これは，OHC のゲインを減少させることで神経系の順応を防ぎ，雑音下での聴取を改善するというアンチマスキングの仮説と合致している．また，聴取困難度実行中の方が，検出閾との相関が高かったことは，能動的な MOCR の制御がアンチマスキング効果に寄与していることを示唆している.

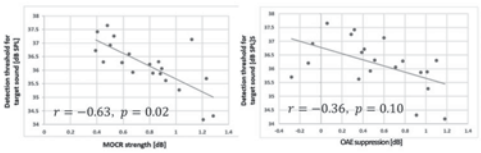


Fig.1：Correlation of listening performance with OAE suppression (left) and latency reduction (right) during detection-in-noise task

2-Q-18

大脳皮質の活性がオリーブ蝸牛束反射に及ぼす影響―聴性誘発電位を用いた検討

Effect of cortical activity enhancement on medial olivocochlear reflex.

☆内山寛大，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へ接続した遠心性神経束であるオリーブ蝸牛束は，強大音に反応し外有毛細胞の増幅を抑制することで内耳を保護している．この反応はオリーブ蝸牛束反射 (MOCR) と呼ばれている．MOCR 強度の計測よって騒音性難聴のリスクを事前評価できる可能性がある．しかし MOCR は計測毎のばらつきが大きく，安定したリスク評価のため MOCR の変動の要因の解明が必要である．

◆ MOCR は認知活動の影響を受けることが報告されている．前報では広範囲な皮質ネットワーク活動を要する暗算課題中の自発脳波と MOCR 強度を計測した．課題実行により，計算処理と関連するδ帯域の活動の増加した一方で，MOCR には減少傾向が見られた．これは，惹起された計算課題を司る皮質ネットワークの活動が，MOCR に影響を与えることを示唆するものの，両者をつなぐ神経基盤は未解明であった．聴覚皮質からオリーブ蝸牛束に遠心性神経が連続していることを踏まえると，計算課題によって惹起された神経活動が，音に関連した皮質活動を修飾し，それが遠心性神経を経由してオリーブ蝸牛束まで及んだ可能性がある．そこで，本研究では，MOCR の誘発に用いたホワイトノイズに同期した聴性誘発電位を測定し，音に関連した皮質活動の変動と MOCR の関係を調べた．

◆ 計算フェーズでは暗算課題が 6 問呈示され，計測フェーズでは 7 秒間の MOCR が行われた．計算・計測フェーズを 1 エポックとし，50 回繰り返した．有意差は見られなかったものの MOCR には課題進行に伴って減少する傾向が見られた．N1-P2 振幅は，課題開始直後は減少したものの，それ以後は課題進行とともに増大した (Fig. 1)．この結果は，暗算課題の実行よって惹起された複合的な皮質ネットワーク活動が，音に対する注意も促進することを示唆している．計算課題の進行に伴う MOCR の減少に，この音に対する皮質レベルの応答性の増大が関与している可能性がある．

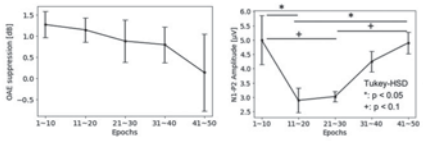


Fig. 1 MOCR (left panel) and N1-P2 amplitude (right panel) for each epoch

2-Q-19

講演取消

2-Q-20

2-Q-20　DNN 声質変換による男女間話者変換の品質に与える音響特徴量の違いに関する分析

An Analysis of Acoustic Feature Differences on the Quality of Inter-Gender Speaker Conversions Using DNN Voice Quality Conversion

☆松久直生，坂野秀樹（名城大院）

◆ 本研究は DNN 声質変換システムを用いて男女間において声質変換をする際に音響特徴量が品質にどの程度影響を及ぼすかを分析したものである．

◆ StarGANv2-VC を用いて JVS コーパスの中から男女各５名を選び，**声質変換前後で平均基本周波数の差，平均ケプストラム距離の差における平均ケプストラム距離の比較と MOS テストを行い，分析した．**

◆ 結果，ケプストラム距離と変換の品質には相関がなく，平均基本周波数の差と品質には正の相関がみられることが確認できた．

◆ **男女間で声質変換を行う際，基本周波数の差の影響を受けやすい傾向があることが考えられる．**

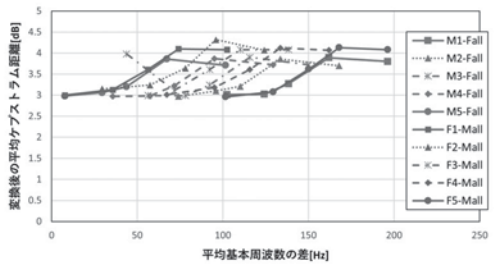


Fig. 1 Cepstrum distance at average fundamental frequency

2-Q-21

2-Q-21　学習データにおける言語の違いによる DNN 声質変換音声の音声特徴量の違いに関する調査

Investigation of difference of acoustic features caused by language difference of training data in DNN voice conversion

☆両角祐二，坂野秀樹（名城大院）

◆ DNN 声質変換モデルの StarGANv2-VC を用い，変換後の音声の音響特徴量が入力音声および学習の際の言語によりどのように異なるか分析

- 使用言語：日本語・タイ語・英語

◆ 客観評価実験：言語によらず女声間の変換でケプストラム距離が小さい

◆ 主観評価実験：言語によらず男声間の変換で MOS 値が高い

◆ 客観評価と主観評価で評価が入れ替わる結果となった

- ２人の音声を結合したような不自然な品質になっていたことに起因する可能性

Table 1 Cepstral distances for three languages.

	Japanese	Thai	English
F-F [dB]	3.89	3.71	3.98
M-M [dB]	4.27	4.30	5.54

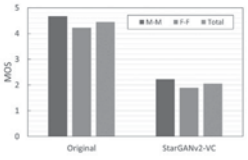


Fig. 1 MOS for Japanese speakers.

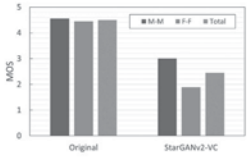


Fig. 2 MOS for Thai speakers.

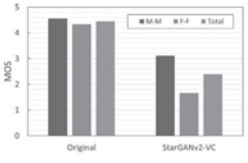


Fig. 3 MOS for English speakers.

2-Q-22

2-Q-22 音声の F_0 推定におけるオフセットの影響について

A Study of the offset on F0 estimation of speech

〇舟木慶一（琉球大）

音声の F_0 推定は、音声分析、音声合成、音声強調などの音声処理から医療診断に至るまで、様々な分野で重要な役割を果たしている。F_0 推定の重要性にも関わらず、その評価には一貫性がなく、公正ではない比較結果となることが多い。この一貫性のなさは、主に F_0 時系列の推定値と真値間の時間的整合性の欠如に起因している。文献 [1] ではこの問題に対処するため、最適なターゲットオフセットを決定することにより、両時系列を時間的に整合させる手法を提案している。本稿では、実際に音声信号を数サンプル前後にスライドさせることによる、F_0 推定精度の変化を調べたので、報告する。

[1] Desheng Wang, et al., “Automatic temporal alignment for pitch estimation evaluation,” Proc. ICASSP2024, Seoul, Korea, Apr. 2024.

2-Q-23

聴覚障害者の異なる速度での文章音読における発話音声の韻律的特徴

Prosodic features of speech sounds in the reading aloud of sentences at different speech rates by hearing-impaired people

〇湯浅哲也（大教大・教）

◆人工内耳装用聴覚障害者４名を対象に，異なる速度での文章音読の音声を録音し，発話速度・音圧・基本周波数の変化を検討した。

◆その結果，以下の３点が明らかになった。

◆①速い発話になると，音圧が上がる（声が大きくなる）。

◆②遅い発話になると，基本周波数が高くなる（声が高くなる）。

◆③普通の速度での発話では，音圧及び基本周波数が他の速度での発話に比べて低い傾向が示された。

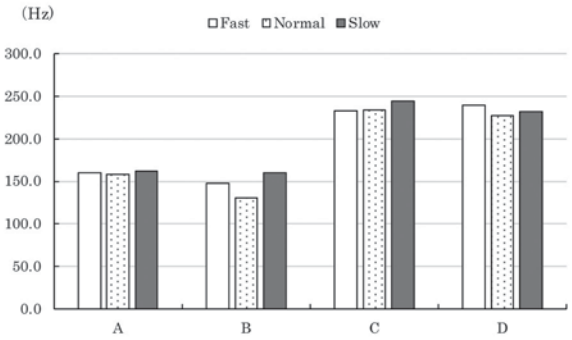


Fig.1 : Fundamental frequency at different speech rates

2-Q-24

講演取消

2-Q-25

2-Q-25 正規分布の確率質量関数表現による主観評価集計法の提案とその予測器学習への応用

Score aggregation via quantized distribution fitting and its application to predictor learning

◎近藤祐斗，亀岡弘和，田中宏，金子卓弘（NTT）

◆テーマ : 音声の主観品質を表す値を自動的に予測するタスクである主観音声品質予測（speech quality assessment: SQA）

◆動向 : 合成音声の SQA 手法として深層 MOS 予測モデルが主流

◆以前の我々の研究 : 新たな主観評価集計法である *N*-lowest MOS を提案し，これを SQA モデルの予測対象とすることで予測精度を向上

◆本研究 :『潜在的な連続値評点を 1～5 に量子化するプロセス』を仮定した集計法を提案（Fig. 1）．*N*-lowest MOS とは異なり，音声ごとに評価数が異なる場合に対しても適用可能で，かつ評価数に応じて値 *N* を設定する必要がないことが特徴

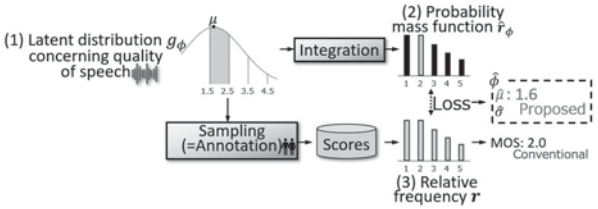


Figure 1 提案法の概略．ここではリスナーが想起する真の主観評価値は(1)潜在的な正規分布に従って生成され，観測される ACR スコアはそれを丸めた値であるとしてモデル化する．(2)確率質量関数と(3)ACR スコアの相対頻度分布との差異に基づいて潜在分布を推定し，そのピークを MOS に代わる主観評価値とする．

2-Q-26

2-Q-26　大規模事前学習モデルによる笑い声表現を用いた speech-laugh 音声の生成

Generating speech-laugh using laughter representation via large-scale pre-trained model

☆木全亮太朗，上乃聖，李晃伸（名古屋工業大院・工学研）

◆大規模音声モデルと FastSpeech 2 をベースに，自然な speech-laugh 音声の生成を目指す．

- ➢ 事前学習された FACodec を大規模音声モデルとして使用
- ➢ FastSpeech 2 内の Variance Adaptor に新たに Laughter Predictor を追加することでフレーム単位での笑い声の制御が可能．
  - ✧ label：データセットに付与された笑い声なら１となるラベル
  - ✧ prob.：laughter detection model により推定された笑いの確率

◆ベースモデルは，メルスペクトログラムを予測する FastSpeech 2 を使用

◆提案手法 (label，prob.)による speech-laugh 音声の自然性，話者類似度はベースラインモデルより高いスコアを示した．

◆また，提案手法による speech-laugh 音声の面白さの評価は提案手法 (label)が最も高いスコアを獲得．

Table 1．各テストセットにおける自然性，話者類似度，面白さの主観評価の結果．平均（左）と 95%信頼区間（右）を示す．

	MOS	SMOS	面白さ
Ground Truth	4.67 ± 0.12	-	4.16 ± 0.17
ベースモデル (label)	2.39 ± 0.20	2.58 ± 0.16	2.60 ± 0.18
提案手法 (label)	**2.77 ± 0.21**	**2.67 ± 0.16**	**2.74 ± 0.19**
提案手法 (prob.)	2.44 ± 0.21	2.58 ± 0.17	2.60 ± 0.20

2-Q-27

2-Q-27　喋り笑い合成に向けた基礎検討

Towards speech-laugh synthesis

☆近藤駿介，近藤和弘（山形大）

◆ 笑い声の研究で笑い声の音声データ、特に喋り笑いの不足とラベル付けの難しさが問題

→　セグメンテーションの基準を考案、各音声に対してラベル付け。

◆ セグメンテーションの新たな基準は以下の２条件。

・笑い声によって発生した明らかなポーズは発話部分とみなす

・呼気の後に付随する吸気音を発話終了時刻とみなす。

→　次の音声までに 400ms のポーズがない場合、その終了時刻を次の音声の開始時刻とする

→　音声長が短くなり、笑い声の後に発話が続くケースを減らす。

◆ 計 3,180 個の音声を「笑い」，「喋り笑い」，「喋りと笑い」，「その他」の４種にラベル付けを実施。

→　９割の音声ラベル付け完了

Table 1　分類済みラベルの構成

ラベル	音声数 (個)	割合 (%)
笑い	1,002	31.51
喋り笑い	784	24.65
喋りと笑い	1.016	31.95
その他	27	0.85
NON_LABEL	351	11.04

2-Q-28

2-Q-28　疑似電気喉頭音声コーパス

Pseudo-Electrolaryngeal Speech Corpus

◎小林和弘，荻田健一（名大・TARVO），
丹羽希碩，Lester Violeta，WenChin Huang，戸田智基（名大）

PES Pseudo-Electrolaryngeal Speech Corpus 疑似電気喉頭音声コーパス
TARVO　名古屋大学

◆喉頭摘出手術は声の発声器官を除去するため、発声機能に障害が残り、手術前の自然な声で話すことができなくなる。

◆発声機能に障害を持つ人（発声障害者）は、電気式人工喉頭やシャントなどの代替発声器により発声することが可能であるが、声質は機械的で自然な声とは大きく異なる。

◆その結果、他者との会話機会が減り、社会的な孤立や就業機会の減少などの問題が生じるといわれている。

◆この問題を解決するために、我々は手術前に声を収録・保存し、喉頭摘出後に患者の声を復元させる「Save the Voice プロジェクト」に取り組んでいる。本稿では、Save the Voice プロジェクトの概要およびプロジェクトを通じて公開された疑似電気喉頭音声（PES: Pseudo-Electrolaryngeal Speec）コーパスについて紹介する。

PES コーパス公開サイト
(https://github.com/k2kobayashi/PES-corpus)

2-Q-29

2-Q-29　模擬難聴システムを用いた難聴状態におけるポップアウトボイスの明瞭性と音響特徴の関係の調査

An investigation of the relationship between intelligibility of pop-out voice in hearing loss condition and acoustic features of the voice by using hearing impairment simulator.

☆上杉亮介，坂野秀樹（名城大院）

◆背景雑音から際立って目立つ音声をポップアウトボイスと呼ぶ．本研究では，このポップアウトボイスが難聴者に対しても明瞭度の高い音声として知覚されるか模擬難聴システムを用いた主観評価実験を行い調査した．

◆実験から得られた主観評価値と音響特徴量との相関分析を行った．

◆分析の結果，基本周波数および動的特徴量と評価値の間に目立った相関は見られなかったが、1kHz から 4kHz の帯域のパワーと評価値の間に強い正の相関がみられた．このことから，高域のパワーが大きい音声は難聴者にとって明瞭度の高い音声として知覚される可能性があると考えられる

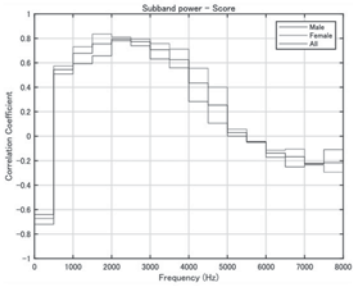


Fig.1：The correlation coefficient between the pop-out score and the subband power of the stimulus used in our experiment.

2-Q-30

2-Q-30　訛り制御音声合成のための話者類似度を利用した擬似同一話者による訛りあり・なしペアデータ作成の検討

Towards the Creation of Paired Accented and Non-Accented Speech Data from Pseudo Same-Speakers Using Speaker Similarity for Accent Control Speech Synthesis

☆高橋玲，能勢隆，伊藤彰則（東北大）

◆本稿では訛り制御 TTS の実現を目指し，擬似的な同一話者による訛りあり・なしのペアデータを作成する手法を提案する．

◆実環境における英語のコミュニケーションを考慮すると，リスニング学習では多様な種類の英語に触れることが重要である．

◆話速や訛りに含まれる特徴を制御可能なリスニング学習支援の実現のため，モーフィング技術を用いた訛りスタイルの制御を目指す．

◆特定の話者の訛り特徴のみを制御するためには，同一話者による訛りあり・なしのペアデータを用いた英語発話コーパスが必要となるが，その収録は容易ではない．

◆本稿では，L1，L2 コーパスから話者の声質を表現する話者特徴量に基づき擬似的な同一話者ペアを作成する．またその中に訛りの強度に一定の差異があるペアが存在することを明らかにする．

選択された擬似同一話者ペアにおける話者類似度及び訛り強度の差異（男女各上位５名）

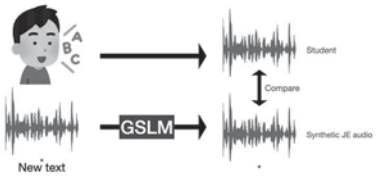
男性話者ペア	類似度	L2.GOP	Libri.GOP	DGOP	女性話者ペア	類似度	L2.GOP	Libri.GOP	DGOP
EBVS.4659	0.850	-4.880	-1.740	3.140	LXC.1422	0.850	-4.747	-2.129	2.618
ASI.3871	0.823	-4.460	-2.958	1.502	THV.5019	0.888	-4.730	-2.704	2.026
HKK.6993	0.872	-4.040	-2.624	1.417	YDCK.8014	0.899	-3.609	-1.860	1.749
HQTV.3060	0.837	-5.031	-3.736	1.295	SKA.1171	0.869	-4.543	-2.801	1.742
BWC.1924	0.853	-4.626	-3.355	1.271	SVBI.8295	0.812	-5.080	-3.476	1.603

2-Q-31

2-Q-31　Analyzing Language Learners' Oral Imitation Through Machine Model of Oral Imitation with Foreign Accentuation

外国語訛りが付与される機械による口頭模倣モデルを用いた英語学習者の口頭模倣に対する分析

☆Xiai Cheng, Kentaro Onda, Daisuke Saito, Nobuaki Minematsu (Univ. Tokyo)



◆ Identify the difficulties learners face in prosody imitation.

◆ Use the GSLM model to predict the prosody imitation difficulty of new practice materials.

Step1: Synthesize Japanese-English audio
American English (AE_i) --> GSLM--> synthetic JE_i

Step 2: Calculate value of correlation coefficient and difference between AE and synthetic JE / Learners' JE.
pitch_corr(AE_i, sJE_i),intensity_corr(AE_i, sJE_i)
pitch_corr(AE_i, LJE_i),intensity_corr(AE_i, LJE_i)

Step 3: Identify the challenges in student imitation through the correlation coefficients between students and AE.

Step 4: Plot scatter diagrams using the
[pitch_corr(AE_i, sJE_i),pitch_corr(AE_i, LJE_i)],
[intensity_corr(AE_i, sJE_i),intensity_corr(AE_i, LJE_i)]
to measure the similarity between the generated audio and learns' audio.

2-Q-32

2-Q-32　詩吟歌唱における歌唱フォルマント

Singer's Formant on Traditional Shigin Singing

☆柳原直也，中山仁史（広島市大院），加藤浩介（大阪大）

◆詩吟は和歌や漢詩を声調と呼ばれる旋律にそって唄う日本の伝統的歌唱である．

◆歌唱フォルマント Singer's Formant F_{SG} は咽頭腔を広げ，高次フォルマントを強く共鳴させることで歌声としての強い特徴を与える．

◆第 1 フォルマントの強度 L_1 で正規化した歌唱フォルマントの強度 $L_{SG}-L_1$ を用いる．

◆歌唱条件で統制した詩吟歌唱データベースを用いて正規化歌唱フォルマントを目的変数，歌唱条件を説明変数とした分散分析を行う．

◆各歌唱条件を変数とした歌唱フォルマントの関係を明らかにした．

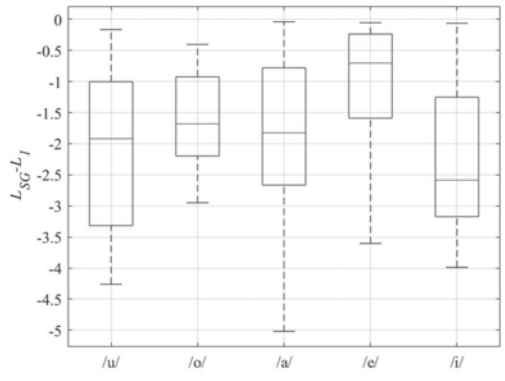


Fig.1: Singer's Formant L_{SG}-L_1 for each vowel.

2-Q-33

2-Q-33　歌唱音声の基本周波数に含まれる微細変動の有無が自然性に与える影響の調査

A study on effect on naturalness by existence of fine fluctuation of fundamental frequency in singing voice.

◎鈴木千文，坂野秀樹，旭 健作(名城大)

◆本研究ではこれまで以下の方法で抽出したノンビブラート音声の基本周波数に含まれる微細変動を観察した

- ノンビブラート音声の基本周波数の微細変動の抽出手順
  - ✧ 基本周波数から回帰直線の傾きである微分信号を算出
  - ✧ 微分信号に 10Hz のハイパスフィルタを用いて微細変動を抽出
- Fig.1 左より、微細変動にビブラートに類似した変動を観察することができる

◆ 次に基本周波数の微細変動に含まれるビブラートに類似した変動のピークに関する処理を行った

- 微細変動の対数スペクトルをフレーム分析し、フレーム内のピークの最大値を抽出した(Fig.1 右)
- 抽出したピーク値がビブラートに類似した変動のピーク値に近いことが分かった

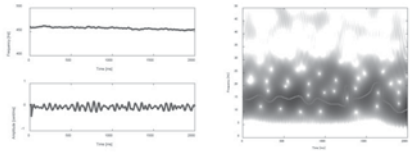


Fig1. Fundamental frequency and fine fluctuation (Left).
Spectrogram of fine fluctuation with peak of log spectrum of fine fluctuation (Right).

2-Q-34

講演取消

2-Q-35

2-Q-35　ワーピングを用いた電気式人工喉頭音声変換システムに適した発話例生成

Optimum speaking example generation with warping for an electrolaryngeal speech conversion system

☆宮下敦志，戸田智基(名古屋大)

◆音声変換システムユーザの入力発話制御をワーピングでモデル化

◆電気音声変換システムに適した発話制御を推定

◆ワーピングによる最適な制御により変換システムの出力音声品質が向上

◆破裂音を重視した発声制御が有効であることが示唆された

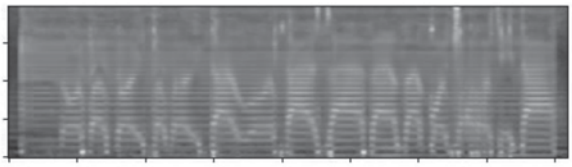


Fig.1 Mel-spectrogram of input electrolaryngeal speech

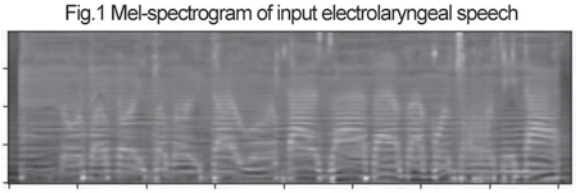


Fig.2 Mel-spectrogram of warped speech

2-Q-36

2-Q-36　製薬分野に特化した音声認識モデル PHASR の開発

井上顧基 (Elith)

◆製薬業界向けの音声認識モデル「PHASR」を開発した。

◆PHASR は、ReazonSpeech v2.0 と Mecon Audio に加え、独自に収集した約 10 時間の製薬ドメインのデータを活用して開発された高性能な音声認識モデルである。

◆PHASR は、特製薬業界における複雑な専門用語や独特の表現に対応できるよう設計されており、業界特有のニーズに応えるために開発されたものである。

◆モデルのサイズは 118MB と非常にコンパクトであり、エッジデバイス上での実行が可能である。

◆ Table 1 に CPU で動作可能な各音声認識モデルにおけるテストセット CER を示す。CPU で動作可能なモデルと他モデルと比較して、PHASR の認識精度が優れていることが確認できた。

◆また、Table 2 に各音声認識モデルのパラメータ数を示す。パラメータ数も 118M と軽量化に成功していることが分かる。

Table 1　各音声認識モデルにおける CER

ASR モデル	CER
PHASR	10.5
Whisper Medium	13.2
Whisper Small	15.9
Whisper Base	24.9
Whisper Tiny	35.2

Table 2　各音声認識モデルのパラメータ数

ASR モデル	パラメータ数
PHASR	118M
Whisper Medium	769M
Whisper Small	224M
Whisper Base	74M
Whisper Tiny	39M

2-Q-37

2-Q-37　wav2vec2.0 による低資源言語の音声認識における学習戦略や認識器の効果に関する研究

A Study on the Effectiveness of Different Recognizers and Training Strategy in Low Resource Speech Recognition Using wav2vec2.0

☆越川崇貴，伊藤彰則，能勢隆(東北大)

◆本研究では事前学習済みの wav2vec2.0 を使用した低資源言語の音声認識において，異なる学習戦略や認識器，フレームワークを適用した際の認識性能への影響について調査した．

◆実験では，学習データ量が 1 時間未満の 8 つの低資源言語において，5 つの学習戦略，LSTM, GRU を含む 9 つの認識器，Encoder-Decoder モデルを適用した際の CER を測定し，それぞれ比較した．

◆実験の結果，学習戦略ではモデルの下部を固定し，認識器としては bi-GRU を用いることで認識性能が改善されることが分かった．

◆また Encoder-Decoder モデルに適用した場合，学習データの多寡にかかわらず認識性能が悪化することが分かった．

Table 1 : CER(%) when changing training strategies, recognizers, frameworks

Training strategies

Update-RL	92.60
Freeze-Transformer	**13.38**
Layer Re-Initialization	13.51
Exclude-CNN	14.01
Update-All	16.88

Decoders

FC layer	**12.55**
TrDec(1 layer)	14.36
TrDec(4 layers)	14.86

Recognizers

FC layer	14.01
3 layer MLP	14.48
sim-RNN	14.96
2-RNN	15.99
bi-RNN	13.47
sim-LSTM	14.15
2-LSTM	17.87
bi-LSTM	13.83
bi-GRU	**13.20**

2-Q-38

2-Q-38　低頻度語のためのプロンプトを活用した音声認識

Automatic Speech Recognition Utilizing Prompts for Low-frequency Words.

◎菅野竜雅，佐藤裕明，佐久間旭，△熊野正，
△河合吉彦(NHK)，小川哲司(早大)

◆Encoder-decoder 型の音声認識モデルで、プロンプトを使って低頻度語など認識が難しい単語を認識する手法を提案。
◆音声認識モデルは、十分に学習されていない単語で、誤認識が多発。
◆提案法では、認識が難しい単語が事前に得られると仮定。
◆得られた単語の発音表記を用意し、モデルの decoder に入力。
◆モデルは該当する発音箇所で、無理に表記を推定せず、特殊トークンを出力。
◆最後に、後処理で特殊トークンを正しい表記に置き換える。

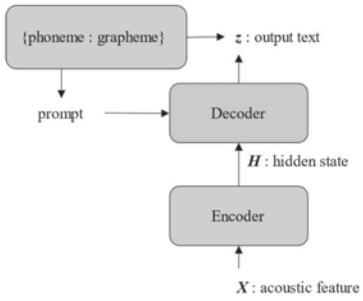


Fig.1 : Proposed method

2-Q-39

2-Q-39　Exploring Pseudo-Label Filtering Methods to Improve Rare Word Recognition

☆ Wen Shen TEO (UEC), Taishi YAMAOKA (Poetics.Inc)

◆Investigates methods for selecting pseudo-labels to improve an existing ASR model's rare word recognition.
◆Compares the importance of morpheme count λ_M and ASR Confidence λ_C as criteria for pseudo-label filtering.

Findings:
◆More data is better for pseudo-label finetuning.
◆Confidence-based filtering is not helpful for rare word recognition
◆Random filtering yielded the best results for both rare word recognition and overall ASR accuracy.

Top % Audio	Epoch	$\lambda_M > \lambda_C$	$\lambda_M < \lambda_C$	Random
Baseline	0	22.00		
1%	16	23.17	25.97	23.76
4%	4	21.51	22.73	**21.22**
16%	1	21.36	22.15	21.26

Table 1: Rare Word Error Rate (Rare WER) comparison from finetuning the baseline model with different pseudo-label subsets.

2-Q-40

2-Q-40　聴覚障害者音声におけるエンコーダ層置換による音声認識の高精度化

Improvement of speech recognition performance for deaf and hard-of-hearing based on encoder layer replacement

☆高橋 快斗（豊橋技科大），若林 佑幸（豊橋技科大），
太田 健吾（阿南高専），小林 彰夫（大和大），北岡 教英（豊橋技科大）

<背景>
◆以前の手法：Pretrain (LTV→DEAF)→ fine-tuning(LTV+JNAS→DEAF)
◆健聴者音声による fine-tuning 時に音響的ドメイン適応効果の忘却
<提案手法>
◆エンコーダ層を置換することにより，大規模コーパスの言語情報と自己教師あり学習モデルの音響情報を保持した音声認識器を作成

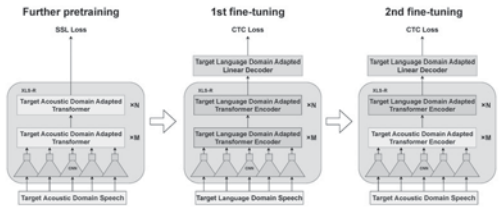


Fig. 1 The flow of constructing the proposed ASR model with gradual fine-tuning and layer replacement

<結果>
◆聴覚障害者音声において，エンコーダ層置換により認識精度が向上

Table 1 Dependence of CER on replacement of XLS-R layers

Layer replacement	JNAS CER	DEAF CER
w/o replacement	**8.3**	23.0
1–6	8.4	22.5
1–12	9.3	**22.1**
1–18	11.9	24.9
1–24 (all)	22.8	32.6
7–12	8.8	22.7

2-Q-41

2-Q-41　複数種類の構音障害音声を用いた wav2vec 音声認識モデル学習の検討

Training of wav2vec model using speech with multiple types of speech disorder

☆大谷赳[1]，相原龍[2]，高島遼一[1]，滝口哲也[1]，△田口進也[2]
([1]神戸大学，[2]三菱電機)

◆本研究では脳性麻痺による運動障害性構音障害者，口唇口蓋裂および舌切除による器質性構音障害者を対象とした音声認識を目的とする。
◆構音障害者の音声認識モデルを学習する際，構音障害者の音声を大量に収集することが困難であるという課題がある。
◆本研究では，利用可能な構音障害者音声を増やすことを目的に，ラベルの無い自由発話音声と，複数の構音障害者音声を学習に利用することを検討する。
◆提案手法では，wav2vec 2.0 の枠組みで，複数構音障害者音声のラベル無し音声を用いた自己教師あり学習を行う（Fig. 1）。
◆実験の結果，評価した全ての構音障害者に対して，構音障害者音声を用いて自己教師あり学習を行うことで認識誤り率が改善されることが確認できた。

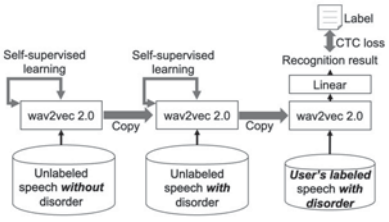


Fig.1 : The proposed model training procedure

2-Q-42

2-Q-42　雑音・無音ラベリングによる音声区間検出が不要な音声認識

Automatic speech recognition without voice activity detection using noise/silence labeling

☆坂井利光，若林佑幸，北岡教英 (豊橋技科大)

目的

◆音声区間検出の誤りは後段の認識処理では回復できない.

◆ノイズも含めて学習させることで、音声区間検出が不要な音声認識モデル作成を試みる.

手法

◆２つの音声サンプルを無音区間を挟んで連結し，さらにノイズを重畳した音声データを学習に使用する.

◆教師用テキストデータは，２つのテキストを結合させ，追加無音区間に対応する箇所に雑音・無音ラベル“＊”を１つ付与したものを使用する.

結果

◆低SNRの検証データにおいて，提案手法はVADを使用した場合よりも高精度であることが確認できた.

Table 1: 各実験条件の CER [%]

Valid type	train data architecture			SNR [dB]			
	Model name	Concat	Noisy	∞	15	10	5
w/ VAD	**n_short**	-	✓	7.88	8.59	9.93	13.89
w/o VAD	**c_long**	✓	-	9.23	10.76	13.02	18.91
	n_short	-	✓	9.25	10.02	11.36	14.95
	n_long	✓	✓	10.15	10.66	11.77	15.08
	n_long+∗	✓	✓	8.73	9.14	9.96	12.48

2-Q-43

2-Q-43　CHiME-7 遠隔発話音声認識タスクに対する大規模言語モデルを用いたNベスト音声認識仮説のリスコアリング

Applying LLMs for rescoring N-best ASR hypotheses of CHiME-7 DASR task

○小川厚徳，加茂直之，松浦孝平，芦原孝典，森谷崇史，叶高朋，傳直弘，マーク・デルクロア (NTT)

◆大規模言語モデル(LLM)を音声認識において利用する最も簡単な方法は，Nベストやラティス形式の複数音声認識仮説のリスコアリング(リランキング)に用いる方法であり，従来研究でその効果が確認されている.

◆しかし従来研究は，読み上げ音声，独話音声，比較的フォーマルな会話音声を認識対象としているため，カジュアルな(非常にくだけた)会話音声に対するLLMを用いたリスコアリングの効果は明らかではない.

◆そこで本研究では，カジュアルな会話音声を認識対象とするCHiME-7遠隔発話音声認識タスクに対して，Llama2-7Bを用いたNベストリスコアリングを行い，特に，LLMのドメイン適応の効果と，リスコアリングを行う際に過去の文脈を考慮することの効果を明らかにした.

◆Llama2は長い文脈を考慮することで，ドメイン適応なしでも，ドメイン適応を行った場合を同等の精度を示すこと，及び，ドメイン適応はLlama2の最高精度を達成するために必要な文脈長を短縮する(計算量コストを削減できる)ことが確認できた.

Table. N-best rescoring results in WERs [%] obtained with Llama2 (w/o domain adaptation) and domain-adapted (by QLoRA) Llama2 of six context lengths L on the CHiME-7 distant ASR task consisting of the CHiME-6, DiPCo, and Mixer 6 datasets.

		CHiME-6		DiPCo		Mixer 6	
Model	L	W/o adapt	QLoRA	W/o adapt	QLoRA	W/o adapt	QLoRA
ASR 1-best	---	26.2	---	25.5	---	15.8	---
Llama2-7B	0	26.1	25.7	26.5	25.7	16.2	15.5
	64	25.5	25.3	25.8	25.4	15.7	**15.4**
	128	25.4	25.3	25.6	**25.3**	15.6	**15.4**
	256	25.4	**25.2**	25.5	**25.3**	15.5	**15.4**
	512	25.3	**25.2**	25.4	**25.3**	15.5	**15.4**
	1024	25.3	**25.2**	**25.3**	**25.3**	**15.4**	**15.4**
Oracle	---	17.2	---	18.0	---	11.6	---

2-Q-44

日本語諸方言コーパスを利用した全国地域方言の言語モデルおよび識別モデルの構築と比較分析

Construction and analysis of language and identification models for regional dialects nationwide using a Japanese dialects corpus

☆神谷悠太，甲斐充彦（静岡大），Raufun Nahar(阿南高専)，中川聖一（中部大）

◆本研究では、日本語諸方言コーパス(COJADS)のカナ書き起こしを利用して全国８地域の方言識別を行った。方言識別を通して方言地域の言語的特徴の自動獲得の可能性を調べた。

◆方言地域の識別手法として、BERT、LSTM、ngram による地域別言語モデルや識別モデルを用いた手法の性能を比較分析した。

◆BERT 単一モデルで識別する手法ではカナ文字のみの長期発話(34 文字以上)に対して識別率 **82.1%** を達成した。(Table 1, Fig. 1)
また、方言音声認識を利用した場合に識別率 **70.7%** を達成した。

Table 1: Comparison of dialect identification rates (%) by long text units across 8 regions

Discrimination Model	
BERT	LSTM
82.1	74.3
Generative Model		
BERT	LSTM	4gram
67.8	78.4	80.8

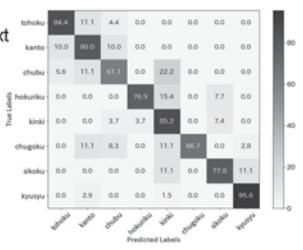


Fig. 1: Confusion matrix of the BERT -based dialect identification model for long text units (longer than 34 characters)

2-Q-45

2-Q-45　TED 講演音声の音声認識性能と音声翻訳性能の関係 —その 2—

Relationship between speech recognition accuracy and translation performance on TED talks — part 2 —

☆増田健人，山本一公，中川聖一（中部大）

◆本稿ではTED 音声講演(IWSLT2016)を Whisper の Medium モデルと Large モデルによって音声認識を行い、単語誤り率(WER)を求めた。

◆テキスト入力と音声入力(音声認識結果)を Transformer 翻訳モデルで翻訳し、翻訳評価値(BLEU)を求めた。

◆Transformer 翻訳モデルにはソース言語とターゲット言語のペアとターゲット言語とソース言語を入れ替えたペアからなるパラレルデータセットで学習された双方向翻訳モデル(ベースライン翻訳モデル)とデータ拡張で学習された双方翻訳モデル(最良翻訳モデル)を使用した。

◆テキスト入力翻訳の BLEU と、音声入力による音声翻訳結果の BLEU を比較した際の低下率(r-rate)と WER の相関を求めた(相関値 0.44)。

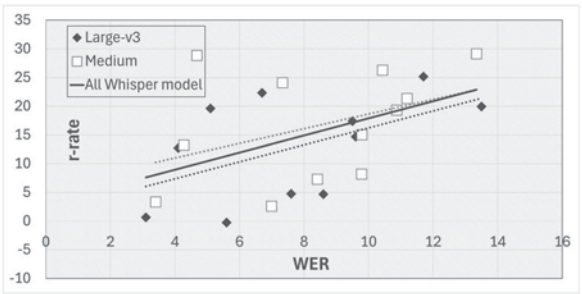


Fig.1: Relation of WER of ASR and reduction rate of BLEU by speech translation to text translation (best translation model)

2-Q-46

2-Q-46　敵対的生成ネットワークを用いた小規模音声データベースの話者年齢推定

Speaker Age Estimation using Generative Adversarial Network with Small-Scale Speech Database.

☆藤居 謙, 西村 竜一（和歌山大）

◆話者年齢埋め込み Age-vector を用いた手法を年齢推定タスクに導入

- 抽出 NN：発話音声から Age-vector を抽出
  AgeVoxCeleb（大規模データ）で学習
- 3 層の全結合層 NN：抽出した Age-vector を入力とし、出力は 155 クラス（年齢と性別の組み合わせ）

◆比較的小規模なデータセットからの識別器の構築を想定

- NN を 2 段に分割することで、大規模に事前学習した抽出 NN によって Age-vector を抽出し、分類器の精度を確保できる

◆識別器を全結合 NN と SGAN で比較

- SGAN：敵対的な学習によりデータ拡張の効果が期待される
- 「年齢誤差（推定年齢と実際の話者年齢の差）」を尺度とした検証
- ほぼすべての年齢層で SGAN モデルの年齢誤差が小さい
  ✧ **SGAN モデルにより年齢推定精度が向上**

Table. 1 全結合NNモデルとSGANモデルの年齢層別平均年齢誤差
（緑：年齢誤差が小さい，赤：年齢誤差が大きい）

	男性		女性	
	全結合NN	SGAN	全結合NN	SGAN
1歳 ~ 4歳	7.10	7.01	6.06	3.07
5歳 ~ 8歳	4.83	2.97	4.44	2.44
9歳~ 12歳	5.73	3.58	4.19	2.63
13歳~16歳	7.08	3.23	5.71	3.45
17歳~20歳	5.82	2.51	7.52	4.47
21歳~30歳	9.49	6.73	11.85	9.88
31歳~40歳	12.97	15.28	17.69	12.01

2-Q-47

2-Q-47　話者埋め込み上の損失に基づく話者認証に対する敵対的サンプル生成手法の検討

A study on generation method of adversarial example for speaker verification based on speaker embedding loss

☆北村悠, 齋藤大輔, 峯松信明(東大)

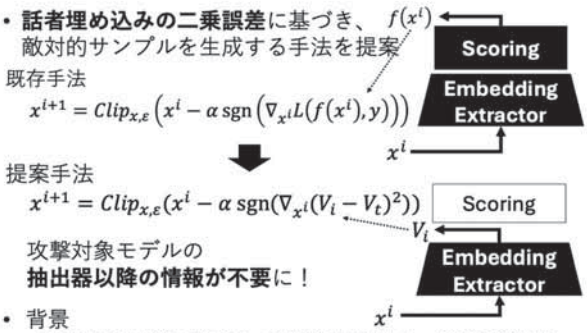
- **話者埋め込みの二乗誤差**に基づき、敵対的サンプルを生成する手法を提案

既存手法

$$x^{i+1} = Clip_{x,\varepsilon}\left(x^i - \alpha \operatorname{sgn}\left(\nabla_{x^i} L(f(x^i), y)\right)\right)$$

提案手法

$$x^{i+1} = Clip_{x,\varepsilon}(x^i - \alpha \operatorname{sgn}(\nabla_{x^i}(V_i - V_t)^2))$$



攻撃対象モデルの**抽出器以降の情報が不要**に！

- 背景
  - 話者認証は敵対的サンプル等に脆弱で、研究が必要
  - 敵対的サンプルは、攻撃するモデルが既知である前提
    →未知のモデルへの攻撃性能の向上を目指す
- 実験
  - 提案手法と既存手法で、3つの話者認証モデルを攻撃

全モデルに対し、高い攻撃成功率！

Method	Victim Model: XV+PLDA	XV+CS	XV+DNN
Proposed	**0.999**	**1**	**0.883**
XV+PLDA	1	0.999	0.838
XV+CS	1	1	0.861
XV+DNN	0.686	0.027	0.889

Fig. 1 : Attack Success Rate against 3 Speaker Verification Models

2-Q-48

2-Q-48　音響的特徴に基づくターンテイキング予測

Turn-taking prediction based on acoustic features

☆杉山雅和, 西村良太(徳島大), 太田健吾(阿南高専), 北岡教英(豊橋技科大)

◆背景

- 音声対話システムにおいてスムーズな対話を実現するため、システムが発話するか否かの判定を高速に行いたい。
- 先行研究では音響特徴量と音声認識結果を併用しており、また、発話の終了判定に長い無音区間を必要としていた。

◆提案手法

- 音響特徴量のみを入力とし、音声認識処理待ちの遅延を省略する。
- 分析対象を 100ms 以上の無音で区切られる発話とすることで、発話終了からモデル予測開始までの待ち時間を短縮する。

◆結果

- 日本語対話に対するターンテイキング予測において、正解率 75%を達成した。

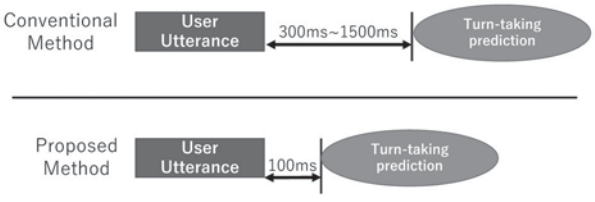


Fig.1: Comparison of turn-taking prediction timing

2-Q-49

2-Q-49　YouTube を利用した大規模会話コーパス構築のための複数話者動画抽出

Extracting multi-speaker videos from YouTube for building a large-scale conversation corpus

☆米山蒼祐, 森大毅（宇都宮大・エ）

◆本研究では、YouTube 動画から大規模対話コーパスを作成するために、話者が複数人登場する動画を取り出す。

◆動画の音声、字幕、画像 (サムネイル、動画内のシーン)、から特徴量を抽出し、ナイーブベイズ分類器で統合する (Fig. 1)。複数話者動画の分類を行なった結果、適合率 0.777、再現率 0.625、F 値 0.693 となった。

$$P(C_k|x_1, x_2, \cdots) \propto P(C_k) P_a(x_a|C_k) P_l(x_l|C_k) \prod_n P_{vn}(x_{vn}|C_k)$$

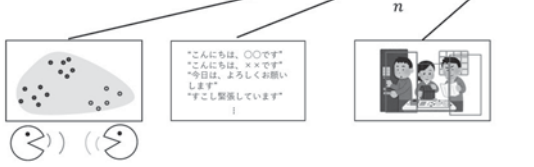


Fig. 1 Integration of acoustic, linguistic, and visual features using naive Bayes.

2-Q-50

2-Q-50　wav2vec 2.0 による言語特徴を考慮した音声対話システムへの入力発話の終了時間推定

Estimation of the end time of input utterances to a spoken dialog system considering linguistic features using wav2vec 2.0

☆金井良賢，若林佑幸（豊橋技科大），西村良太（徳島大），
北岡教英（豊橋技科大）

目的

◆先行研究 [1] で音声対話システムに入力された発話がいつ終了するのかを事前に推定する手法が提案されている．

◆この実験では言語特徴を考慮して，推定精度の向上を試みる．

手法

◆システムに入力された過去の発話情報を用い，現在の発話が終了するまでの残り時間を推定可能な深層学習モデルを構築する．

◆音声認識用に Fine-tuning された wav2vec 2.0 から得られる特徴量を用いて実験を行う．

結果

◆言語特徴を考慮した特徴量を用いることにより，推定精度が向上した．

Table 1: Experimental results of utterance end time prediction. FT indicates Fine-tuning.

Feature		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
MFCC		29.1	28.0	29.1	28.0
CNN		31.2	29.8	31.2	29.8
wav2vec 2.0 w/ FT	hidden states	**37.0**	**35.8**	**37.0**	**36.1**
	logits	36.9	35.6	36.9	36.0
	symbol sequence	27.4	25.8	27.4	23.4
wav2vec 2.0 w/o FT	hidden states	35.1	33.8	35.1	34.0

[1] 金井ら，“円滑な音声対話システムのための発話終了時間の事前推定”，音講論集（春），pp. 961–964, 2024