3-1-1

3-1-1　フェーズドアレーパラメトリックスピーカにおける周波数変調波の位相反転に基づく大音圧狭指向性ビームの形成

High-power Narrow-edged Beamforming Based on Phase-inversion of Frequency-modulated Wave in Phased-array Parametric Loudspeaker

◎耿　毓庭(立命館大), 中山　雅人(阪産大), 西浦　敬信(立命館大)

パラメトリックスピーカ(PAL)は超音波の直進性を利用ことで狭指向性を実現できるが，小型 PAL では狭指向性の実現が難しい傾向がある．我々はこれまで，PAL から放射する振幅変調の変調波に対し，位相を反転することで，狭指向性ビームの生成手法を提案してきたが，復調音圧の低下とサイドローブの発生という問題がある．そこで，前者の解決に向けて，復調音圧が高い周波数変調方式に着目し，位相反転を利用することで，大音圧と狭指向性の両立を目指す．本稿では，周波数変調波のスペクトル構造に基づき，位相反転方法を検討し，超音波と復調音の両者の位相打消しを利用できる条件を２つ提案する．具体的には，提案手法の条件１では周波数変調波の位相偏移を反転し，条件２では振幅と位相偏移の両方を反転する．実測による評価実験の結果，提案手法は狭指向性が実現できていることを確認した．また，提案手法で形成した指向性は，目的信号の周波数依存性が低く，低周波音を再生する際にも狭指向性を形成できる(Fig.1).

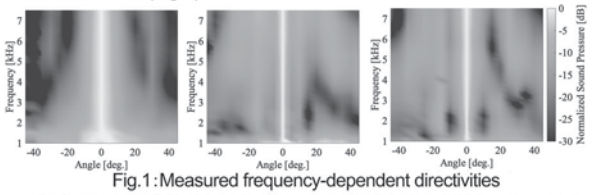


Fig.1 : Measured frequency-dependent directivities
(Left: common PAL driven by frequency modulation; Middle: proposed method with condition 1; Right: proposed method with condition2)

3-1-2

3-1-2　バーガース方程式がもたらす非線形歪の補償

Compensation of nonlinear distortions caused by Burgers' equation

☆髙橋威安，島内末廣，青木茂明(金沢工大)

◆鋭い指向性を有することで知られるパラメトリックスピーカは、超音波が空気伝搬過程において受ける非線形性により、可聴音が復調される再生技術であるが、復調時に可聴音の音質の劣化が問題となる。

◆本研究では、パラメトリックスピーカの歪補償の基礎検討として、空気伝搬の非線形性の最も簡単なモデルとされるバーガース方程式に注目した。

◆歪補償を実現するために、LSTM (Long Short-Term Memory)に基づくニューラルネットワークを適用した。

◆歪補償の一例として、1.0 [kHz] と 2.5 [kHz] の混合信号に対する補償の有無の様子を Fig.1 に示す。

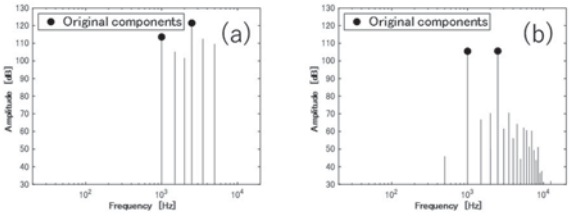


Fig.1 : Spectrum comparison of output for mixed signal (1.0+2.5 [kHz]). (a) Without compensation and (b) With compensation.

3-1-3

3-1-3　Acousto-optic sound-field reconstruction based on envelope circle sampling projections

○ Phuc D. Nguyen, Kenji Ishikawa, Noboru Harada, Takehiro Moriya (NTT)

In this study, we propose a novel reconstruction method based on an envelope circle sampling scheme, employing circular harmonics expansion. The proposed method successfully reconstructs the sound field even when the sound source is located inside the reconstruction area. This proposed technique facilitates the visualization of sound radiation from a loudspeaker with reduced effort.

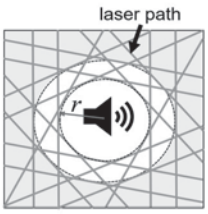


Fig. Circle envelope sampling

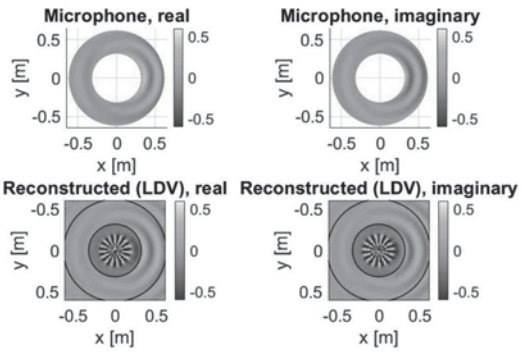


Fig. Sound field reconstruction

3-1-4

3-1-4　光周波数による音圧計測

Sound pressure measurement by optical frequency

○石川憲治，△白木善史，守谷健弘(NTT)，石澤淳(日大)，日達研一，小栗克弥(NTT)

光周波数を用いて音圧を測定する新手法を提案

- 音によって生じるファブリペロー共振器の共振周波数変動を光周波数コムを用いて精密測定
- 音圧単位パスカルの直接実現方法として度量衡応用を期待

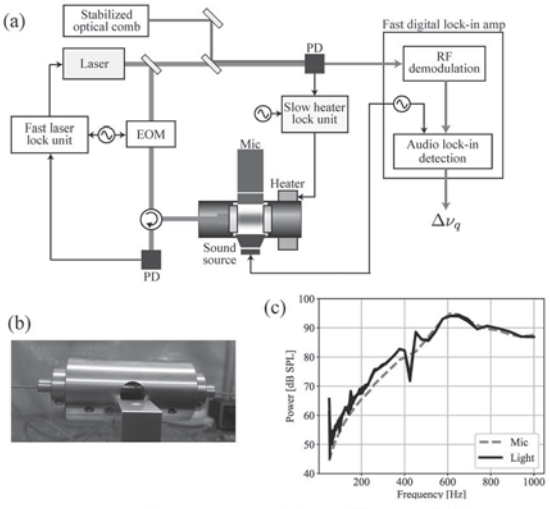


Fig.1 (a) 測定システム，(b) 共振器の外観，(c) 測定結果

3-1-5

3-1-5　周波数領域カルマンフィルタを用いた非定常信号の短時間区間信号予測に関する検討

A study on short-time signal prediction of non-stationary signals using frequency domain Kalman filter

◎小塚詩穂里，伊藤弘章，鎌土記良(NTT)

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

◆対象騒音が非定常である場合，その構成要素の相対配置やアルゴリズム遅延，二次系の遅延などによる消音性能劣化を防ぐため，参照信号を予測することは有望である.

◆計算負荷の軽量なカルマンフィルタに着目し，時間領域カルマンフィルタでは数サンプル先の予測が限界だが時間周波数領域でカルマンフィルタによる予測を行うことで短時間区間先の参照信号を予測可能な手法を提案し，シミュレーション実験により性能評価した.

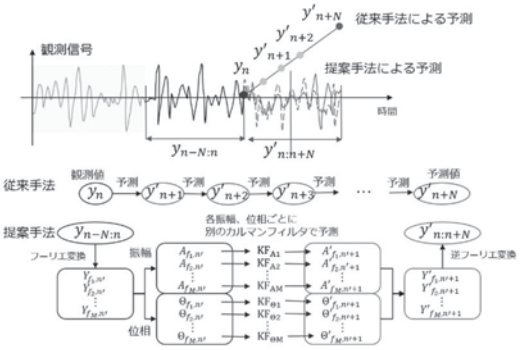


Fig.1 : Overview of the proposed system

3-1-6

3-1-6　フィルム状圧電スピーカを用いた能動騒音制御技術のセミオープン型ワークブースへの適用検討

Active noise control system using piezoelectric film speaker for sound leaking from semi-open workbooth

○大戸康平(日東電工)，辻一晟，梶川嘉延(関西大)

◆回折音を広い面積（体積）で消音可能なフィルム状圧電スピーカを用いた能動騒音制御技術（ANC）をセミオープン型ワークスペースに適用し、利用者の音声漏洩を防止する ANC システムの検討を行った。

◆回折音の影響が大きい 100～700Hz に対し、図１、２のような配置で case(1-2-2)フィードフォワード型 ANC を用いて制御を行った。

◆ANC によって 6dB 以上消音できた領域を R.A(-6dB)と定義し評価を行った。その結果、図２のようにブース入口前方の 60×80×80cm の空間において提案システムは 14.1%の消音領域改善を達成した。

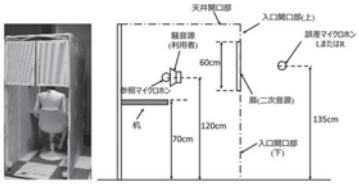


図１　システム配置図（左：写真、右：断面図）

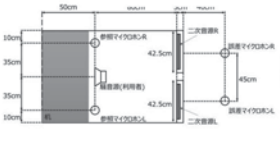


図２　システム配置図（俯瞰図）

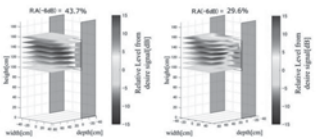


図３　消音領域（左：フィルム状圧電スピーカ、右：ダイナミックスピーカ）

3-1-7

3-1-7　フィルム状圧電スピーカを用いた個人ブース ANC におけるバーチャルセンシングの導入に関する検討

A Study on Introduction of Virtual Sensing in Personal Booth ANC Using Piezoelectric Film Speakers

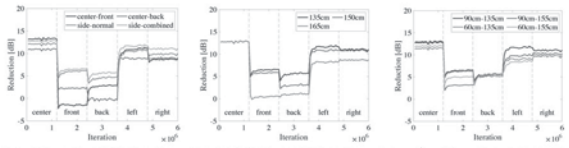
☆辻一晟，大戸康平(日東電工)，梶川嘉延(関西大)

◆両開きの開閉扉に二次音源としてフィルム状圧電スピーカが搭載された，テレワークブースを模した半開放型の個人ブースに，ANC を導入した個人ブース ANC システムについて検討した.

◆本システムは，ブース内の音声を，マイクロホンを設置できない人の行き交うブース外で消音するため，バーチャルセンシングを導入する.

◆低減対象となる音声は，人の動きによって発生源が変動することから，変動に対してロバスト性が保たれるマイクロホン配置について検討した．本稿では騒音源を中央，前後への移動，中央位置での左右への放射角度の変動における，騒音低減量について比較した.

◆その結果，Fig.1 に示すように参照マイクロホンを騒音源が中央の位置の左右に２つ配置して経路を結合させた場合，誤差マイクロホンを 135cm の高さに配置した場合，バーチャルマイクロホンを高さ 135cm・誤差マイクロホンとの距離 90cm に配置した場合が最もロバストとなり，その理由は，この配置が各経路の位相のずれが最も小さいためであると考えられる.



(a) 参照マイクロホン　(b) 誤差マイクロホン　(c) バーチャルマイクロホン

Fig.1 : 各マイクロホン配置における騒音低減量

3-1-8

3-1-8　３次元空間の騒音源移動を考慮した ANC システムに関する検討

Study on ANC System Considering Noise Source Movement in 3-D Space

☆豊岡祥太，梶川嘉延（関西大）

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

◆本稿では，固定フィルタ ANC システムに音源定位を用いることによって，所望地点に誤差マイクロホンを設置することなく騒音源移動にロバストなシステムを提案する.

◆提案法では，４つの参照マイクロホンからの受音信号から相互相関法を用いて３次元の騒音源位置を推定し，騒音源位置に対応した騒音制御フィルタを選択することによって ANC システムを動作させる.

◆Fig. 1 に誤差マイクロホン地点の時間波形を示す．提案法では，騒音源移動の直後や誤ったフィルタを選択した場合を除いて，約 20dB の高い騒音低減効果が得られていることがわかり，誤差マイクロホンなしでも動作する.

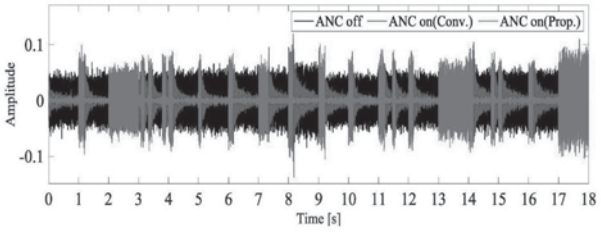


Fig.1 : 提案法と従来法のフィードフォワードANCシステムの時間波形

3-1-9

3-1-9　*n-n-n* マルチチャネル能動騒音制御システムの一構成法

A Method of configuring *n-n-n* multi-channel active noise control system

〇藤井健作（コダウェイ研），△棟安実治（関西大・工），苣木禎史（千葉工大）

マルチチャネルシステムでは，騒音制御フィルタの係数は騒音検出マイクロホンで採取された１次騒音間の無相関成分を利用して行われる．このことは逆に，無相関成分が含まれない１次騒音を採取した騒音検出マイクロホンは利用の必要がないことを意味している．

図１は，１次騒音から分離された相関成分について実行される提案法の構造を示している．この構造において $G(z)$は，２次系行列から計算された余因子行列から構成される簡易逆フィルタシステムを挿入することによって騒音制御フィルタと誤差マイクロホンが１対１で結ばれる経路の伝達関数を表している．この 1-n-n システムを適用することで相関成分が低減される．更に，残された各騒音制御フィルタの独立成分についても相関成分を取出して図１に示す演算を繰り返せば，騒音は更に低減され，最後には無相関成分だけが残る．そのように残された無相関成分の全てについて図１に示す演算を実行すれば騒音検出マイクロホンから誤差マイクロホンに到来した１次騒音は全て低減される．

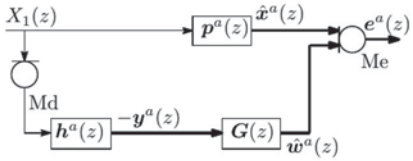


Fig.1：Proposed configuration of 1-n-n system controlling correlation component.

3-2-1

3-2-1　コンクリート非破壊計測のための非接触音響探査法に関する研究

-複数台の LDV を用いた移動計測の基礎検討-

Study on noncontact acoustic inspection method for non-destructive testing of concrete - Basic study of mobile measurement using multiple LDVs

〇杉本恒美，中川　裕，杉本和子，上地　樹（桐蔭横浜大・院），
△黒田千歳，△歌川紀之（佐藤工業㈱），△二瓶靖和（富士フイルム㈱）

◆少子高齢化の進展に伴い，点検作業の省力化・自動化が求められているため、低速でも構わないから，非接触音響探査法を用いた移動計測が実現できないかとの要望が増えてきた。

◆そこで，走査機構を持たない複数の LDV を移動台車に搭載して移動させ，後方から音波照射を行った場合の基礎検討を行った。

◆直径 300 ㎜，深さ 40 ㎜の円形模擬空洞欠陥（25 ㎜厚の発泡スチロール）を対象として移動計測実験が行われた（計測領域図 Fig.1(a)）。

◆音圧 120dB 時の送信信号と受信信号の相関値のエネルギー表示を用いた映像結果（Fig.1(b)）より，低速であれば移動計測が可能であることが明らかになった。

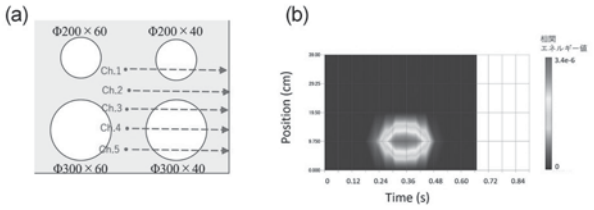


Fig.1 (a) Measurement area, (b) defect detection by correlation energy value during movement measurement (0.48 m/s).

3-2-2

3-2-2　加振信号を用いた CNN による外壁タイルの健全性評価

Soundness evaluation method by CNN using excitation signal for tile exterior wall,

〇西井朋也（東洋建設・技研），森淳一，森長誠（神奈川大学）

◆建物の外壁に用いるタイルの剥離・剥落事故を未然に防ぐため，タイル施工後の健全性に関する診断技術が重要である。

◆本研究では試験体を製作し，タイルに振動を与え，その加振応答から，畳み込みニューラルネットワーク（CNN）による接着剤の塗布状況を識別した。

◆1 枚のタイルに対し，6 個の受信センサ（M1～M6）を搭載した装置を用い，計測を行った。各受信センサにて記録したデータはそれぞれ Fig.1（a）～（f）のように表した。

◆Fig.1 を学習データとした識別では正解率は 95.8％であった。

◆各受信センサの結果（Fig.1）を１枚の画像データに結合したものを学習データ（Fig.2）とした識別では正解率は 99.5％であった。

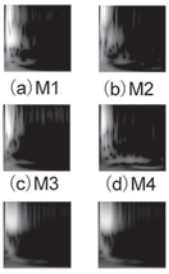


Fig.1：Example of scalograms

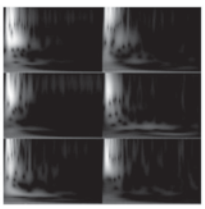


Fig.2：Example of Image-integrated

3-2-3

3-2-3　圧縮センシングを用いた空中超音波伝搬画像の再構成

Reconstruction of an airborne ultrasound propagation image using compressed sensing

☆濱田郁哉，清水鏡介（日大院・理工），大隅歩，伊藤洋一（日大・理工）

◆空中非線形超音波源走査法を用いた非破壊検査方法の高速化について研究を行っている．提案手法の高速化を妨げる根本的な要因は計測点数の多さにあるが，計測点数と可視化精度はトレードオフの関係にあり，単純に減らすことはできない．

◆近年，少数の計測点数より所望の結果を再構成する圧縮センシングが提案されている．

◆本検討では，高調波を含む空中超音波音場の計測に圧縮センシングを適用することを試み，少数の計測点数により再構成した伝搬画像について検証した．

◆Fig. 1 に音源のパイプ開口端から放射された音場を示す．図(a)にフルサンプリング結果，図(b)に再構成結果(フルサンプリング結果のうち 25%の結果から再構成)を示す．

◆結果より，25%の実験結果を用いて伝搬像を再構成できることを確認した．

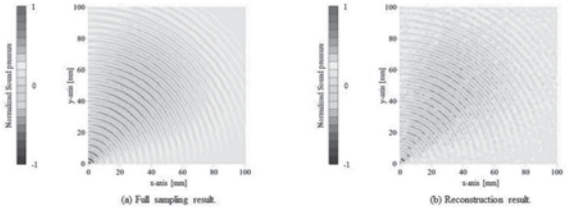


Fig .1 Propagation of airborne ultrasound (81.6 kHz).

3–2–4

3-2-4　せん断波速度の解析条件がミクロ構造の評価能に与える影響のシミュレーション検証

Simulation verification of the effect of analytical conditions of shear wave velocity on the ability to evaluate microstructures

☆大里晃大[1]，大栗拓真[2]，神山直久[2]，平田慎之介[3]，吉田憲司[3]，山口匡[3]
（[1]千葉大院・融合理工，[2]GE ヘルスケア・ジャパン，[3]千葉大・CFME）

◆Shear wave velocity (SWV) の解析条件がミクロ構造の評価能に与える影響について，脂肪量違いの肝臓を模擬した２次元空間モデルを対象として，弾性FDTD計算機シミュレーションを用いて検証した．

◆せん断波は脂肪滴の境界において散乱や屈折の影響を受けるため，脂肪量が増加するにつれて，せん断波の波面の乱れが大きくなり，波面の間隔が狭くなることが確認された．(Fig. 1(a))

◆SWV 評価の分解能が十分細かいと脂肪滴の影響を追跡することができるが，SWV 評価の分解能を粗くすると平滑化され，SWV 評価の結果が安定化した (Fig. 1(b))．臨床用診断装置でも同等の効果を踏まえた評価がなされていると想定される．

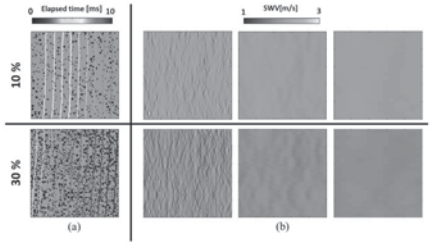


Fig.1 (a) Shear wave propagation from 4.5 ms to 10 ms and (b) SWVMAPs with analysis resolution of 10, 100, and 300 μm, respectively

3–2–5

3-2-5　高周波超音波イメージングにおけるビームステアリングと振幅統計量の関係性の評価

Evaluation of the relationship between echo-envelope statistics and beam steering in high-frequency ultrasound imaging

☆奈良 千尋，吉田 憲司，山口 匡，平田 慎之介（千葉大）

◆医用超音波画像のテクスチャ（振幅分布）を統計的に解析することで，組織性状診断を行うことができる．

◆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手法を検討している．

◆リニアアレイプローブを用いて，±15°の範囲を 3°間隔でビームステアリングを行って超音波画像を 11 枚作成した．各角度の３次モーメントマップおよびそれらをコンパウンドしたもの Fig.1 に示す．

◆Fig. 2 のように解析領域（図中三角形内）の３次モーメントを比較すると，統計ゆらぎを低減することができた．

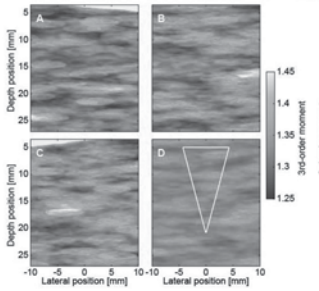


Fig.1 : 3rd-order-moment maps of ultrasound images, the steering angles of -15° (A), 0° (B), 15° (C), and the compounded one (D).

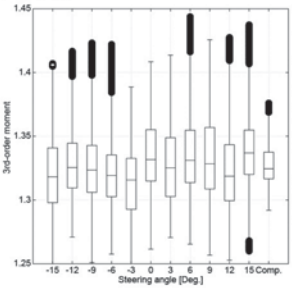


Fig.2 : Box-and-whisker plot of 3rd-order-moments in the regions of interest of ultrasound images.

3–2–6

3-2-6　多種の散乱体間の干渉状態が後方散乱特性に与える影響に関する検討

Study on the influence of interference state between multiple types of scatterers on backscattering characteristics

☆沓沢駿人（千葉大・院融合），平田慎之介，吉田憲司（千葉大・CFME），
△Emilie Franceschini（Aix-Marseille Univ.・LMA），山口匡（千葉大・CFME）

◆二種の散乱体が混在したファントムを計算機上で作成し，数理的検証を行うとともに，同条件で実体ファントムを作成し，35 MHz の単一凹面振動子を用いて後方散乱係数を評価した．

◆数理的検証では各散乱体の理論値の合算値と一致したことから，複数種の散乱体からのエコーの相互干渉が影響を及ぼしあっている可能性が示唆された（Fig. 1 (a)）．

◆実測評価においては強散乱体からの信号が支配的であり，弱散乱体からの信号の干渉が後方散乱特性に与える影響は小さいと考えられる（Fig.1 (b)）．

◆実測では，計測に使用した振動子の周波数帯や S/N の影響が大きいと想定されるため，今後はより高強度での観察が可能な高周波リニアアレイプローブを用いた検討を行う．

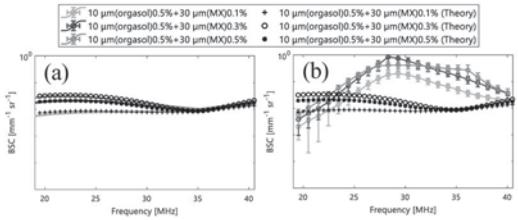


Fig. 1: Frequency dependence of BSC of two-scatterer mixed phantoms (a) in mathematical evaluation (b) in measured evaluation

3–2–7

3-2-7　バースト波支援動的造影超音波による単一マイクロバブルの可視化

Visualization of a single microbubble by burst-wave-aided contrast enhanced active Doppler ultrasnography

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

◆これまでに、リンパ管の可視化方法として，音響放射力で移動させた造影剤をドプラ法により可視化する方法を提案

◆造影剤の破壊を抑制しながら移動させる方法として、移動用の低振幅バースト信号と画像化用パルス信号を交互に送信する方法を提案

◆流路ファントムを用いた実験により、提案法の有効性を検証（造影剤の数密度が点拡がり関数あたり１個以下の条件で実施）

◆従来の送信方法（PID 法）と比較した結果、画像化用パルス波の音圧振幅が同じ条件であれば、コントラスト比において提案法が優れ、造影剤の破壊も生じないことを確認

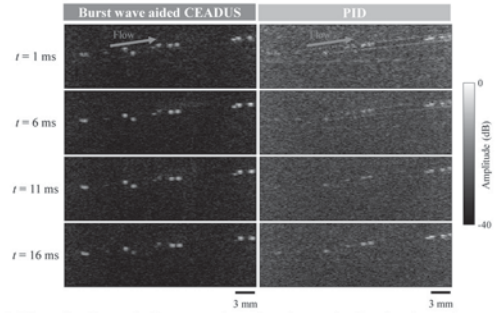


Fig.1 : Visualization of ultrasound contrast agents flowing in a channel by burst-wave-aided CEADUS and conventional PID method. t = 0 denotes the time when the transmission sequence is started.

3-2-8

3-2-8　振動型触覚センサの設計指針の検討 ―硬い対象物の測定の場合―

The study of design guidelines on vibratory tactile sensor

○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，硬い対象物を測定する場合の設計指針について実験的に検証したものである。Fig.1(a),(b)は本研究で使用する縦振動子触覚センサで，それぞれ通常型及びホーン型縦振動子を用いた構成である。Fig.2 は，縦振動子触覚センサを試験片(Aluminum(A5052) ヤング率 71GPa，Steel (SPCC)ヤング率 206GPa)に接触させた時の振動子の共振周波数の変化量を測定した結果である。ヤング率の大きい硬い試験片ほど，共振周波数の変化量は大きくなる。また，通常の縦振動子と比較し等価質量が小さいホーン型縦振動子を用いた方が共振周波数の変化量は大きくなることが実験的に明らかとなった。

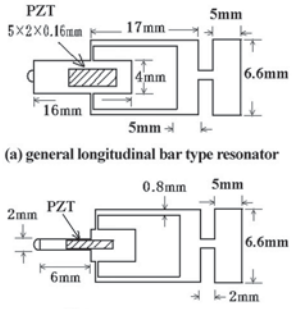


Fig.1. Construction of vibratory tactile sensor.

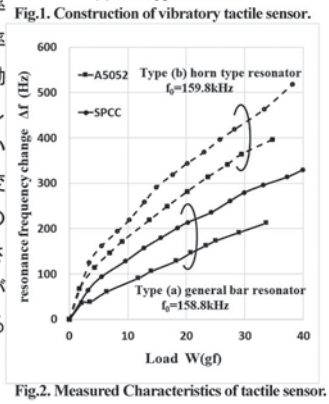


Fig.2. Measured Characteristics of tactile sensor.

3-2-9

3-2-9　超音波たわみ定在波を用いた α-酸化鉄(III)コロイドの複屈折変化

Birefringence change of α-iron oxide (III) colloids by ultrasound resonant standing waves

☆黒田悠真[1]，△江本顕雄[2]，小山大介[1]（[1]同志社大，[2]徳島大）

◆**機械的機構を必要としない，超音波振動を用いた複屈折の空間勾配**を生じさせる**新たな方法**を検討した．

◆矩形 α-酸化鉄(III)コロイドセル（Fig. 1(a)）へ**超音波たわみ定在波**を印加することで**複屈折の空間勾配**を生じさせることに成功した．

◆**複屈折の空間勾配**はたわみ定在波の振動分布に依存せず，**自己組織化**により**セル中心で極大**をとった（Fig. 1(b)）．

◆**複屈折 Δn の変化**は**超音波強度の平方根**に比例した．

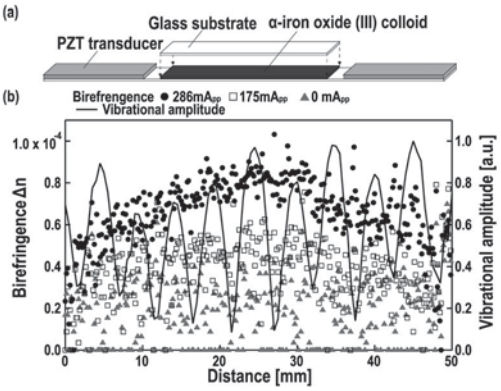


Fig.1(a) Ultrasound colloid cell and (b) distributions of birefringence and vibrational amplitude at 56 kHz.

3-2-10

3-2-10　多重エコー集約による基板付き圧電薄膜の電気機械結合係数 k_{eff}^2 の新規評価法

New method for extracting k_{eff}^2 from HBAR using multiple echo integration

☆島野耀康（早稲田大），矢田部浩平（東京農工大），柳谷隆彦（早稲田大）

◆電気機械結合係数 k_{eff}^2 とは圧電体における電気エネルギーと機械エネルギーの変換割合であり、デバイス応用において最も重要な評価指標の一つである。

◆一般的には k_{eff}^2 の測定では基板を除去する必要がある。そこで本報告では、基板付き薄膜共振子(HBAR)に対し、実際に除去せずとも信号処理を用いて基板なし自立薄膜共振子(FBAR)特性を擬似的に表現する新手法を提案する。本手法によって擬似的に表現された FBAR 特性に対して従来の共振反共振法によって k_{eff}^2 が評価できることをシミュレーション、実測共に確認した。

◆具体的には Fig. 1 に示すように、基板に漏れた弾性波のエネルギーを信号処理により、電磁結合信号と同時刻に圧電層に戻せば、FBAR 特性を擬似的に表現することができ、共振反共振法が使用可能となる。

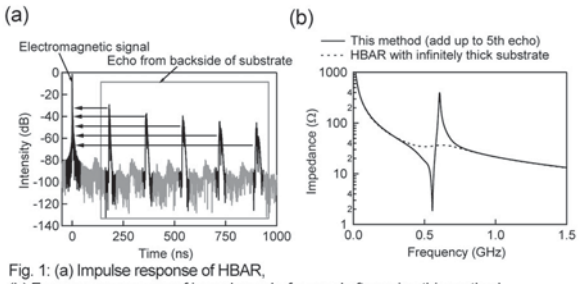


Fig. 1: (a) Impulse response of HBAR, (b) Frequency response of impedance before and after using this method

3-2-11

3-2-11　超音波ゲルレンズの光学特性評価

Evaluation of optical characteristics of the ultrasonic gel lens

☆中村光佑，原田裕生，△三木陽斗，黒田悠真，小山大介（同志社大）

◆超音波ゲルレンズは，１枚のレンズで凹凸両用の焦点変化が可能であり，従来のレンズ機構に比べて薄型化が可能である．

◆超音波ゲルレンズの光学特性改善に向けて，様々な共振周波数を用いた際の光学特性について検討を行った．

◆周波数および電圧の増加と共にゲル変形は急峻となり，焦点距離は短くなる．しかし，急峻なレンズの形状変化は球面収差の増大を引き起こす．(Fig. 1, Fig. 2)

◆今後，レンズを通してターゲットを撮影し，レンズ像からレンズの光学特性評価を行う予定である．

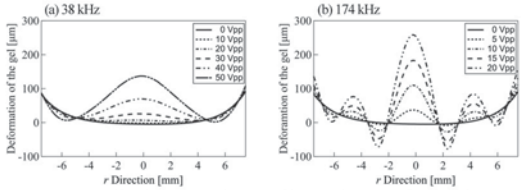


Fig. 1　Deformation of the surface of the gel.

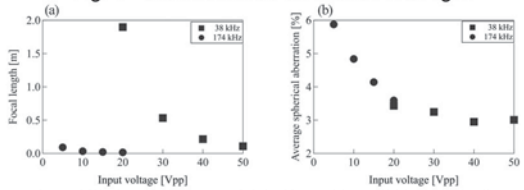


Fig. 2　(a) Focal length and (b) Average spherical aberration.

3-2-12

3-2-12　クント管内の粒子に作用する２種の音響放射力と縞の形成

The primary and secondary acoustic radiation force acting on particles and the formation of striation in a Kundt's tube

○笹川民雄(新潟南高)

◆クント管内の１個の発泡スチロール球の運動を調べ，運動の駆動力が primary acoustic radiation force であることを確認した。

◆管軸方向に一直線上に複数の球を置くと，球は音圧の節を中心として特定の間隔で管軸方向に並ぶ（Fig.1）。この現象を primary acoustic radiation force と secondary acoustic radiation force のつり合いの関係及び球の運動方程式を用いて明らかにした（Fig.2）。

◆複数の球が接近している場合，secondary acoustic radiation force（引力）の作用で付着して管軸と垂直方向に並んで列をつくる。複数の列は管軸方向に secondary acoustic radiation force（斥力）を及ぼし合い互いに遠ざかるが，primary acoustic radiation force とつり合う位置に静止して安定な縞が形成されることが分かった。

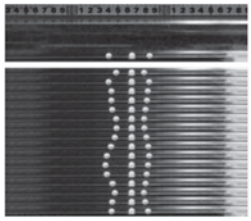


Fig.1 : The motion of three spheres. The center sphere is fixed at a pressure nod.

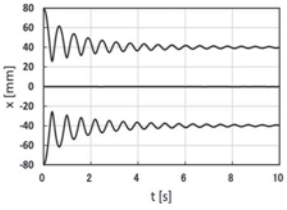


Fig.2 : The position-time graph of three spheres. (radius 3.0 mm, density 24 kg/m^3)

3-2-13

集束空中超音波ビームの放射圧による弾性表面の変形−非線形伝搬を含めての理論予測

Deformation of Elastic Surface by Radiation Pressure of Focused Airborne Ultrasonic Beams − Theoretical Prediction by Taking Account of Nonlinear Propagation of the Beams.

○鎌倉友男，酒井新一（スマートエーイー）

◆凹面開口集束音源の幾何焦点を半無限弾性表面に合わせ，超音波ビームを音響的剛なその弾性表面に垂直入射させたとき，音響放射圧によって生ずる表面の変形を，Boussinesq の変位解を用い，更に放射圧の大きさに重要な影響を与える超音波の非線形伝搬を含めて求めた。

数値計算結果は以下の通りである。

◆焦点に向かうほど顕著に発生する超音波ビームの非線形減衰は弾性面の放射圧変位に大きく影響する。

◆その影響を定量化する係数として以下の N factor を導入する。

$$N_f = \frac{u_{z,m}(0)}{u_z(0)}$$

$u_{z,m}(0)$：非線形伝搬を考慮しての変位，$u_z(0)$：線形伝搬のときの変位

◆一般に集束音源の半開口角や焦点距離などの条件で N_f の大きさは変化するが，特に周波数の影響は大きい。

◆音響放射パワーの増加で，一時的に線形理論で予測される変位よりも僅かに増すこともあるが，多くの場合 $N_f < 1$ であって，線形予測値よりも下回る。

3-2-14

3-2-14　近距離場音波浮揚する小径穴付き円柱周辺の音響流の流速分布観察

Observation of velocity distribution of acoustic streaming around a cylinder with small holes in near-field acoustic levitation

☆唐木澤直樹（室蘭工大･院），青野浩平（精電舎電子工業），
青柳学(室蘭工大･院)

◆振動面(φ10mm)と平面の間で浮揚する小径穴付き円柱(φ10mm)の周囲に発生する音響流の流速（流線）分布を観察した。

◆Fig.1 のようにアクリル容器内をトレーサー粒子煙で充満させ，穴付き円柱の横穴にシート光を照射し，高速度カメラで撮影した。

◆Fig. 2 は穴付き円柱周辺の粒子挙動を連写速度 5,000 fps で撮影した結果である。領域(A)では振動面と円柱のギャップ間に空気は流入せず，領域(B)および領域(C)では音響流の発生が確認できた。

◆Fig. 3 は横穴周辺の噴出流の粒子挙動を連写速度 10,000 fps で撮影した結果である。横穴からの噴出流は周囲の空気を引き込んで発生することが分かった。

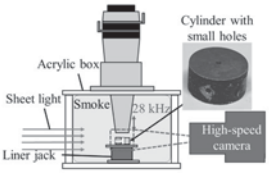


Fig. 1 Recording method.

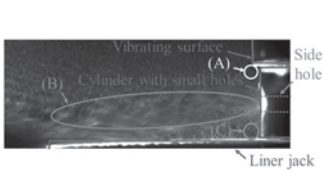


Fig. 2 Particle behavior around cylinder with small holes (Continuous shooting speed: 5,000 fps)

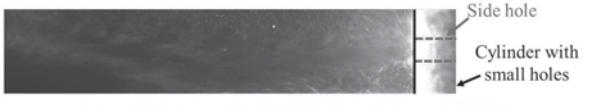


Fig. 3 Particle behavior around the side hole (Continuous shooting speed: 10,000 fps)

3-4-1

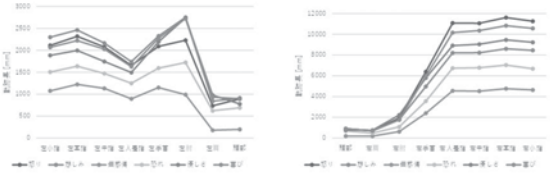
3-4-1　モーションキャプチャを用いた感情表現による中国琵琶の演奏動作の解析*

* Motion Analysis of Chinese Pipa Performance by Emotional Expression Using Motion Capture

○三戸勇気，田葱（日大･芸術），山口達也（日大･理工），
駒澤大介，川上央（日大･芸術），篠田之孝（日大･理工）

筆者らは，モーションキャプチャシステムを用いた鍵盤楽器の演奏動作の計測ならびに解析を行ってきた[1]-[4]．東洋楽器と西洋楽器における感情表現による演奏動作の解析を目的に，東洋の撥弦楽器である中国琵琶の演奏動作の計測を行った．今回，中国琵琶の演奏動作における身体各部の関節部位などの軌跡長を算出し，中国琵琶の感情表現における演奏動作の解析の検討を行った．

Fig.1 の(a), (b) は身体の左側, 右側の関節部位等の鉛直方向の軌跡長の結果である．



(a) Left side of the body　　(b)Right side of the body

Fig.1 Trajectory length in Z-axis

3-4-2

3-4-2　クロスグリップでのマリンバ演奏時における腕のしなりと音圧レベルの関係

Relationship between arm flexion and sound pressure level when playing the marimba with cross grip

☆竹下和秀（国立音楽大学大学院），三浦雅展（国立音楽大学）

◆マリンバでの４本マレット奏法において，腕の連動は重要視されるポイントの一つである。しかし、腕の連動について触れられている先行研究は存在しておらず，指導者や演奏家の感覚的な指導から抜け出せていない。

◆本研究では，日本で広く普及しているトラディショナルグリップとバートングリップの演奏時おける肘の角度が最大角になる時刻と打叩動作の手首の角度が最大角になる時刻との時間差分と当該打叩動作に対応する音圧レベルとの関係性について異なるのかを比較検討を行なった。

◆トラディショナルグリップ及びバートングリップの打叩動作における肘の角度が最大角になる時刻と手首の角度が最大角になる時刻の時間差分と当該打叩動作に対応する音圧レベルの分布を Fig.1 に示す。Fig.1 から，トラディショナルグリップの奏者は時間差がバートングリップに比べ小さく，腕の振りとしてはしなるような動きが少なく，かつダイナミックレンジの広い音圧レベルが確認された。一方でバートングリップでは，腕のしなりが大きいものの，ダイナミックレンジが狭い様子が確認された。

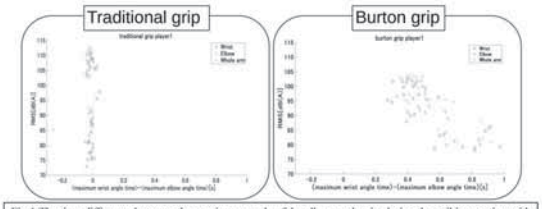


Fig.1 The time difference between the maximum angle of the elbow and wrist during the striking motion with traditional grip and Burton grip, and sound pressure levels

3-4-3

3-4-3　スネアドラム演奏におけるロール奏法の動作と演奏音の関係

A relation between motion and sound on playing the roll performance using the snare drum

○三浦 雅展（国立音楽大学）

◆スネアドラムによるクローズドロール演奏を対象とし，スティックの動きと音の関係について述べている。

◆ロール演奏におけるドラムスティックの先の動きをモーションキャプチャによって記録し，動作の記録を行ない，録音されたロール音の音響信号と動作の対応関係を調査している。

◆どのロール音においても，振幅包絡のパワースペクトルにおいて，おおよそ 40Hz にピークを持つことが明らかになった。

◆左右手のバウンスから得られる変動スペクトルは 16-26Hz であるため，左右手のバウンス以外の要因（スナッピー等）が示唆される。

◆なお，左右手のストローク間隔の変化，バウンド数の変化と音響信号に特に対応関係は見られず，一貫して 40Hz 付近に見られる点が興味深い結果であった。

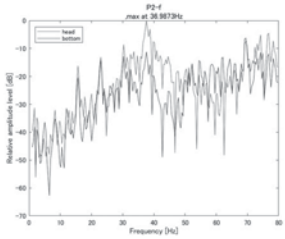


Fig.1 Fluctuation Spectrum on playing *p* by P2

3-4-4

3-4-4　鏡面反射を用いる膜振動の簡易計測に基づく膜鳴楽器のチューニング支援

Tuning support for membranophones based on simple measurement of membrane vibration using specular reflection

☆井上茜（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）

◆背景

膜鳴楽器のチューニングは微妙な音程の聞き分けなどを要するため困難である．したがって，チューニングを支援するシステムが必要であり，そのためには膜の振動を計測することが有効である．

◆目的

膜全体の振動を計測可能とする簡便な計測手法を提案する．本手法では，膜鳴楽器の膜面が鏡面反射することを用いる．本稿ではティンパニの (1,1) モードの振動モード形状を可視化することに取り組んだ．

◆結果

鏡面反射を用いて Fig. 1 の方法で膜振動計測を行い，Fig. 2 のように (1,1)モードの形状を可視化できた．節の場所を参考にすることで，モード形状の非対称を直すことができると考えられる．

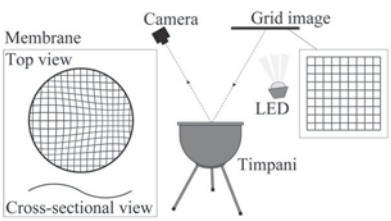


Fig.1 : Schematic diagram of arrangement.

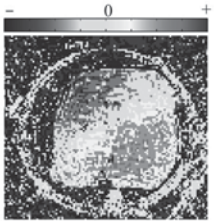


Fig.2 : Visualized shape.

3-4-5

3-4-5　ティンパニ演奏における打点とバチの制御による音響的効果および音楽的効果*

Acoustic and musical effects of controlling the beating point and mallets in timpani performance

○深町浩司（愛知県立芸術大学）

◆ ティンパニは**「ノイズと純粋なピッチという、相反する領域を自在に行き来するユニークな太鼓」**である。

◆ 15 世紀以降の西欧の戦場や宮廷などで鳴らされた**強い打音を現代のティンパニ演奏で活用**して**多様な音色を創出**する。

◆ 現代のティンパニメソードの主流は打面の半径手前 1/3～1/4 あたりを「良い音が出る打点」だとしている。本発表は打面の半径手前 1/10～1/2 を打点として使い、**打音（ノイズ）を積極的に用いる奏法による音色変化の効果**を示す（Fig.1）。また、**バチを持つ指の動作による微細な音色変化**の効果も示す。

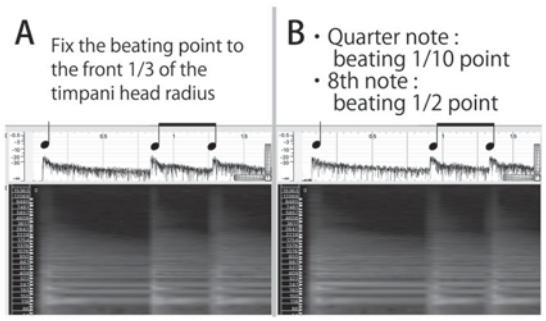


Fig.1 : Change the beating point to express the difference between notes by changing the attenuation and timbre of the sound.

3-4-6

3-4-6　歌唱指導による歌声の音響的特徴と体内運動の変化の検討

A study of changes in acoustic characteristics and body movements of singing voice by singing instruction

◎高橋純(大阪芸大), 志々目樹, 戸田菜月, 竹本浩典(千葉工大)

◆歌唱中の体内運度は外部から観察できないため，指導者は学習者の歌声から発声器官の状態を判定し，体内運動に対するイメージや感覚を示す用語を用いて指導している。

◆本研究では，声楽未経験の学習者が１年間の歌唱指導を経て，歌声の音響的特徴と歌唱中の体内運動がどのように変化したかを検討した。

◆その結果，訓練後は口の開きが大きくなり，口腔と咽頭が拡大し，喉頭が下降した。また，歌い手のフォルマントが含まれるスペクトルの2−4 kHz の帯域におけるパワーが相対的に上昇した。

◆また，指導者は歌声から喉頭の上下動などの学習者の体内運動を推定できることも示された。

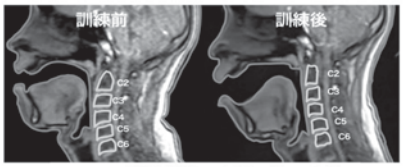


Fig.1: 訓練前後の声道形状の比較

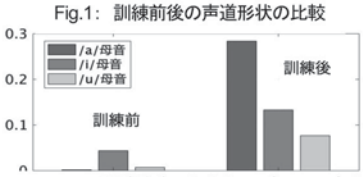


Fig.2: 訓練前後の2-4kHzのパワーの変化

3-4-7

3-4-7　リアルタイム MRI を用いたオペラ歌唱における声量を調節した際の身体制御

Examination of the body control for changing voice volume in operatic singing using real-time MRI

◎戸田菜月, 竹本浩典(千葉工大), 高橋純(大阪芸大)

◆オペラ歌手は，豊かな響きを保って声量を調節することが可能である。

◆声量は声門下圧や声帯の状態，声道の形状によって変化するが，歌唱中に肺や声道形状をどのように制御しているのか明らかではない。

◆本研究では，テノールとバリトン各１名が声量を調節して歌唱した際の胸腹部と頭頸部をリアルタイム MRI で撮像し，肺圧や声道形状の変化が声量とどのような関係があるのか検討した。

◆歌唱中の声門面積が一定として推定した肺圧の変化パタンは，デクレッシェンドでは声量の変化パタンと一致していたが，クレッシェンドでは一致しなかった。よって，必ずしも肺圧で声量を制御しているとは言えなかった（Fig. 1）。

◆声道形状の主成分分析の結果，喉頭の上下動が声量に関係する制御であった。これは，この制御が披裂軟骨の内転度合いを変化させることで声量を調節している可能性があることを示す。

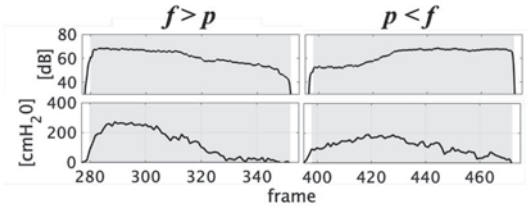


Fig.1 : Vocal intensity (top) and estimated lung pressure (bottom) during singing for a baritone

3-4-8

3-4-8　クラリネット用リードに対する演奏者による評価を再現する機械学習モデルの検討

A machine learning model to replicate evaluations by performers on clarinet reeds

☆鈴木真琴(国立音大大学院), 三浦雅展(国立音大)

◆クラリネットの演奏には葦材のリードを伝統的に用いる。葦材は収穫後様々な加工が施されクラリネット専用のリードとなるが，入手前に試奏することはできない。要因として，葦材は一度水や唾液につけると劣化が始まってしまうこと，消毒を行なうと演奏感が変化してしまうことなどが挙げられる。リードは一般に 10 枚一組で販売されているものの，演奏適性にはばらつきがあり，現状奏者は購入したリードを一度演奏して使用可能か判断することを余儀なくされている。

◆本報告ではリードの外観を撮影し CNN に入力することで，演奏することなく奏者による評価を再現すること目的とする。本調査の検証に用いたリードの一部を図 1 に示す。

◆まずリードを表，左，右，裏の四方向から撮影した画像をそれぞれ入力した。その結果，右方向から撮影した画像を用いた学習に対し，いずれも正解率に有意差が認められた。

◆次に，リードの枚数を増やして表方向から撮影した画像を用いて学習を行なった結果を目視による分類と比較した。機械学習の評価指標をそれぞれ求め，目視においても同様に求めた。結果を図 2 に示す。その結果，目視と CNN は同程度の判断ができることが示唆された。

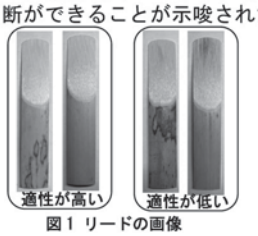


図 1 リードの画像

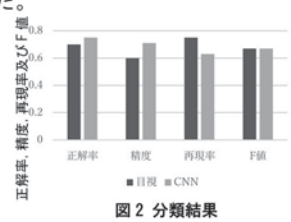


図 2 分類結果

3-4-9

3-4-9　演奏中の擦弦パラメータ計測及びその時刻歴提示

Measurement of bowing parameters and presentation of time histories during performance

◎猪崎遼大(筑波大・院・シス情工), 若槻尚斗, 水谷孝一, 善甫啓一, 海老原格(筑波大・シス情系)

◆楽器練習は演奏動作などのフィードバックを用いることが重要

◆楽器に対して与えている制御パラメータの計測を応用することで演奏練習の効率の向上が期待できる.

◆擦弦楽器において，弓圧や擦弦位置などの擦弦パラメータを計測する試みがあるが，演奏練習の観点では最適ではない.

◆チェロの響板に設置されたカメラモジュール及び，モーションキャプチャを用いて擦弦パラメータの計測を行った.

◆計測環境や解析手法などにおいて，改善の余地がみられた.

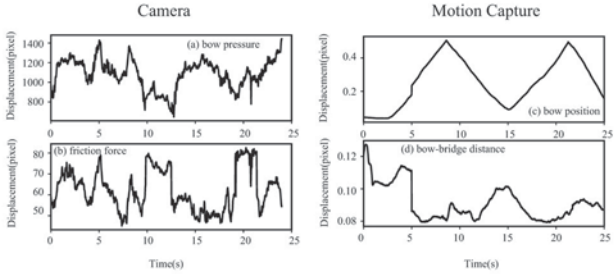


Fig.1: Measured bowing parameter

3-4-10

講演取消

3-4-11

3-4-11　楽器開発の現場における技術と演奏家の接点

Relation between technology and music players in the scene of developing musical instruments.

〇森隆志（ヤマハ株式会社）

◆楽器の開発において暗黙知の伝承が困難になり，科学的なアプローチの重要性が増しているが，一方で，演奏家の評価も依然重要である。

◆演奏家の評価は「音を聞いて感じる」というような一方通行の行為ではなく，聞く・感じるといった感覚を元に運動を通じて楽器を操作し，またその音を聞くというフィードバックループのような構造である。

◆楽器開発においては演奏者の評価などを元に重要な要素を抽出し，効果的な近似手法を見つけ出すということが求められる場合がある。

◆楽器や人の科学と共に“演奏”という行為が解明されることで，楽器や音楽に新しい可能性が見えてくると考えている。

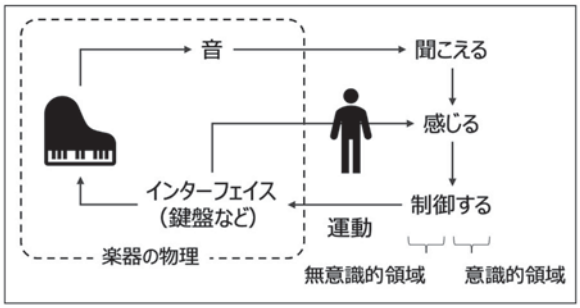


Fig.1：Relation between the instrument and the player during the performance

3-4-12

3-4-12
信号包絡線を用いた打撃音の客観評価指標の提案とタイムストレッチング性能の評価

Signal-envelope-based objective index for evaluation of time stretching methods

◎赤石夏輝，山田宏樹，矢田部浩平（農工大）

タイムストレッチング手法の客観評価
► 真の信号が存在しないため難しい

従来の位相の整合性に基づく客観評価
► 打撃音に対する性能を適切に評価できない

元信号とストレッチ信号の概形を包絡線で抽出して包絡線の概形を比べる評価手法を提案

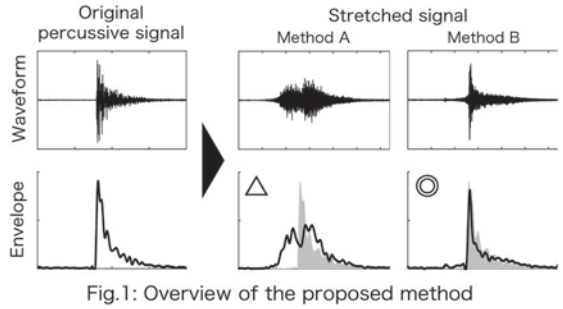


Fig.1: Overview of the proposed method

3-4-13

3-4-13　ゲームの映像から BGM を予測するための機械学習モデル

Machine Learning Models to Predict Background Music from Game Videos

☆林龍星，北原鉄朗（日大）

◆はじめに：ゲームのシーンと BGM には強い関係があると考えられる．そこで，ゲームのシーン（映像）から BGM を検索するシステムを提案する．

◆提案手法：ユーザはゲームの映像を入力する．システムは CNN で映像をフレームごとに次元削減する．さらに，LSTM で次元削減した映像を音響特徴に変換する．そして，予測した音響特徴に類似した音響特徴を持つ BGM を検索する．

◆評価実験

- 『Undertale』と『クロノ・トリガー』を学習したモデルにそれぞれ『OMORI』と『ロマンシング サ・ガ3』を入力する．
- 予測した音響特徴と実際の音響特徴を比較した．『Undertale』を学習したモデルは『OMORI』の戦闘シーンの音響特徴を予測できた．『クロノ・トリガー』を学習したモデル『ロマンシング サ・ガ3』の探索シーンの実際の音響特徴を予測できた．
- 異なるシーンを入力すると，異なる BGM を検索するか確認した．『Undertale』を学習したモデルに異なるシーンを入力すると，同じ BGM を２回検索した．『クロノ・トリガー』を学習したモデルに異なるシーンを入力すると，同じ BGM を１回検索した．

◆おわりに：本研究はゲームの映像と BGM に強い関係があることを前提としている．この前提について妥当性を検証する必要がある．

3-4-14

3-4-14　小節間の類似度に基づくポピュラー音楽音源からのリフ抽出

Riff extraction from popular music sources based on measure-to-measure similarity

☆宮武碧，下倉良太，△飯國洋二（大阪大）

◆リフ（refrain）は曲の前奏や間奏等で演奏される印象的なフレーズで，曲が登場した年代の曲調や雰囲気を表す特徴量と言える．

◆ギタータブ譜からリフを抽出する先行研究があるが，タブ譜が手に入らない曲やリフがギターで表されない曲では抽出できない．そこで，本研究では**ポピュラー音楽音源から**リフ抽出を行う．

◆小節（音楽を区切る単位）ごとのスペクトログラムの類似度を定義することにより，相関行列を得る．例えば，リフの繰り返し単位の長さが２小節の場合，リフで相関行列にチェック柄が現れる．（Fig.1）

◆チェック柄のフィルタ（Fig.2）を相関行列に畳み込み，最大値をとるインデックスよりリフが何小節目から始まるかを特定できる．

◆今後は，チェック柄に限定しない，相関行列からリフを取り出すフィルタの設計が必要である．

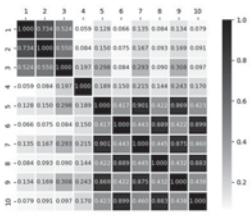


Fig.1：Correlation matrix (ray / BUMP OF CHICKEN)

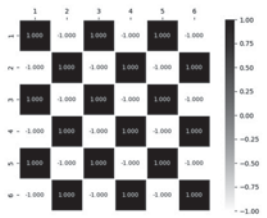


Fig.2：Riff extract filter

3-4-15

3-4-15 クロマグラムと CNN を用いたギター音源からベース音源の生成の試み

Attempt to generate bass sound source from guitar sound source using chromagram and CNN.

☆香西智雄，北原鉄朗（日本大学大学院総合基礎科学研究科）

◆目的:ギター音源からベース音源の生成。

◆提案手法:CNN を用いて、音源のクロマグラムから学習させる。

◆データセット: Cakewalk by BandLab を用いてギターおよびベースパートの自分で作成(20 曲)。BPM は 120，小節数は 4 小節 (8 秒) 固定。

◆学習データ: 上記のデータセットのうち 10 曲。

◆テストデータ:上記のデータセットのうち学習データ以外の 10 曲に加え、The Beatles の 3 曲 (『ヘルプ!』『恋を抱きしめよう』『デイ・トリッパー』)

◆結果: 一部についてはコード進行に沿ったベース音が生成された。

◆今後は実音源を学習データとして進めていきたい。

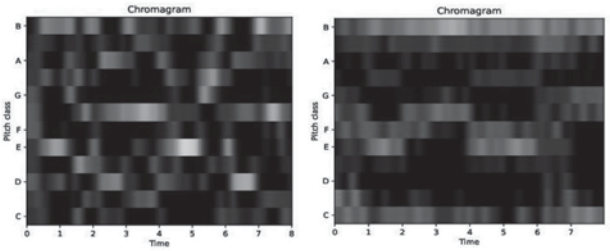


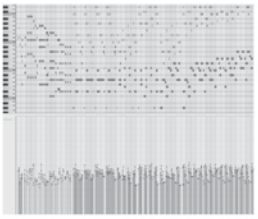
Fig1: Chromagrams of the input guitar sound (Left) and the generated bass sound (Right) of "Day Tripper"

3-4-16

3-4-16　Piano Velocity Prediction Using a Seq2Seq Model with Attention Mechanism

☆Taein Kim (Inha University), Yunho Kim (Andong National University)

Digital music scores can offer precise time and velocity information for each note but setting each velocity value by hand can be time-consuming or require high piano skills. In this paper, we propose a deep learning model to predict the velocity of piano notes using the Seq2seq model with Luong Attention and the MAESTRO dataset. The proposed model consists of a loss function including cosine similarity that treats note elements (TimeDiff, NoteNumber, Length) as vectors with a single direction toward velocity. The qualitative metric indicated that the model is not yet strongly preferred over human performance. In the quantitative assessment, the proposed model presented a better error rate than other previous studies but achieving a competitive F1 score is remained as further research.



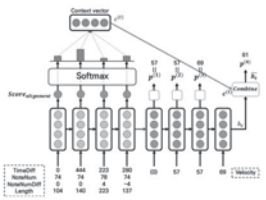


Fig.1 (Left)：Velocity prediction results of H4-att model
Fig.2 (Right): Diagram of model architecture

3-4-17

3-4-17　禁則違反回数の少ない和声自動生成のための禁則違反予測モデル

Error prediction model for harmony generation with fewer parallel errors

○唐津裕貴，深山覚（産総研）

◆目的：深層学習モデル DeepBach による和声生成結果の品質向上

◆課題：和声学の初歩的な禁則違反回数が多い

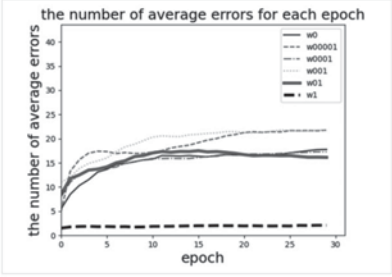
- 初歩的な禁則：声部間の音程の連続，並達など
- 禁則違反に対するペナルティがない

◆方法：学習済みの禁則違反予測モデルつき DeepBach を検討

- 損失関数 L＝L_1＋w＊L_2
- L_1：DeepBach による損失
- L_2：禁則違反予測モデルによる損失（禁則違反のペナルティ）
- w＝0, 0.0001, 0.001, 0.01, 0.1, 1 でそれぞれ性能比較

◆結果：w＝1 のモデルで禁則違反回数の減少を確認

- 禁則に関連しない面での和声，対位法的な処理に課題が残った



各生成エポック時点でのそれまでの生成エポックを通じた禁則違反回数の平均値

3-4-18

3-4-18　独自非和声音データセットを用いたニューラル言語モデルに基づく旋律分析

Melody analysis based on neural language models using the classical non-chord tone dataset

☆髙橋拓椰(電通大)，金子仁美(藝大)，
嵯峨山茂樹(東大/電通大)，中鹿亘(電通大)

本研究では、楽譜、和声情報から楽譜内の各音符の非和声音分類を含む旋律分析ためのニューラル言語モデルを提案する。提案モデルは、我々がこれまで公開してきた楽譜、和声、非和声音ラベルを含む独自の非和声音データセットによって学習される。さらに、精度向上のため、データセット内のサンプル数の不均衡性、サンプル数不足などの問題に有効な学習戦略についても比較検討した。このようなモデルは、人手で行われていた非和声音データセット作成の作業を支援したり、音楽教育における和声学学習の補助などへの応用が期待される。

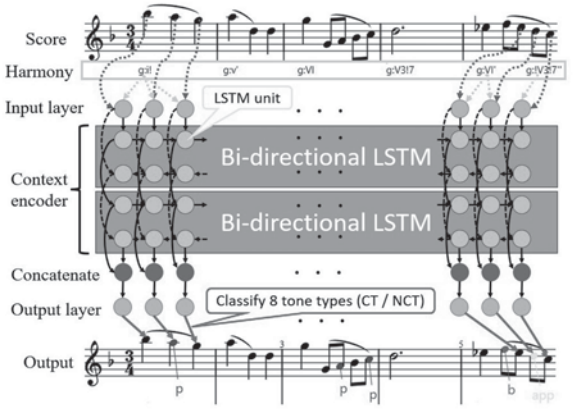


Fig. 1. Proposed model for identifying tone types based on score and harmony information.

3-4-19

3-4-19　盛り上がり度に基づくループシーケンサ：Web アプリ化のためのフロントエンドの実装

Loop sequencer based on the degree of excitement: Implementation of front-end for web application

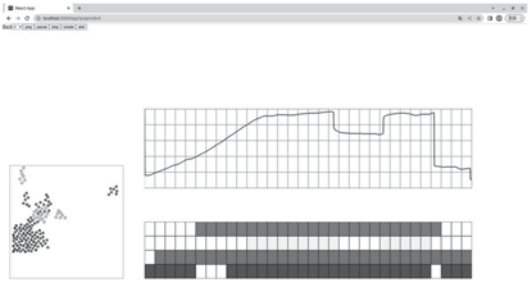
☆岩本祐輝(日大院･総基科研)，尾上洋介，北原鉄朗(日大･文理)

◆我々は以前から，ユーザが描いた盛り上がり度曲線に基づいて音素材を自動選択するループシーケンサの開発を進めている．しかし，スタンドアローンアプリであるため，使用するにはインストールが必要であり，自動的なログ収集も容易ではなかった．

◆そこで，本システムの Web アプリ化を進めている．前報では，サーバサイドでの API の設計・実装について述べた．本稿では，フロントエンド側の設計・実装について述べる．具体的には，React を用いて Web ブラウザ内で実行可能な GUI を作成し，Redux Toolkit を使って，UI コンポーネントのイベント処理および状態更新を実現する．

◆ローカル上で，現在実装されているものは，いずれも正常に動作した．

◆一般公開可能な Web アプリの実現に向けて，認証機能，データベース，ユーザの操作ログの収集機能の実装に取り組む．



3-5-1

3-5-1　EV 加速を想定した聴覚の臨界帯域に基づく非定常マスキング特性

Unsteady masking characteristics based on the auditory critical band assuming EV acceleration

☆柴橋和真(中央大院)，△金堂雅彦，戸井武司(中央大)

◆本研究は，電気自動車(以下，EV)の車室内で知覚されるモータ回転次数やギア噛み合い次数による定速時の定常純音，および加速時の非定常純音すなわち周波数掃引音において，暗騒音からの可聴突出量 L_P に基づくマスキングラインを，人間の可聴周波数全域で実験的に明らかにした。

◆さらに，聴覚の臨界帯域に基づき，周波数軸から Bark 軸に変換した時の純音移動速度として定義する臨界帯域通過速度 v_{CBpass} [Bark/s]という新たな観点から，周波数掃引音に対する聴覚メカニズムを解明した。

◆聴覚メカニズムの異なる領域別に，決定係数が 0.95 以上の高精度な近似式を採用し，Fig. 1 に示す(a),(b),(c),(d)の目的変数の順に任意の掃引速度をモデル式に代入することで，100~10000 Hz の周波数領域において，周波数毎の可聴突出量 L_P の予測値が一意に求まる。これより，EV 定速時および加速時に想定される全ての周波数掃引速度に対応した非定常マスキング特性の定式化が実現した。

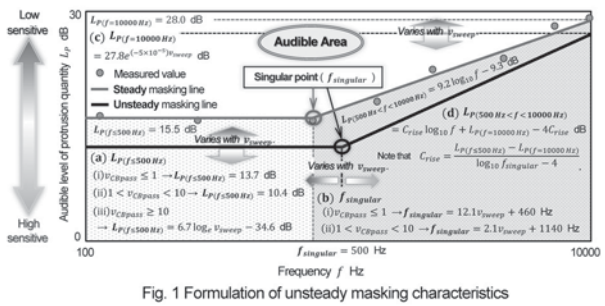


Fig. 1 Formulation of unsteady masking characteristics

3-5-2

3-5-2　EV 走行音に対する順応と選択的注意

—第2報　周波数帯域の検討—

Adaptation and selective attention to EV vehicle sound

☆亀山勇希，石光俊介(広島市大)，△古高啓介，△藤井泰人(東洋紡)

◆本研究では，順応及び選択的注意に着目し，脳波の事象関連電位(ERP)から走行騒音に対する注意機能を定量することを目的とした．第１報では，前段階として，走行音の音圧レベルの変動に対する ERP の影響について調査した．本稿では，運転中に注意の向きやすい走行音の周波数特性を明らかにすることを目的とした．

◆実験では，走行音の音圧レベルの変動に対するトップダウン型注意に着目し，３つの周波数帯域の走行音成分を変動させ，それに対する ERP の変化を観測した．

◆その結果，走行音の低周波成分より高周波成分(1 kHz 付近)の方がより注意機能が働きやすいことが分かった．

◆今後の課題としては，1 kHz 以上の帯域を追加することによって，高周波帯域での傾向を調査することが挙げられる．

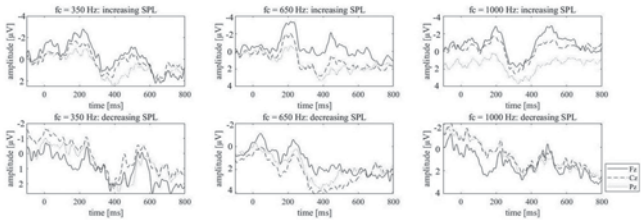


Fig.1 : Results of experiment. Grand-average ERP waveforms with increasing and decreasing SPL in the frequency bands at center frequencies of 350, 650, and 1000 Hz. The P3 component appears early at a center frequency of 1000 Hz.

3-5-3

3-5-3　自動運転中の車内環境におけるBGMのシャープネスの変化が反応時間と主観的印象に及ぼす影響

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

☆尾崎真弘，星野博之（愛知工業大学）

◆目的　本研究では、自動運転中の車室内を模した環境下で一定間隔ごとにシャープネス値を変化させた音楽を聴取しながら視覚反応タスクを行い、ドライバを快適かつ覚醒度の高い状態に保つクラシック音楽を明らかにする。

◆方法　防音室内において高速道路走行中の映像と、ピアノ音の原曲またはその原曲から高周波成分を 10, 20, 30 秒ごとに強調・減衰した音データ 7 個を 10 分間流し、反応時間のタスク評価と、「不快-快適」などの SD 法による印象評価を 11 項目に対して行う。

◆結果　全体 10 分間の反応時間（被検者 12 名の平均）を図 1 に示す。また、No.4 の音の経過時間と反応時間の推移を図 2 に示す。高周波成分を 20 秒ごとに減衰させた音に対しては覚醒維持効果の減衰が大きく見られた。

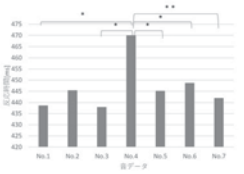


図 1：全体 10 分間の反応時間

(No.7：原曲、No.1/2：10秒間隔強調 / 減衰
No.3/4：20秒間隔強調 / 減衰、No.5/6：30秒間隔強調 / 減衰)

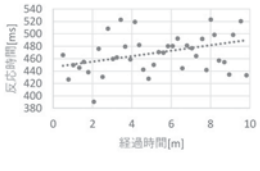


図 2：No.4 の音の経過時間と反応時間の推移

3-5-4

3-5-4　変調感への影響を考慮した快適な協和音付加条件の考案

Consideration of comfortable consonance addition condition based on the effect on modulation feeling

◎広沢隼土（中央大院），△和久井康平，
松田道昭（荏原製作所），戸井武司（中央大）

◆機械動作音が持つ振幅変調する純音成分(以下, 変調純音)に対する付加音の条件が変調感, 快適感に与える影響を明確にし, 快適な協和音付加条件を考案する。

◆変調周波数 70 Hz の高速変調純音，4 Hz の低速変調純音に対して音付加前後で相対評価を行った。その結果，変調感は付加音の音量による影響が大きく，Fig. 1 に示すように音付加後のラウドネスが 11 sone である付加条件 Def よりも大きな音量で付加することで，振幅変調を認知しやすい低速変調純音に対しても変調感が改善する。また，付加音の周波数比が協和音ならばドとソのような 2 音の組み合わせでも快適感が向上する。

◆2000 Hz の変調純音に対して検証実験を行った。高速変調純音には音量が Def から付加音のラウドネスレベルを 5 phon ずつ増加させ，周波数比が不協和音(DisH)，協和音(ConH)として付加した。低速変調純音には音量が Def，周波数比が高速変調純音と同様に不協和音(DisL)，協和音(ConL)として付加した。その結果，Fig. 2 に示すように ConH，ConL で快適感が向上し，選定した付加条件における有効性が確認できた。

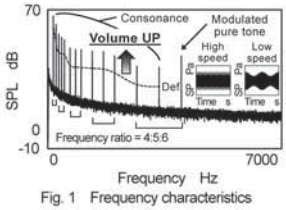


Fig. 1　Frequency characteristics

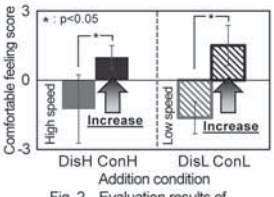


Fig. 2　Evaluation results of verification experiment

3-5-5

3-5-5　高齢者見守り支援のための認知機能に基づくサウンドデザイン

Sound design for living support for the elderly based on human cognitive functions

○有光哲彦，テオリン・アクセル（フィート），岡本怜奈（IC）

◆要介護者の自立および介護者のゆとり創出を目的に，在宅高齢者の介護者に対し見守り支援のためのヒアリング調査を実施し，介護の課題を顕在化した。顕在化した要介護者の中核症状や周辺症状に応じて，音により人の心理・生理へ作用する支援の概要を Fig.1 に示す。

◆人には同調機構が備わり，体内時計は日照と食事の刺激で同調する。時間感覚の喪失は生活の乱れが一要因と考え，網膜刺激および内蔵刺激を考慮して自律神経系へ作用させる旋律による同調支援音をデザイン検討した。

◆旋律の音色は「BPM」「和音」「周波数ピーク」「調子」および「明度」として制作デザイン要件を定義し，8 種（①起床→②朝食→③生活行動→④昼食→⑤身体活動→⑥夕食→⑦風呂→⑧就寝）の 5 s 程度の音を制作した。

◆明度，主旋律 BPM，および調子の関係を Fig.2 に示す。明度は④昼食時を最高，⑧就寝時を最低にして循環させた。主旋律 BPM は①起床時から上昇，④昼食時で最高，⑧就寝時へ下降した。調子は午前を上り調子，午後を下り調子とし，主旋律の終盤に前後周波数の比で表現された。相関係数は明度と BPM では 0.96，明度と調子とでは 0.80 と高い傾向が見られた。

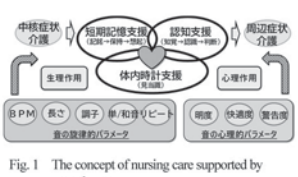
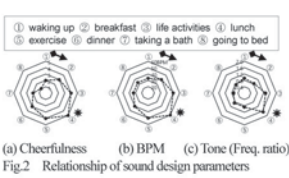


Fig. 1　The concept of nursing care supported by sound cues



(a) Cheerfulness　(b) BPM　(c) Tone (Freq. ratio)

Fig.2　Relationship of sound design parameters

3-5-6

3-5-6　印象評価および生理指標に基づく体内時計を考慮した生活機能音の評価

Evaluation of everyday life functioning sound-cues in consideration of the body clock based on impression evaluation and physiological index

☆河原怜雄（中央大院），有光哲彦，テオリン・アクセル（フィート），
岡本怜奈（IC），戸井武司（中央大）

◆本論文では，体内時計を考慮して制作された 8 種類の生活機能音（①起床，②朝食，③生活行動，④昼食，⑤身体活動，⑥夕食，⑦風呂，⑧就寝）を印象評価と生理指標の両面より評価し，時間感覚の支援に対し有用な音デザインパラメータを推定する。

◆網膜刺激，活動刺激，内臓刺激を考慮した形容詞対の 7 段階絶対評価を行い，因子分析を Table 1，覚醒因子得点を Fig. 1 に示す。覚醒因子より，網膜や活動を刺激することで，体内時計と実時計が同調できると示唆された。

◆心電情報によるローレンツプロットに基づく覚醒度合いを Fig. 1 に示す。印象評価と生理指標で対応が見られ，評価音が「朝・昼」，「夜」の時間帯の活動を促すグループに分類できた。

◆評価音を聞いた際の時間帯の印象(朝，昼，夜)で分類し，「なぜそう感じたか」を記述してもらうことで時間帯の印象を定量化した。

◆重回帰分析により，時間帯の分類に対し有用な音デザインパラメータが，BPM，Tonality，Loudness，低周波数帯域の音の調子であると推定できた。

Table 1　Factor analysis

	(+) 形容詞対 (−)	因子 覚醒	活性	平穏
●	明るい−暗い	0.79	0.39	−0.04
▲	軽快な−荘重な	0.77	0.08	−0.18
▲	陽気な−陰気な	0.74	0.40	−0.10
●	すがすがしい−うっとうしい	0.74	0.12	0.17
	…	…	…	…
■	潤った−乾いた	0.24	0.70	0.31
■	温かい−冷たい	0.17	0.69	−0.13
	…	…	…	…
■	いらだつ−心休まる	−0.00	0.03	0.72
	寄与率	37 %	16 %	10 %
	累積寄与率	37 %	53 %	63 %

●網膜刺激 ▲活動刺激 ■内臓刺激

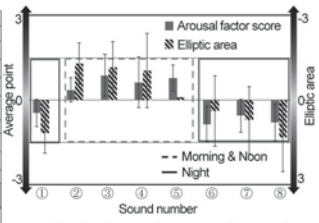


Fig. 1　Average factor scores for arousal factors and elliptic area

3-5-7

3-5-7　サイン音のデザインプロセスに在る「知」の考察－プロと初心者の比較分析－

Consideration of Knowledge Presented in Design Process of Auditory Signs
－Comparative Analysis between Professionals and Beginners－

◎村木亜未香（小野測器），山内勝也（九州大・芸工）

◆背景

新たにサイン音をデザインする際に，具体的にどのようにして音を作れば良いのかという知見が少ない

◆目的

サイン音デザインの場面に存在する「知」について考察を行う

◆方法

プロのデザイナーと初心者のデザインプロセスを比較した

- デザインのお題：九州大学大橋キャンパス内で用いることができる「芸工」独自の電子決済サービスの決済音
- 作業の様子を映像で記録，作業中の思考についてインタビュー
- デザインフローを図にまとめ，各デザインプロセスを分析・比較

◆結果

- プロと初心者の共通点より

  「決済音っぽさ」と具体的な音の特徴を結びつけていた点が共通していたことから，日常的な経験がデザインを行う上での「知」に結びつき得ることが示唆された

- プロと初心者の相違点より

  作業とその結果として生まれる音のイメージ・試行錯誤を行う際の引き出し・言語や視覚情報を音の特性に変換する術など，プロの方がより多く保持している「知」の存在が示唆された

3-5-8

風鈴の音色に関する研究
各種風鈴の音響特性と心理的評価の関係について

A study on the sound of wind chimes; the relationship between acoustic characteristics and psychological evaluation

○土田義郎（金沢工大）

◆風鈴の単発の音について音響分析と心理的評価を行った。

○減衰時間（60dB 減衰時間）：金属が長め

○ラフネスとうなり：ラフネスはうなりの影響を受ける。

○スペクトル重心と高調波比：形状の影響が大。周波数特性の複雑なものは高調波比が大きい。

○心理的評価との関係：対数減衰時間は評価性・明瞭性・複雑性の 3 項目と関連。減衰時間の長い物が好まれやすく、明瞭かつ複雑に聞こえる。スペクトル重心には対数減衰時間・高調波比とともに複雑性との相関がある。高調波比は明瞭性・繊細性・複雑性との相関がある。

seq.	名称	基本周波数(Hz)	スペクトル重心(Hz)	高調波比	減衰時間(60s/dB)	ラフネス	うなり(Hz)
1	江戸風鈴　小丸	2209	6635	3.00	0.65	0.036	22.9
2	江戸風鈴　大丸	2151	4684	2.18	0.49	0.061	12.6
3	江戸風鈴　ひょうたん	2808	5591	1.99	0.26	0.030	14.4
4	風鈴（作家作）	2914	3412	1.17	0.73	0.053	-
5	加賀風鈴	2563	8197	3.20	0.67	0.031	7.0
6	クリスタルガラス風鈴	1627	4889	3.00	3.33	0.061	5.7
7	南部風鈴	2687	3475	1.29	6.51	0.010	16.6
8	呼び鈴型風鈴	1084	7731	7.13	1.98	0.045	-
9	火箸風鈴1本（明珍）	1960	7176	3.66	4.08	0.049	0.9
10	松虫風鈴（柏木）	4169	6543	1.57	1.42	0.251	78.9
11	鈴虫風鈴（柏木）	2943	6394	2.17	1.85	0.045	20.0

3-5-9

3-5-9　楽曲および動画の主成分分析の累積寄与率と楽曲が動画から受ける印象影響度との関係

Relationship between cumulative contribution rate of PCA of music and video and the degree of impression influence from media

☆藤井香奈江，脇田由実（大阪工大院・RD 工）

◆音楽と映像を組み合わせて視聴したときにどちらの印象が強くなるかについて分析を行った。

◆音楽と映像の各メディアの特徴量における**主成分分析の累積寄与率の立ち上がりと組み合わせたときの印象影響度の関係**を分析した。

◆音楽に映像を付与した時における音楽の印象変化の受けやすさは Fig.1 のように**ΔMFCC の低次の累積寄与率が高く単純な音楽ほど映像から受ける印象変化が大きくなる**傾向があった。

◆同一楽曲で楽器構成を変え累積寄与率を比較した結果、打ち込みと生演奏の音色の差や音数により情報量が変化することを指示した。

◆映像の RGB 値、オプティカルフローに対する累積寄与率と音楽へ与える印象変化の大きさの関係を調べた結果、低次の累積寄与率が大きく単純な映像ほど音楽に印象変化を与えやすい傾向であった。

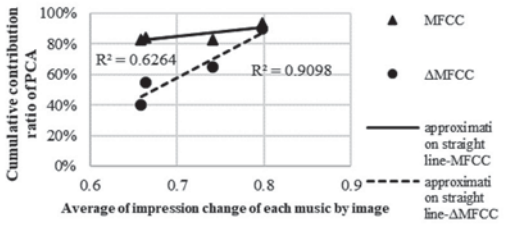


Fig.1：Correlation between impression change and CCR for each music

3-6-1

3-6-1　（招待講演）体力のとらえ方 ～音響学の展開に資する～

[Invited talk] How to Understand Physical Fitness and Sports Medicine
- To Contribute toward the Development of Acoustics -

北川 薫（中京大学）

◆体育・スポーツ界の体力の考え方を筆者の解釈を中心に述べる。

◆これまで、体力を肉体の機能的側面と performance 的側面のあること、体力には心（精神）の働きを無視できないことが指摘されている。

◆また、体力にも健康にも決定的な定義はなく、体力を身体能力の意味に用いるなら、体力に対する概念は精神力であることが指摘されている。

◆筆者は、我が国での健康の概念は人間をひ弱にする、という観点およびWHO の健康の定義から健康を体力とすべきと考える。また、体力要素とその測定方法をまとめており、今後の発展が見込まれる。

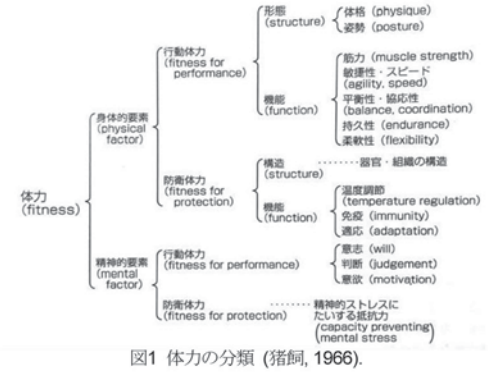


図1 体力の分類（猪飼, 1966）.

3-6-2

3-6-2　スポーツパフォーマンスと音

Sports performance and sound

〇仰木裕嗣(慶大)

◆スポーツに限らず，ヒトの身体運動に伴って生じる音はその動作に由来していることから，動きに関連する何らかの情報を持っていることは容易に推測できる。ヒトは運動によって生じる「出力」である音を，発生と同時に「入力」として聴き取り，これによって運動パフォーマンスを認知し解釈しているといえる。これに対して他者，あるいは機械的な外部の音源によって生じる音を入力としてそれを運動に反映させることも数多く試みられている。自身の情報が「出力でもあり，入力でもある」のか，あるいは「他者の音源情報が入力」であるのか，のいずれかによって運動パフォーマンスと音との関係性は整理できる。本稿では，スポーツパフォーマンスと音との関係についていくつかの事例を紹介すると共に今後の課題について論じたい。

3-6-3

3-6-3　体操競技における音響学的特徴 —音とパフォーマンスに着目して—

Acoustic camera analysis of vault performances to determine acoustic features measured in Gymnastics Competitions

〇久永将太(慶應大・総)，浅野恵子(順天堂大・医)，上田麻理(神奈川工科大・情)

◆昨今、スポーツ競技とその周辺で産出される音との関連性がパフォーマンスの向上に影響する点に着目した研究が注目されつつある。

◆今回、体操競技における音響測定を行い、特に「踏切音」、「着手音」、「着地音」に焦点を当て、「音とパフォーマンス」に着目した。

◆日本代表経験のある選手を含む20~25歳の男子選手４名に実施した。

◆瞬間的に大きな音が発生する「踏切音」及び「着地音」の音源を可視化し、音源位置を視覚的に確認することができた。

◆跳躍板を踏み切った時の「踏切音」では、前方入りの方が後方入りよりも高値であることが示唆された。

◆マットに足が着地した時の「着地音」では、価値点の高い跳越技において高値であることが示唆された。

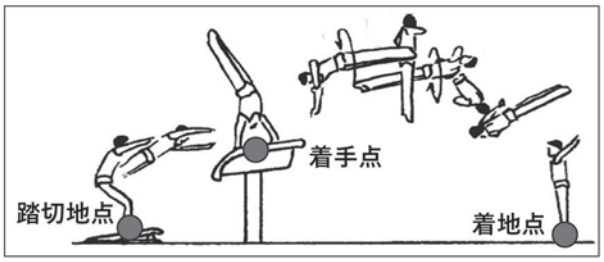


Fig.1：Acoustic camera analysis of vault performances.

3-6-4

3-6-4　体操競技者における単独発話言語の使用についての分析 —「ハミング」、「独り言」、「内的発話」に着目して—

An analysis of the effective use of self-single non-dialogue by gymnasts: a focus on "humming," "self-talking," and "inner speech".

〇久永将太(慶應大・総)，浅野恵子(順天堂大・医)

◆ハミングは口からではなく鼻から出る音で、歌詞はない曲、音を指す。

◆独り言は会話の相手が存在せず、発声を伴う言語を口にする行為、呟く「ことば」を指す。

◆内的発話は自分自身への問いかけや指示出し、思考の整理や調整などに用い、外言音声を伴わない発話を指す。

◆スポーツ競技者が自らパフォーマンス向上のために単独で言葉を発する効果について、アンケート調査を実施した。

◆体操競技選手 45 名 (18~26 歳の男性 23 名、女性 22 名) に実施した。

◆各単独発話行為とも、頻度は高く「毎日」行ない、この発話行為を行うことで、「感情が落ち着く」との回答が多くみられた。

◆各単独発話間で比較すると、男女とも内的発話が最も使用されていた。

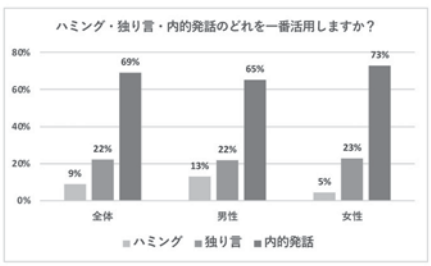


Fig.1：The most frequent usage among humming, self-talk, and inner speech.

3-6-5

3-6-5　バレーボール競技の聴覚情報利用に関する基礎的検討その 4

-プレー時の発生音の分析と競技者の主観的印象-

Examination of Auditory Utilization by Position in Volleyball Part III - Analysis of sounds generated during play and subjective impressions of competitors

☆高橋隼人(神奈川工科大)，三浦貴大(産総研)，上田麻理(神奈川工科大)

本研究では，大学生のバレーボール競技者を対象に，サーブとスパイク時の発生音の測定と，ミート具合やインパクトの強弱などに関する主観評価を実施した．

参加者のスパイクサーブの打球音の 1/3 オクターブバンド分析の結果と，ミートの良し悪しに関する主観評価の相関係数を求めた．結果より，ミートが上手く出来た選手が評価した時は，打球音の約 300~800 Hz 成分は有意に小さく，約 6000~12000 Hz 成分は有意に大きくなっていた．

今後は打球音の判別可否に関して検討を行うとともに，距離・方向定位の訓練に基づく競技力向上の可能性も検討する．

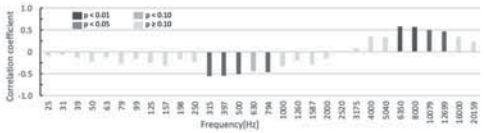


Fig.1: 1/3 オクターブバンド周波数と，主観評価項目における相関

Fig.2：サーブ，スパイクにおける 1/3 オクターブバンド分析結果

3-6-6

3-6-6　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 3 −聴覚を用いたトレーニング方法の提案と評価−

A basic study on the importance of using auditory information in FPS games Part3

Proposal and evaluation of training methods using auditory cues

☆青木大成，大矢敬之，△鳴海颯一郎，塩川茂樹（神奈川工科大），
三浦貴大（産総研），上田麻理（神奈川工科大）

◆我々は2020 年から FPS ゲームにおける聴覚情報に関するインタビュー調査や視覚を制限した実験を行い，FPS ゲームにおける聴覚情報の重要性や，プレイヤーレベルによって聴覚情報を手がかりにプレイしている程度に差があることが確認できた．

◆聴覚情報を活用することで e スポーツの競技力向上を目指し，本研究では FPS ゲームに対する聴覚トレーニングを提案するための予備的な実験を行った．またトレーニングを想定した実験で難易度を分けるため，ゲームキャラクタの足音を分析した．Fig 1 に分析結果を示す．

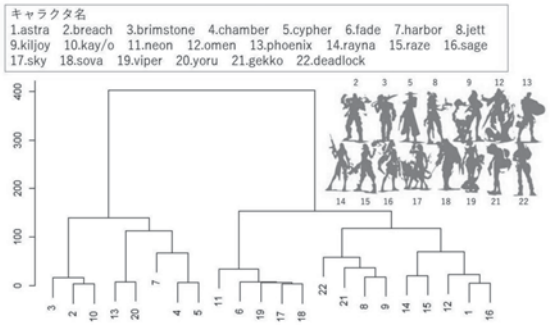


Fig. 1 FPS ゲーム「VALORANT」の全キャラクターのデンドログラム

3-6-7

3-6-7　野球のバッティング練習時の環境音の影響

Effects of environmental sounds on baseball batting practice

○春日秀雄，△馬城遼暉，△小田村涼，△新田晃司，△一色正雄，上田麻理，
△田中博（神奈川工科大）

◆音楽は，スポーツ選手を心理的にリラックスさせたり，モチベーションを向上させたりすると言われている。

- ➢ 練習中に音楽を流す野球チームは多い

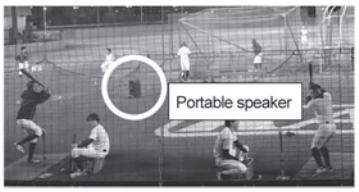


Fig. 1 Batting practice scene

◆異なる環境音で，バッティングにおける打ちやすさやモチベーションなどの変化を調査

- ➢ 3種類の環境（『いつもの音楽』，『クラシック』，『無音』）で実験を実施

✧ 打ちやすさは，『いつもの音楽』がもっともよい結果に

✓ ただし，『クラシック』と『無音』の間には有意差は確認できなかった

⇒ 聞きなれない音楽でもネガティブな影響を与えるとは限らない

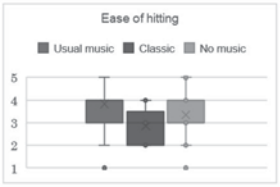


Fig. 2 Questionnaire result: ease of hitting

3-6-8

3-6-8　野球競技における打球音の判別に関する聴能訓練

Hearing training for discriminating the sound of a batted ball in baseball games

☆茂木優真，高橋隼人，二村亮平，（神奈川工科大），
三浦貴大（産総研），△山口雄大（日本体育大学），
春日秀雄，△新田晃司，上田麻理（神奈川工科大）

◆本研究では球種ごとの正否について調べた．
結果，フライはゴロよりも判別しやすい可能性が示唆された．

3-7-1

3-7-1　環境騒音の印象評価に視覚情報が及ぼす影響に関する評価実験

Subjective test on the effect of visual information on impression of environmental noise

○秋山あさひ（東大大学院），中貴一，坂本慎一（東大生研）

◆環境騒音の評価には、聴覚だけでなく視覚情報も影響を及ぼすが、その影響の大きさは環境音の種類に依存する。

◆その違いをより詳細に検討するため、航空機(ac)、道路交通(rd)、鉄道(rw)、風車(wt)に対するアノイアンス評価と SD 法による印象評価を行った。それぞれの環境騒音について映像有り、映像無し(無響室の映像)という条件を提示した。

◆実験には、6 チャンネルの立体音場再生システムとドーム型スクリーンを備えた視聴覚シミュレーションを使用した。

◆映像がある場合にはアノイアンスが低減した。

◆環境騒音の種類によって被験者が抱く印象は異なり、映像の有無によっても安定さや明瞭さという印象が変化した。

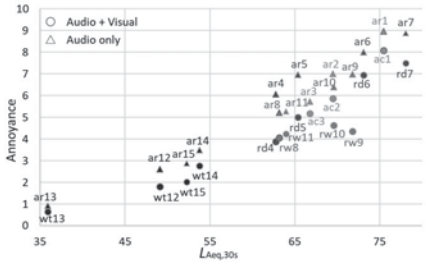


Fig. 1 Annoyance at each $L_{Aeq,30s}$ of the stimulus

3-7-2

3-7-2　低周波数成分を含む交通騒音の主観評価・その１ －圧迫感・振動感及びイベント回数の影響－

Subjective evaluation of traffic noise with low frequency components: effects of the oppressive or vibratory feeling and number of sound events

○横島潤紀（神奈川県環境科学セ/神奈川大），森長　誠（神奈川大），
牧野康一，土肥哲也，横山　栄，小林知尋（小林理研），山崎　徹（神奈川大）

◆周波数成分を含む新幹線鉄道騒音を用いた主観評価実験を実施した。
◆音圧レベル，低周波数成分の有無，騒音イベントの発生回数などの要因が，圧迫感・振動感の発生に及ぼす影響を検討した。さらに，圧迫感・振動感の存在も加えて，新幹線鉄道騒音に対する評価（気になる程度）に及ぼす影響も併せて検討した。
◆ロジスティック回帰分析の結果から，圧迫感の有無に及ぼす影響は，低周波数成分の有無（LF）が最も強い影響を及ぼしていた（Table 1）。
◆重回帰分析の結果から，気になる程度への影響は，圧迫感の有無（Oppression）が最も強い影響を及ぼしていたが，低周波数成分の有無（LF）の影響は有意でなかった（Table 2）。

Table 1: Result of logistic regression analysis.

Item	OR	95%LCI	95%UCI
L_{Amax}	1.174	1.096	1.257
N	1.349	1.053	1.728
LF	2.729	1.433	5.195
Gender	1.144	0.542	2.415
Age	1.013	0.985	1.042
Sensitivity	1.200	0.558	2.581

Table 2: Result of multiple regression analysis.

Item	B	β	p
L_{Amax}	0.075	0.190	<0.001
N	0.314	0.199	<0.001
LF	0.109	0.028	0.584
Oppression	2.133	0.454	<0.001
Gender	-0.567	-0.142	0.015
Age	-0.005	-0.030	0.559
Sensitivity	-0.243	-0.061	0.306

3-7-3

低周波数成分を含む交通騒音の主観評価・その２ －低周波数成分による「気になる」印象の違い－

Subjective evaluation of traffic noise with low frequency components: effects of low-frequency components on annoyance

○森長誠（神奈川大），横島潤紀（神奈川県環境化学センター／神奈川大）
小林知尋，横山栄，牧野康一，土肥哲也（小林理研），山崎徹（神奈川大）

◆回転翼機の音と新幹線の音を題材として，低周波数成分の大小によって騒音全体の「気になる」印象が変化するか検討した．
◆A 特性音圧レベルが等しい場合に，低周波数成分をカットしていないオリジナル刺激の方が，ハイパス刺激よりも「気になる」と回答される割合が高いことが確認された．
◆一対比較における２つの刺激の主観的等価点に該当するときの L_{EA} の差は Fig. 1 に示す通りである（z = 0.5 の時の DLEA）．
◆ロジスティック回帰分析の結果，一対比較において，オリジナル刺激を「気になる」音として選択する割合は，A 特性音圧レベルの差だけでなく，40 Hz と 50 Hz のエネルギー合成値も有意に関連していた．

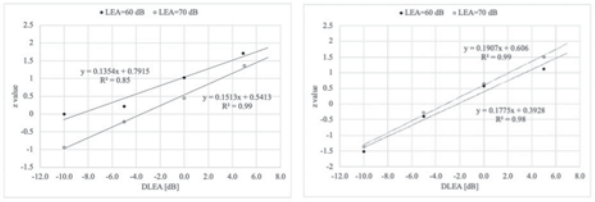


Fig.1：Level differences between stimuli in the pair comparisons and the percentage of respondents who chose the original stimulus as a more

3-7-4

3-7-4　血流改善効果を持つ低周波音の同定

Identification of low-frequency sound that increases blood flow

○岩崎成仁，香川匠（名古屋大院），大神信孝（藤田医大/名古屋大院），
加藤昌志（名古屋大院）

◆100 Hz 以下の低周波音に長時間曝露されることにより、健康障害が誘発される可能性が指摘されている。一方、低周波音を短時間曝露された場合の健康への影響は不明であった。そこで、本研究においては、低周波音を短時間暴露した際の生理作用について検討した。
◆デジタルサウンドジェネレーターを用いて、60 Hz から 150 Hz までの低周波純音を合成した。各周波数の低周波純音を 2 分間 85 dB で手掌に曝露し、皮膚血流を測定した。
◆60 Hz から 100 Hz の低周波純音は、いずれもコントロール群と比較し、有意な血流増加を認めた（Fig 1）。
◆本研究により、低周波音には血流改善効果があることが明らかとなった。今後はこの成果を職場、住宅といった日常生活空間へ応用して、健康増進がはかれる環境を作りたい。

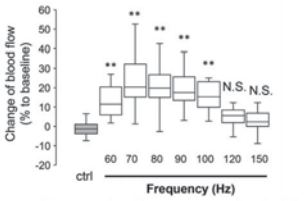


Fig.1：Cutaneous blood flow in hands.
Deng et al., Sci Total Environ. 2022
(Elsevier より図の掲載許可あり)

◆今後の課題: 生活空間へ応用するためには、離れた位置からでも、対象(ヒト)に向かって一定の周波数の音を正確に届ける必要がある。音を高い指向性をもって対象へ届ける技術が必要である。この点について、医工連携研究を進められるパートナーをみつけたいと願っている。

3-7-5

3-7-5　運搬可能な低周波音体感装置の開発 —「低周波音体験車」の使用事例紹介—

Development of a transportable low-frequency sound sensation chamber: Case studies of the "Low-Frequency Sound Experience Vehicle"

◎岩永景一郎，土肥哲也（小林理研）

◆トラックで各地に移動できる「低周波音体験車」を開発 (Fig.1)
◆最大３名同時に 100 dB を超える低周波音を体験可能 (Fig.2)
　① 感覚閾値を超える超低周波音
　② 低周波音で感じる圧迫感・振動感
　③ 実環境で収録した低周波音を含む騒音
◆ 各地のイベントなどで 1,000 名以上が低周波音を体験

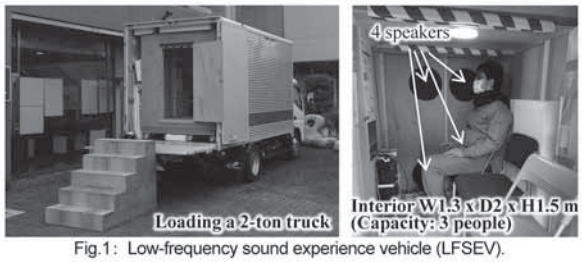


Fig.1：Low-frequency sound experience vehicle (LFSEV).

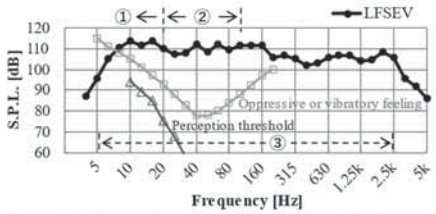


Fig.2：The maximum S.P.L. that can be reproduced by LFSEV.

3-7-6

3-7-6　発話時の音環境に着目した飛沫粒子数の予測モデルの検討

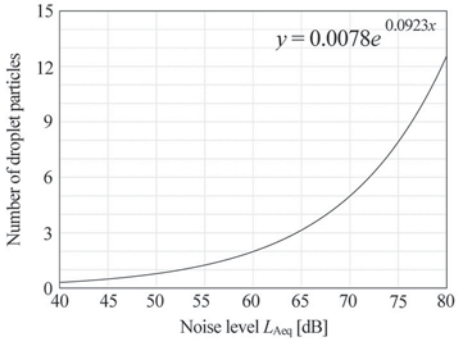
Study on a model for predicting the number of droplet particles focused on the sound environment during speech.

○辻村壮平(茨城大院・理工研), 木谷恒陽(茨城大院),
△李艶栄(茨城大院・理工研)

発話者の聴覚フィードバックを考慮した会話時の飛沫粒子の放出量の予測モデルを構築するため，騒音レベル 50～70 dB の 5 dB ごとに連続した 5 水準の条件で発話を伴う心理音響実験を実施した．飛沫粒子の数を目的変数とし，騒音レベルを説明変数として回帰分析を行った結果，下の知見が得られた．

◆ 騒音レベルと飛沫粒子数の関係について，以下のような予測式が得られた（$r^2 = 0.90$）．

◆ この予測式を用いて，発話時の騒音レベルと飛沫感染リスクの関係を発話時間ごとに算出した．



Figure：Prediction model for number of droplet particles using noise levels

3-7-7

3-7-7　新型コロナウイルス感染症によるステイホーム中の音環境アンケート調査 －マンションとアパートの比較－

Sound environment questionnaire survey during stay home due to COVID-19 －Comparison between condominiums and apartments－

○冨田隆太(日大・理工), △阿部今日子(日大・芸術)

◆筆者らは，はじめて緊急事態宣言が出され，解除された直後の 2020 年 6 月に共同住宅の音環境に関するアンケートを実施した結果を既報にて報告した。

◆本報では，マンションとアパートの居住者反応を明らかにする目的で，2020 年 6 月に実施したアンケート調査のクロス集計結果を報告する。

◆Fig. 1 に他の住戸や外からの音について聞こえる程度を質問した結果を示す。本報では，マンションとアパートでの聞こえる・気になる・うるさい・問題があるの指摘が異なること等が示唆された。

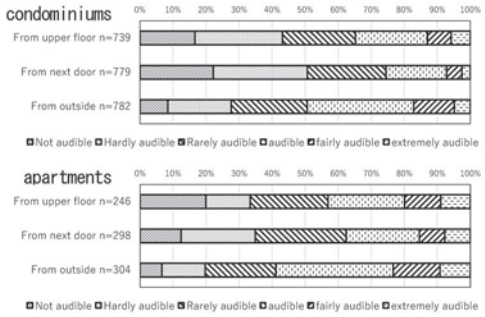


Fig. 1　Hearing level during stay home

3-7-8

3-7-8　質問紙を用いた高齢者のフレイル予防のためのスクリーニング調査 –音声コミュニケーションの関わりとフレイル予防のための調査–

Screening surveys in the elderly for preventing frailty using a questionnaire - A study on relationships between sound communication and frailty prevention -

☆高橋達樹，上田麻理(神奈川工科大)，廣江正明(小林理研)

- 日本の超高齢化を受けて，内閣府下で自治体等が実施している高齢者の生活実態調査結果からフレイル予防のスクリーニングの可否を検討するため，首都圏の地方自治体の調査結果から質問票および調査結果の洗い出しを行った．
- 本報では，過去３ヵ年分の「介護予防・日常生活支援総合事業対象者」からの回答を一括分析した．
- フレイルに係る回答から総合・運動・栄養＋口腔・閉じこもり・物忘れの各フレイルを求め，性別，年齢，体格(BMI)，家族構成，介護等の必要性，暮らし向きの他，耳の聞こえや病気との関連を分析した．
- 総合フレイルは耳の病気を除く 7 つの属性との間に統計的に有意な関連性があった．

Table 1　総合フレイルと属性や耳の聞こえ・病気との関連性

	総合	運動	物忘れ
性別	0.610 (0.000**)	1.441 (0.000**)	0.978 (0.750)
年齢	1.111 (0.000**)	1.087 (0.000**)	1.045 (0.000**)
体格指数	2.083 (0.000**)	1.049 (0.678)	1.107 (0.311)
家族構成	1.869 (0.000**)	0.979 (0.806)	1.282 (0.001**)
介護等	7.319 (0.000**)	5.079 (0.000**)	1.806 (0.000**)
暮し向き	0.579 (0.000**)	0.591 (0.000**)	0.888 (0.095†)

† $p<0.10$, * $p<0.05$, ** $p<0.01$

3-7-9

3-7-9　居住者視点に基づく調査票による社会反応の基礎検討

Basic study on community responses using a survey sheet based on residents' perspectives

☆古味 由惟(神奈川大院・工学研), 横島 潤紀(神奈川大学・工/神奈川県環境科学センター), 森長 誠(神奈川大学・工), 須田 直樹(小野測器),
梅﨑 良樹(社会システム), 山崎 徹(神奈川大学・工)

これまで筆者らは，道路交通騒音に対する社会反応に関する研究を行ってきた．本報では，既往研究で得られた結果をもとに調査票を作成し，社会調査を新たに神奈川大学分野横断型研究推進事業にて，実施した結果を報告する．具体的には，ロジスティック回帰分析を適用し，騒音曝露量からアノイアンスおよび生活妨害に対する影響を住宅種別で示す．加えて，従来の調査と本調査の結果を比較する．結果として，図１より従来の調査結果では，騒音曝露量から聴取妨害に対する影響の傾向が住宅種別で異なっていたが，本調査の結果では戸建て住宅，集合住宅ともに同様の傾向であることがわかった．今後，追加調査の結果も踏まえ，分析を進める予定である．

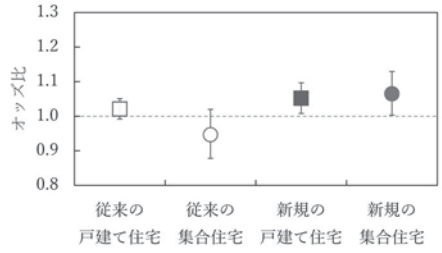


Fig. 1 Odds ratio and 95% CI results for logistic regression analysis (listening disturbance)

3-7-10

3-7-10 Perceived Affective Quality Scale でサウンドスケープの回復的特性の評価は可能か？

Can the Perceived Affective Quality Scale for soundscape research evaluate the restorativeness of soundscapes?

○永幡幸司，△齊藤みのり，△舟山大輝，△石井卓実（福島大学）

【本研究で用いた尺度】

◆Perceived Affective Quality Scale (PAQS): 個々のサウンドスケープが持つ人（々）の感情を変える能力（Affective Quality）を快（PLEASANTNESS）と覚醒度（AROUSAL）の２次元で評価する尺度。ISO/TS 12913-2 で提案された尺度そのままだと，正しい意味での Affective Quality の尺度となっていないことに留意が必要である。本研究では，Nagahata による日本語版 PAQS[1)]を用いた。

◆Perceived Restorativeness Scale (PRS): Kaplan が唱える Attention Restoration Theory（ART）を基に作成された，環境の持つ回復特性（Restorativeness）を評価する尺度。「逃避」「魅了」「まとまり」「視野」「適合」の５つの下位尺度を推すものと，「逃避」「魅了」「まとまり」「視野」の４つの下位尺度を推すものが存在する。本研究では，芝田らによる日本語版 PRS[2)]を用いた。

【本研究の結論】

◆PRS は「逃避」「魅了」「まとまり」「視野」の４つの下位尺度を採用することを支持する結果が得られた。

◆４次元で表される回復特性を，２次元で表される PAQS のモデルで適切に評価することはできない。

1) K. Nagahata: A trial of translation of the perceived affective quality attributes for soundscape research to Japanese. Appl. Acoust. (in print)
2) 芝田征司，畑倫子，三輪圭子: 日本語版 Perceived Restorativeness Scale (PRS) の作成とその妥当性の検討, MERA Journal, 21, 1-10, (2007).

3-7-11

3-7-11　少数マイクロホンによる遠距離音源定位の精度検証

Accuracy of Estimation of Distant Sound Source Locations using a small number of microphones

○山﨑 大，△古谷 政樹，△宮田 陽水，小池 義和（芝浦工大）
△清水 悦郎，△梅田 綾子　（海洋大）

◆視界が悪い状況における定位技術として、音源定位が注目されている。

◆従来の研究では、多数のマイクロホンを利用して、より精度が高い音源位置推定技術が研究されている。

◆本研究では、２つのマイクロホンで構成されたアレイによる音源位置推定精度について検証を行っている。

◆遮へい物の少ない屋外にて移動音源の音源定位を行い、cm 級衛星測位結果との比較を行い、考察を行った。

◆音源到来方向推定の最大誤差は 10° 程度であったが、距離推定には大きな誤差が見られ、さらに到来方向推定精度を向上させる必要があると考えられる。

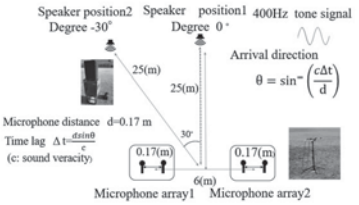


Fig.1 Experimental setup of sound source localization using two microphone arrays.

Table 1: Estimation results

Pos.	Setup	DoA (deg.)	Distance (m)
1	Mic	-3.4	54.1
	GNSS	6.8	25.7
2	Mic	-26.3	72.0
	GNSS	-25.2	25.6

3-7-12

3-7-12　空間フーリエ変換と平面波合成を用いた音源移動に伴う振動音源の振幅変調に関する基礎検討

A basic study on amplitude modulation of vibrating sound source due to sound source movement using spatial Fourier transform and plane wave synthesis

☆牧野 裕介、大谷 真、高野 靖（京大院・工学研）

◆ 音源の移動によって曲げ波からの放射音が変化

◆ 平面波合成にて音圧分布を計算（Fig.1）

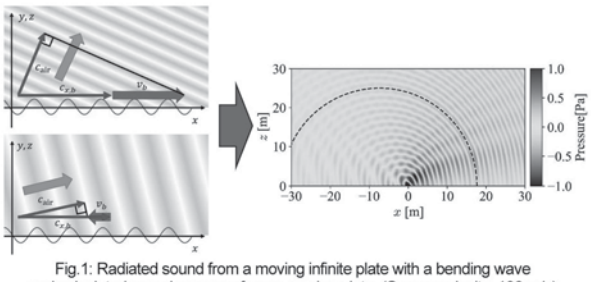


Fig.1: Radiated sound from a moving infinite plate with a bending wave and calculated sound pressure from a moving plate. (Source velocity: 100 m/s)

◆ Doppler factor $(1 - M\cos\theta)^{-1}$ と放射音圧との関係（Fig.2）

→理論より過大評価？

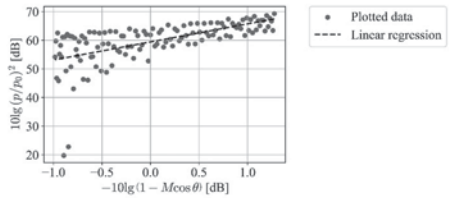


Fig.2: Calculated sound pressure level and Doppler factor. (Source velocity: 100 m/s)

3-7-13

3-7-13 耳栓・イヤーマフ併用による遮音性能の測定

Sound attenuation provided by a combination of ear-plugs and ear-muffs.

○横山 栄，小林知尋（小林理研）

高騒音化での長時間作業を行う際等には、難聴を防止する観点から、耳栓やイヤーマフ等の聴覚保護具を着用する事が有用である。今年４月に「騒音障害防止のためのガイドライン」が約 30 年ぶりに改訂され、聴覚保護具については JIS T 8161-1:2020 の試験方法により測定された遮音値を目安に必要かつ十分な遮音値のものを選定することが明記され、非常に強烈な騒音に対しては耳栓と耳覆いの併用が有効であるとも解説されている。現在、耳栓やイヤーマフをそれぞれ単体で装着した際の遮音性能測定については、上記の JIS に試験方法が規定されているが、併用時は対象外で、測定事例も少ない。そこで、本報では、耳栓とイヤーマフを併用した際の遮音性能測定に上記の JIS による試験方法を応用する事を試みた結果を報告する。

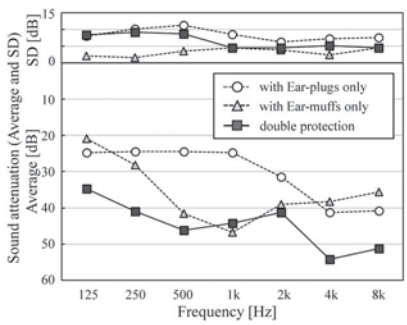


Fig.1 : Sound attenuations compared the double protection and ear-plugs/ear-muffs only.

3-7-14

3-7-14　HATS を用いた聴覚保護用耳栓の遮音性能特性測定の試み

Attempt of sound attenuation measurement of hearing protection earplugs using HATS

○稲永 潔文（サザン音響）

◆作業環境における騒音暴露による聴力損失を防止するため、様々な遮音性能を持つ耳栓や聴覚保護具が用いられている。そしてその性能特性の評価には Real Ear Attenuation at Threshold（以下 REAT）と呼ばれるヒトの聴覚閾値の差を測定する主観評価法が採用されている

◆一方で、近年ノイズキャンセル型ヘッドホンおよびイヤホン（以下NCHP）が普及し、これらの騒音低減量等の性能評価法が JEITA RC-8142A として規定されている。

◆従来、官能検査で行われてきた聴覚保護具の遮音性能特性測定を、前報の耳覆い型聴覚保護具に引き続き、耳栓も物理測定で行うことを試みた。

◆聴覚保護用耳栓には、3M 社“E-A-R TM”318-4001（紐付き）”等を、HATS にはサザン音響 “SAMAR Type 4500”、また測定用音場には RC-8142A の規定を満たすサザン音響“小型疑似拡散 Box Type1020”を用いた。

Fig.2a Externals of HATS used for measurement without hearing protection earplugs & 2b with protection earplugs

◆測定の結果、耳栓の挿入圧により遮音特性が大きく変化するため分散も大きく、従来主観評価値とは異なる結果となった。

◆これは、規格により耳道が内径 7.5mm の硬い円筒で実耳とは条件が異なる事が要因と思われ、耳道部改良が望まれる。

◆実耳に耳栓を押し込む場合は、低域で生理的雑音によるマスキングが生ずることによる遮音値増大も考えられる。

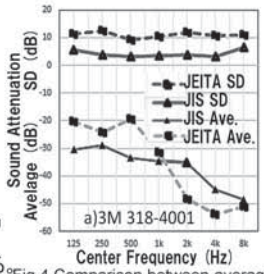


Fig.4 Comparison between average and standard deviation of interception characteristic measurement results of earplugs for various aural protection by JIS and JEITA method.

3-8-1

3-8-1　室内音場の狭帯域分析について －成長・減衰過程の 1/N オクターブ分析－

Narrow-band analysis of sound fields in rooms - 1/N octave analysis for growth and decay processes -

○山田祐生（竹中技研）

◆室内音場の成長・減衰過程では，入力信号の断続にともない，音響モードに対応した入力周波数以外の成分も励起される.

◆この現象と帯域フィルタ (BPF) の効果に関する知見は帯域分析に際して有用であるが，検討例は多くない.

◆今回あらためて狭帯域分析の問題として定式化し，検討を行った.

◆実際に室の減衰過程の分析を試み，BPF の影響，スペクトルの広がり等について確認した.

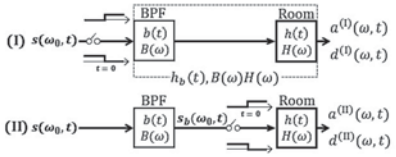


Fig.1：Sound fields and measurement systems.

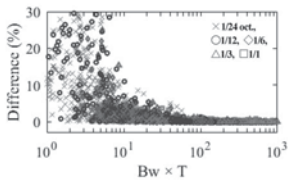


Fig.2：Difference of reverberation time in the system I and II for various rooms.

3-8-2

3-8-2　周期チャープ信号の同期加算を用いた高 SN 比遮音測定法

High SN ratio sound insulation measurement method using synchronous addition of periodic chirp

○小柳慎一郎，山田祐生（竹中技研）

◆実建物における遮音計測法として cyclic chirp 信号を用いる手法を提案する. 本手法では同期加算と計測信号の短縮により SNR が改善し，高遮音構造の計測でも大きな音源を必要としないメリットがある.

◆周期が短いほど SNR は改善するが周波数分解能が低下する問題がある. 伝達関数のスペクトルがガウシアンに従うと仮定すると，フーリエ変換のスペクトルの個数と計測誤差の関係はカイ二乗分布を用いて説明できる. 例えば 9 本のスペクトルを計測対象の周波数帯域に組むことで約 95%の確率で誤差が 3dB 以内となる.

◆提案手法を用いて実建物の遮音性能を計測した. 一例目は音源室(室容積 320m3)，受音室(175m3)の界壁を共有する居室間で性能は D-55 であった. 従来法では SNR が確保できなかった帯域でも提案手法では暗騒音が低減し SNR が確保された。二例目は 50m 程度の離隔のあるホールと会議室間の計測結果である。従来法は D-70 以上，提案手法では D-85 以上の遮音性能が確認できた.

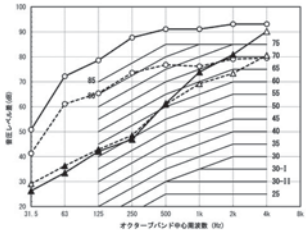


Fig.1：Sound level difference levels measured by the proposed method and the conventional steady noise method: Dotted lines indicate conventional methods, and white markers indicate data with insufficient SNR.

3-8-3

3-8-3　小型模型を用いた円形空間内の音場解析

Analysis of the sound field in circular spaces using small scale models.

☆今西麻優子，野澤遙，鈴木七彩，及川靖広（早大理工）

◆研究内容：**光学的音響計測手法を用いた，円形模型内部の音場計測**

◆円形空間には以下のような音響的問題が存在する。

- 焦点・デッドスポットが存在して音の**拡散性が悪い**
- 「**囁きの回廊**」現象が起こる

◆本研究の目的・手法

- 円形空間の**音場の計測** → PPSI で円形模型内部の音場を可視化
- 円形空間の**音響的問題の解決** → 円形模型に拡散体を設置
- Schroeder が提案した，原始根に基づいて生成される数列のランダム性を利用した拡散体を円形空間の壁面へと応用した。

◆計測結果

- 原始根に基づく拡散体を設置 → **囁きの回廊現象が改善**された。

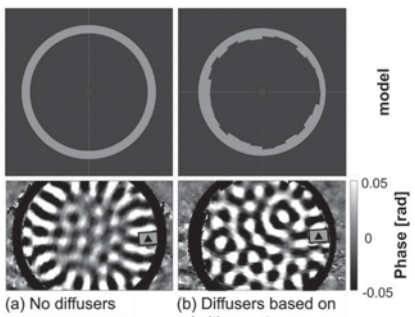


Fig.1：Models and results. (▲:sound source)

3-8-4

3-8-4　アンサンブル平均を用いた吸音特性の現場測定法　−センサーの現場校正について−

An in-situ measurement method using ensemble averaging technique for sound absorption of materials
-- Effectiveness of on-site sensor calibration -

○大鶴　徹, 富来礼次, 岡本則子(大分大・理工)

◆2マイクロホンによるアンサンブル平均を用いる材の吸音特性測定法（EA_{pp}法）について検討

◆EA_{pp}法へエレクトレットマイクロホンを適用し、グラスールの吸音特性を残響室で測定

◆校正前、並びに、平均操作適用前の吸音率はばらつきが大きい

◆現場校正と 1/3 オクターブバンド平均操作の適用で、既往の参照値へ概ね近似する吸音率が得られた : Fig.1 参照

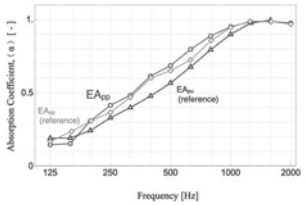


Fig.1: One-third octave-band mean values of absorption coefficient ⟨α⟩ of GW50 measured by EA_{pp} method with sensor-calibration using electret microphones (red line). Reference values measured by EA_{pp} method using condenser microphones (orange line) and those by EA_{pu} method using a PU-sensor (blue line) are plotted together for comparison.

3-8-5

3-8-5　帯域制限空間の視点から見た周波数領域音場

Frequency-Domain Sound Field from the Perspective of Band-Limited Functions.

○岩見貴弘，尾本章（九大・芸工）

◆我々は既報で，周波数上限のある一般の空間次元の瞬時音場を球状帯域制限空間の元とみなし，この空間の再生核の線形和で音場を表現することにより幾つかの高速な信号処理手法を提案している。

◆本稿では，一般の空間次元の周波数領域音場に着目し，ある周波数の音場の集合は帯域制限の視点から考えると，スペクトルを周波数に依存した半径の超球面に制限した関数と一致していることを示し，この空間の再生核を導出する。

◆導出した再生核は，3 次元に限定すれば植野らの Helmholtz 方程式の解集合に適切な内積及びノルムを導入した再生核 Hilbert 空間の再生核と同値である。

◆本稿の再生核を用いることにより，2 次元及び 3 次元音場（場合によってはより高次元の音場）を同様の枠組みで扱うことができる。

3-8-6

3-8-6　周波数領域における一般指向性マイクのモデル化

Modeling of General Directional Microphone in the Frequency Domain.

○岩見貴弘，尾本章（九大・芸工）

◆我々はこれまでに軸対称な指向性マイクの入力信号を球面調和領域でモデル化している。しかし，これは軸対称な指向性に限定されることに加え，全てのマイクを同径方向に向けることを制約されていた。

◆本稿では，一般の空間次元における任意の指向性マイクをモデル化する手法を提案する。このモデル化は指向性関数を多項式表現に基づく微分作用素に置き換えることにより実現される。

◆周波数領域音場の再生核表現に対して提案のモデル化を適用し，音場推定の数値シミュレーションを行うことにより性能を検証した。

◆提案モデルを用いた音場推定結果は，全てのマイクを全指向性と仮定した場合の推定結果より良い結果を示した。

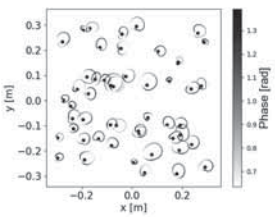


Fig.1　Microphone placement and directivity.

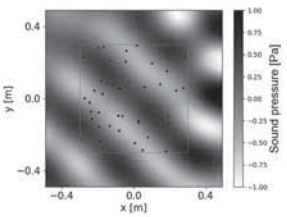


Fig.2　Estimated sound field using proposed method.

3-8-7

3-8-7　マイク指向性推定の一検討

A study of microphone directivity estimation

☆加藤真木，岩見貴弘，尾本章（九大・芸工）

◆周波数領域における再生核を用いた音場表現に対し，マイク指向性関数を斉次多項式で表して微分作用素に置き換える指向性マイクモデル化手法が提案されている。

◆球面調和関数を定義域を球面上に制限した斉次の調和多項式とみることで，指向性マイクのモデル化に適用する。

◆球面調和関数は正規直交基底であるためロバストな指向性推定が期待できるほか，調和性を考慮している分少ない基底で表現が可能である。

◆シミュレーション実験を行い，ロバスト性と実行時間において球面調和関数を基底に用いる優位性を確認した。

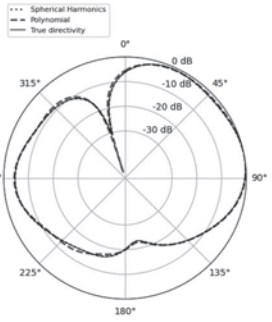


Fig.1　Estimated Directivity

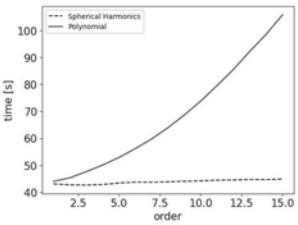


Fig.2　Runtime for directivity estimation

3-8-8

3-8-8　比較的小さな一般空間の音環境評価手法の検討

Examination of sound environment evaluation method for relatively small general space

○大山宏，山下晃一（日本音響エンジニアリング），△若杉浩一（武蔵野美術大），△藤原義一（大建工業），△山田聖士，△松井香代子，△齊藤治宣（内田洋行）

◆室内の音場の特性を表す指標はこれまで提案されてきてはいるが，それらはコンサートホールのような大空間を想定したものが多く，比較的小さな一般空間の音場の状態の良し悪しを表せるような音響指標があれば，より音響的に快適で豊かな空間の実現の一助となるのではないかと考える。

◆インパルス応答波形を高速時定数処理したレベル波形（残響減衰波形）は反射音列のレベル変動が可視化され，直感的に分かりやすくなっていることに着目し、これらの波形の違いを数値化するために，波形の近似直線との偏差の平均値（Fig.3）を 計算し比較してみたところ（Fig.4），それらの部屋を体験した人間の聴感印象と相関がある可能性があること（この値が小さい方が印象がよく高評価）が分かった。

◆音場条件の異なる（空室，吸音，音響調整家具）同容積の小部屋における被験者実験結果は，音声・音楽の聴取における聴感評価と残響減衰波形から導かれた定量値（残響減衰変動）との相関があり，小空間での聴き取りやすさの評価の指標として使える可能性を得た。

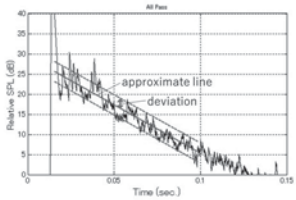


Fig.3：Approximate line and deviation of reverberation attenuation waveform

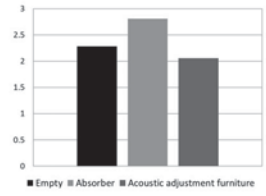


Fig.4：Deviation from the approximation line of the reverberation decay waveform (reverberation decay fluctuation)

3-8-9

3-8-9　残響の多い保育室への吸音材仮設の効果に関する現場実験（第２報）

A field experiment on the effect of sound absorption installed to a highly reverberant kindergarten classroom: A second report.

○川井敬二，△松藤優紀，△石澤碧，△津留芽実（熊本大）

◆保育施設を対象とした実地視察の際に見つかった響きの多い保育室において，仮設吸音材の取り付け・取り外しを繰り返す現場実験を２年間にわたり継続し、室内音を録音した。

◆１年目は吸音のない条件で４月から 11 月まで過ごした後に吸音材を仮設し，２年目は４月に吸音材を仮設したまま 11 月まで過ごした後に吸音材を取り外す，という初期条件設定とした。

◆分析は，1)自由遊びでの子どもの発声に着目した室内音圧レベルと，2)子どもと先生の歌声の発声レベルの２項目について，吸音材の仮設そのものによる低減量(約 4dB)を補正して，吸音の有無の条件間で発声音量を比較した。

◆結果として，１年目の最初の吸音なし条件（N21-1）で自由遊び・歌声とも発声音量が大きく，それが吸音の仮設（A21-1）で大きく下がった（Fig.1）。

◆一方で，その後の条件下では，レベル補正後の発声音量は，吸音の有無によらず大きな変化は見られなかった。

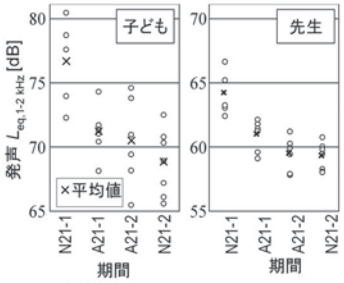


Fig.1：
Daily voice level of singing together before lunch. Levels in N periods are adjusted to A(absorbed) periods.

3-8-10

3-8-10　実教室における児童の音声了解度評価

Speech Intelligibility Assessment in Japanese Elementary School Classrooms

○エバンズ直子（TOA(株)/大阪大院・基礎工），△長崎育恵（大阪市教育センター），△橋本峻，△加藤翔，△國光妙子(大阪教育大・附属天王寺小)，湯浅哲也(大阪教育大・教)，△阪本浩一（大阪公立大院・医学），△清野健（大阪大院・基礎工）

◆児童が教員からの音声による指示や情報を正確に聴き取れることは、学校教室における学びの前提条件として不可欠である。

◆海外には教室内音環境の改善策のためにガイドライン等が整備・認知され、教員が授業中にマイクやスピーカーを使用する国も多い。

◆日本の学校教室で授業中に実施できる、新たな日本語音声了解度評価方法を検討し、実教室で小学校６年生の児童を対象に検証を実施した。

◆得られた 85 名のデータについて、教員がマイクと調整されたスピーカーを使用しない場合（Fig.1、“off”）とする場合（同、“on”）に関し分析した結果、教員が授業中にマイクと調整されたスピーカーを使うと、児童の日本語音声了解度が向上すると推定できる。

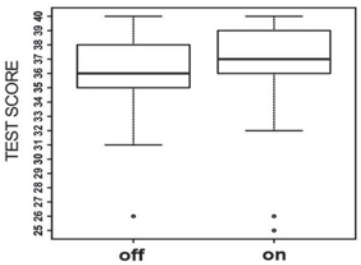


Fig.1：Speech intelligibility test score comparison; with microphone and loudspeakers (right - “on”), and without (left - “off”)

3-8-11

3-8-11　木造及び RC 造住宅における二重窓の低周波防音効果の比較検証

Comparative examination of low-frequency soundproofing effects of double window system in wooden and RC houses

◎劉金雨，佐久間哲哉（東大・工），△鐘築吉文，△秋光昂，塚本陽平（YKK AP 株式会社）

◆既報の木造試験家屋に加えて，既存の木造及び RC 造住宅における現場測定と実験室測定を行い，二重窓の低周波防音効果を調べた。

◆Fig. 1 に，各室における二重窓の準音響透過損失値を示す。軽量の木造では外壁透過が大きいために二重窓の効果は殆ど得られないのに対して，RC 造では実験室と同程度の効果が得られるが，遮音性能が著しく低下する室内奥行き方向の固有周波数付近ではむしろ逆効果となることが明らかとなった。

◆また，Fig. 2 に 100~200 Hz 帯域では，窓シャッターと外窓による共鳴透過により遮音の低下が生じることが確認された。

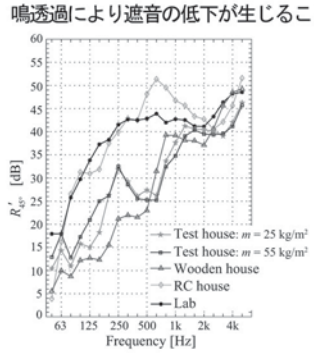


Fig.1：$R_{45°}$ results among different wall structures for double window.

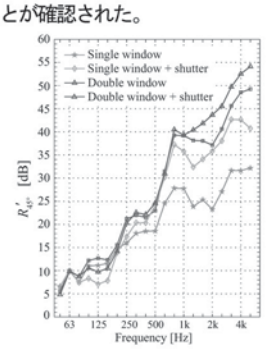


Fig.2：Effect of the shutter in the wooden house.

3-8-12

3-8-12　斜入射音響透過損失の実験室測定における音源条件に関する数値解析的検討

Numerical investigation for source conditions on laboratory measurement of oblique-incidence transmission loss

☆山崎泰知(大成建設/東大・新領域)，田中ひかり，増田潔(大成建設)，
井上尚久(九大・芸工)，佐久間哲哉(東大・工)

◆斜め入射条件におけるガラス板の遮音性能について，実験室実験及び数値解析により検討を行った．

◆Figure 1 に示すように，実験室実験と数値解析の結果は概ね一致した．

◆音源からガラスまでの距離が異なる条件及び平面波入射条件で行った数値解析結果（Fig. 2）の結果，コインシデンス限界周波数以上の帯域において，入射条件により値が大きく異なることが示された．

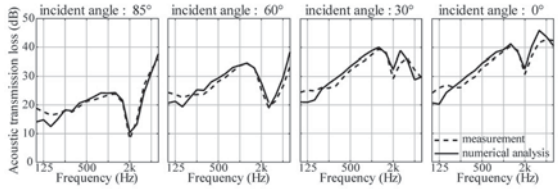


Fig. 1 : Oblique-incidence transmission loss of numerical analysis and measurement (2.25 m sphere wave)

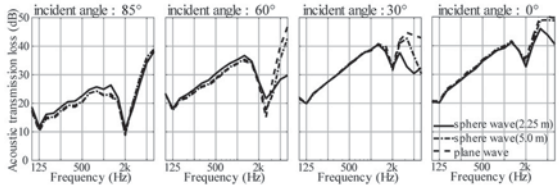


Fig. 2 : Numerical analysis of oblique-incidence transmission loss under sphere waves (2.25 m / 5.0 m) and plane wave conditions

3-8-13

3-8-13　VR 平面空間における先行音効果の検証—空間内各位置の評価順序を考慮した音の方向感特性—

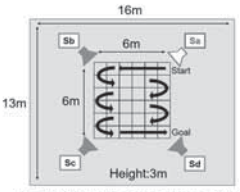
Investigation of the precedence sound effect in VR plane space - Characteristics of sound direction when considering the evaluation order of each position in space -

☆会田征太郎(日大院・理工)，大隅歩，伊藤洋一(日大・理工)

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

◆これについては，対象となる建物空間に本システムを設置するときの各種音響条件の設定に，VR 技術の利用を提案している．

◆本報告では、構築した VR 平面空間における各位置での音の方向感評価について，評価順序の影響について検証を行った．



実験条件
① 一定の順序に従って評価
② ランダムな順序での評価

Fig. 1 : Schematic view of the experiment.

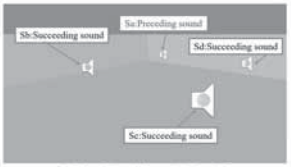


Fig. 2 : Visual information in VR.

3-8-14

3-8-14　動的 HRTF の音楽練習 VR 空間への適用

Dynamic HRTF application for music ensemble training in VR space

〇棚瀬 廉人(ヤマハ)，△貝原正三(新日フィル)，
△下釜恵黎也，小谷創(クレッセント)

◆VR コンテンツ制作における立体音響技術の使用例として、音楽合奏練習の事例を紹介する。仮想ホール空間内の自由な場所で、プロ演奏のインタラクティブな聴取や合奏練習をすることができる。

◆コンテンツは新日フィルが制作した。演奏のボリュメトリック映像に、演奏収録音を重ね、CG のすみだトリフォニーホールと合成した。

◆プログラムはゲームエンジン Unity で制作し、VR ゴーグルや Windows で使用することができる（販売は VR ゴーグル用のみ）。

◆演奏の収録には通常のステレオマイクのほか、立体音響マイク（ViRealMic）を用いた。VR 空間内の再生では、位置に応じて動的に変化する HRTF により方向感を再現した。

Fig.1 : VR Apps. for music ensemble training.
(Planned and produced by New Japan Philharmonic Orchestra)

3-9-1

3-9-1　自己教師あり学習モデルを用いた zero-shot 音声合成の性能評価

Performance evaluation of zero-shot text-to-speech synthesis with self-supervised speech-representation model

◎藤田健一，芦原孝典，金川裕紀，森谷崇史，井島勇祐（NTT）

◆自己教師あり学習（Self-supervised Learning; SSL）により得られた音声表現抽出モデル（SSL モデル）を zero-shot 音声合成に活用した手法において，SSL モデルが異なるデータで学習された時の音声合成の品質に与える影響を評価

◆比較の概要
- SSL モデルとして，それぞれ，日本語話し言葉コーパス（CSJ），LibriSpeech，ReazonSpeech で学習された HuBERT を利用し性能を比較

◆zero-shot 音声合成の性能評価を実施
- 最も大規模な ReazonSpeech を用いると，メルケプストラム距離と音素継続時間長の生成誤差が改善することを確認
- CSJ，LibriSpeech の比較から，SSL モデルの学習データの言語依存性が小さいことを確認

Model	MCD	Duration (ms)
HuBERT(CSJ)	5.26	15.0
HuBERT(LibriSpeech)	5.24	15.4
HuBERT(ReazonSpeech)	5.14	14.3

Table 1 : 客観評価結果. DurationはRMS誤差

3-9-2

3-9-2 Hi-Fi-CAPTAIN：NICT 高品質長時間日英男女パラレル音声合成コーパス

Hi-Fi-CAPTAIN: High-fidelity and high-capacity speech synthesis corpus developed by NICT

○岡本拓磨[1]，志賀芳則[2,1]，河井恒[1]
([1]情報通信研究機構，[2]現所属：東京電機大学)

・**音声合成用公開コーパスHi-Fi-CAPTAIN**
- ・CC BY-NC-SA 4.0ライセンスにて2023年8月に無償公開
- ・対話調
- ・サンプリング周波数48 kHz，24 bit，リニアPCM
- ・アメリカ英語：女性15.1時間，男性15.0時間 (ともに14,000文)
  うち12,988文がパラレルコーパス
- ・日本語：女性23.3時間(19,056文)，男性22.3時間(19,058文)
  うち18,855文がパラレルコーパス
- ・ESPnet2-TTSで学習・合成可能なレシピ同梱
  - ・End-to-end(E2E)テキスト音声合成：JETS
  - ・E2E-sequence-to-sequence声質変換：VITS-VC，JETS-VC

https://ast-astrec.nict.go.jp/en/release/hi-fi-captain/

3-9-3

3-9-3 Coco-Nut: 自由記述文による声質制御に向けた多話者音声・声質自由記述ペアデータセット

Coco-Nut: Corpus of connecting Nihongo utterance and text toward prompt-based control of voice characteristics,

☆渡邊 亞椰，高道 慎之介，齋藤 佑樹，辛 徳泰，猿渡 洋
(東大院・情報理工)

◆声質を表現する自由記述文による話者コントロールが可能な TTS，Prompt TTS の実現を目標とする．

◆Prompt TTS の学習に用いることのできるオープンかつ多話者の音声-声質自由記述ペアデータセットが現状存在しないため，構築する．

◆Web 上の音声を収集し，音質・発話内容の保証をした上でクラウドソーシングにより声質自由記述を収集した．

◆収集データの多様性について評価を行ったほか，自由記述の傾向について分析を行った．

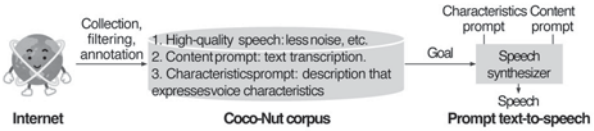


Fig.1：Overview of Coco-Nut corpus and Prompt TTS

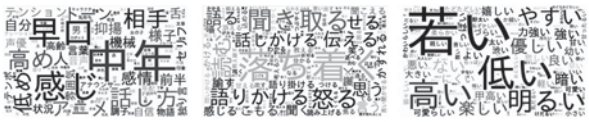


Fig.2：Word cloud of characteristics prompts (noun, verb, adjective)

3-9-4

3-9-4 複数話者テキスト音声合成における話者ベクトルの安定化に関する検討

A Study on Stabilization of Speaker Vectors in Multi-Speaker Text-to-Speech Synthesis

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

◆x-vector を話者埋め込み情報とした複数話者 TTS において同一話者であっても発話ごとの変動による影響があることを確認した。(日田ら，音学シンポジウム，2023)

◆話者埋め込み情報の次元削減したのち，代表ベクトルを用いた量子化による同一話者の発話ごとの変動に頑健な音声合成を検討した。

◆話者埋め込み情報として GST(Global Style Tokens)を使用した。

◆主成分分析により，次元数を 8、16、32 次元に削減した。

◆代表ベクトルは話者ごとの平均 GST を用いた場合と k-means 法を用いて生成したセントロイドを用いた場合の２つを検討した。

◆メルケプストラム歪みによる評価において，次元削減および代表ベクトルを用いたモデルは従来手法と比べ，分散の値が低いことが確認できた。

model	GST(256dim)	GST(8dim SA)	GST(8dim k)	GST(32dim SA)
average	6.77 ± 1.22	7.05 ± 1.11	7.07 ± 1.30	7.04 ± 1.16
dispersion	1.49	1.23	1.69	1.34

▲メルケプストラム歪みの平均と分散(SA:話者ごとの平均GST、kk-means 法のセントロイド)

model	GST(8dim SA)	GST(8dim k)	GST(32dim SA)
ratio(%)	23.7	25	27.7

▲．同一話者の発話の中で最も多く割り当てられた代表ベクトルの割合

3-9-5

3-9-5　55 スタイルを選択可能なテキスト音声合成のための制御インタフェースの設計

Design of a control interface for text-to-speech synthesis with 55 styles selectable

☆中邑草太，藤本崇人，橋本佳，南角吉彦，徳田恵一

◆テキスト音声合成
- ➢ 人間の肉声に近い音声品質の実現に成功
- ➢ 発話スタイルや抑揚の制御，感情表現など，表現力豊かな音声合成の研究が進められている
- ➢ 多様なスタイルを表現可能な音声合成システムの実現が課題

◆制御インタフェースの設計
- ➢ 発話スタイルを指定して音声合成を行うアプリケーションを作成
- ➢ 5 種類のスタイルを行う制御インタフェースを設計
- ➢ どのインタフェースがスタイル選択に使用しやすいかを評価
  - ✧ 実際にアプリケーションを使用し，文章にスタイルを選択
- ➢ 作成した文章についても MOS 試験を行い評価

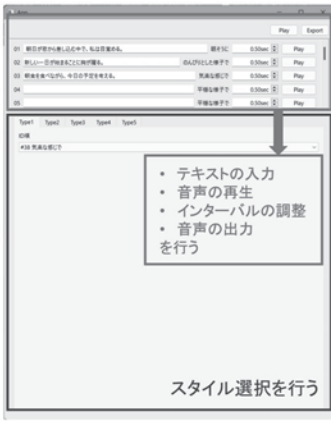


Fig1：Application

3-9-6

3-9-6　話者識別高速化に向けたクラスタリングの計算量削減

Computational Complexity Reduction for Clustering in Speaker Diarization

☆山下孔明，下倉良太，△飯國洋二（大阪大院・基礎工学研）

◆本講演では，話者数と話者情報を用いない，Speaker Diarization という話者識別技術を扱う。具体的には，クラスタリングベースのものを扱い，クラスタリング部分の計算量を減らす方法を提案する。

◆当該技術において，行列の固有値などを用いて定義される数列の最小値を，反復計算によって求める処理が現れる。提案手法では，当該数列が次の Fig.1 のように，初項付近で最小となる傾向があることに注目し，反復計算の回数を定数回に設定した。

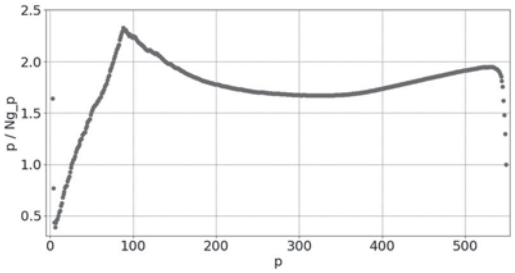


Fig.1 : Behavior of sequence discussed in our presentation

◆提案手法は従来手法と同等以上の精度を維持しながら，処理時間を数パーセントにまで削減した。

◆講演終盤では，現在着手している固有値計算における計算量削減の可能性について言及する。

3-9-7

3-9-7　感情を想定した発話スタイル依存型話者照合

Utterance-style-dependent Speaker Verification by Utilizing Emotions

☆高山響（静岡大），西田昌史（静岡大），柘植覚（大同大），黒岩眞吾（千葉大），西村雅史（愛知産業大，静岡大）

話者照合に関する研究では、リプレイ攻撃や音声合成、声質変換などのなりすまし手法に対する対策が急務となっている。本研究では、なりすまし音声に対する頑健な手法として、照合音声に感情音声を使用し、音声に含まれる話者の特性だけでなく感情も認証の鍵として用いる「感情を想定した発話スタイル依存型の話者照合手法」を新たに提案する。

本稿では、提案手法を x-vector を使用した感情の照合と話者の照合の２段階の照合システムとして実装し、感情音声コーパスである Japanese Twitter-based Emotional Speech corpus (JTES) を使用して４種の感情それぞれを認証の鍵とした時の評価実験を行った。実験結果 (Table 1) より、提案手法 (Proposed method) は感情情報を適切に利用できれば、感情の照合を行わない従来の話者照合 (Conventional method) よりも EER を大幅に削減できることがわかった。また、x-vector が話者照合だけでなく、感情の照合においても有用な特徴量であることを明らかにした。

Table 1 : EERs of conventional method without emotion verification and proposed method assuming that emotion verification is fully correct

	Anger	Joy	Sadness	Neutral	Average
Conventional Method	1.23	2.01	1.10	1.63	1.49
Proposed Method	0.60	1.13	0.78	0.93	0.86

3-9-8

3-9-8　自己教師あり学習モデルに基づくなりすまし攻撃が話者照合システムに与えるリスクの評価

Risk assessment of spoofing attacks using self-supervised learning models on speaker verification systems

☆伊藤葵（法政大・情報科/日立），堀口翔太（日立）

◆ SSL モデルを用いた話者照合システムに対するなりすまし攻撃検出器に対し，攻撃側も SSL モデルを用いた際の性能を評価した。

【防衛側】実発話 or なりすまし音声の二値分類

【攻撃側】SSL モデルを活用し，音声合成や声質変換によって生成されたなりすまし音声由来の話者特徴量が，実発話由来の話者特徴量と似るよう強化モデルを学習

⇒ 攻撃側も SSL モデルを用いることでなりすまし検出の性能が劣化

⇒ 攻撃側と防衛側が使う SSL モデルが同じ場合と異なる場合でなりすまし検出の性能劣化に大差はない

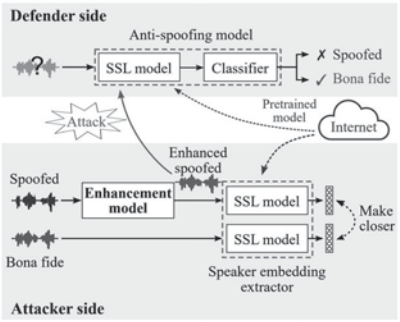


Fig.1 : Overview of evaluation flamework

3-9-9

3-9-9　発話交替頻度を導入した実対話音声に対する Speaker Diarization のためのデータ生成

Generating speaker diarization data for conversations by introducing frequency of speech alternation.

☆市川奎吾，上乃聖，李晃伸（名工大）

◆End-to-End Neural Diarization (EEND) モデルの学習データ生成法 Simulated Conversations 法に発話交替頻度の導入を提案する.

◆発話間の話者交替回数を制御することで，話者交替が激しい音声データへの EEND モデルの性能向上をねらう.

◆実験結果より SC 法と比較して Missed Speech の性能の改善がみられた.

Table 1: DERs (%) evaluated on the CALLHOME. DER consists of Missed Speech (MS), False Alarm (FA), Speaker Confusion (SC).

Dataset	MS	FA	SC	DER
SM method ($\beta = 5$)	4.2	7.9	1.3	13.5%
SC method	6.1	3.1	1.9	**11.1%**
Proposed method	**5.1**	4.8	2.5	12.5%

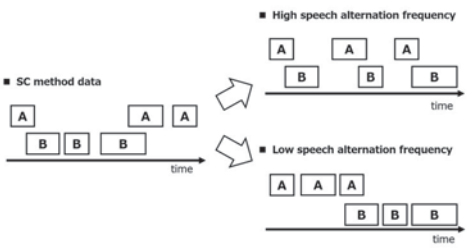


Fig.1 : Proposed method

3-9-10

3-9-10　感情ラベル列推定の知識蒸留を用いた音声感情認識

Speech Emotion Recognition
with Knowledge Distillation of Estimating Emotional Label Sequences

☆永瀬 亮太郎, 福森 隆寛, 山下 洋一（立命館大）

◆ **感情ラベル列推定の知識蒸留によるモデルの軽量化**
- wav2vec2.0 などの大規模モデルは感情ラベル列推定に**有効**
  - ✓ 一方でモデルサイズの増加と共に計算機コストも増加
- 知識蒸留を用いて教師モデルと**同等の性能**かつ教師モデルよりも**小規模なモデル**を構築

◆教師モデルに対する生徒モデルの圧縮率を変えモデルを学習・評価
- ３通りの圧縮率（**52.1%**，**28.1%**，**16.1%**）の生徒モデルを学習
- 通常の学習に加え知識蒸留をすることで教師モデルの性能に近い生徒モデルの構築を実現
  - ✓ （52.1%の場合）教師モデルと生徒モデルの認識率の差が**最小 0.2 ポイント**まで減少
  - ✓ （16.1%の場合）知識蒸留を利用しなかった場合の生徒モデルの認識率よりも**最大 6.5 ポイント**向上

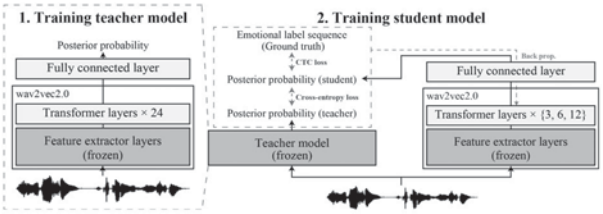


Fig.1 : Method of knowledge distillation for estimating emotional label sequence

3-10-1

3-10-1　アカゲザルの声帯−仮声帯物理モデルを用いた発声の不安定化に関する研究 −ヒトの言語進化に関する一考察−

Voice instability observed in a vocal-ventricular fold physical model of rhesus macaque (*Macaca mulatta*): An insight into evolution of human language

☆吉谷友紀, 上田可奈子, 小畑大樹（立命館大）,
△西村剛（京都大）, 徳田功（立命館大）

◆アカゲザルの摘出喉頭のマイクロ CT 画像から作製された物理モデルを使用し，仮声帯による発声の不安定化について調査した。

◆仮声帯の内転レベルが異なる物理モデル（±0〜−４（小さいほど内転強））を作製し，物理モデルに与える流量を 0〜100 L/min まで 20 秒間かけて増加させ，状態の変化を調べた。

◆各モデル条件において，全体の振動時間における（安定な）声帯振動，（安定な）声帯−仮声帯振動および不安定な振動の割合を調べた。

◆仮声帯の内転が強い条件で，不規則な振動が多く現れた。

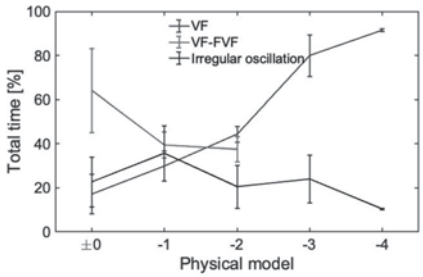


Fig.1 : Total oscillation time (vocal fold oscillation, vocal-ventricular fold co-oscillation and irregular oscillation) observed under different adduction condition.

3-10-2

3-10-2　声帯２質量モデルにおける直交２方向の剛性比が振動特性に及ぼす影響

Effects of horizontal-to-vertical stiffness ratio on the vibration characteristics of a two-mass vocal fold model.

☆黒澤大樹（豊橋技科大）, 吉永司（大阪大）, 飯田明由（豊橋技科大）

◆現在広く用いられている声帯 2 質量モデルでは実際の声帯と異なり鉛直方向の変形が固定されているため，それが音の発生と振動特性に及ぼす影響は詳しく調べられていない．

◆本研究では，質量の鉛直方向の移動を可能にし，ばねとダンパを接続した．同じ方向のばねの定数の比は一定とし，鉛直方向のばねの比を変化させ，その影響を調べた．

◆円筒の質量と接線により滑らかな声帯表面を表現し，気流は１次元の Bernoulli の定理，剥離位置は境界層理論により計算した．

◆基本周波数(f_0)は水平方向の剛性(k_{1y})増加に伴って増加し，鉛直方向の剛性(k_{1x})には影響されないが，k_{1x} のわずかな増加で f_0 が急激に減少する条件があることがわかった．その条件では振動モードが変化し，粘膜波動の様式に大きな影響を与えることがわかった．

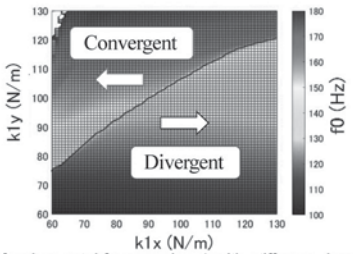


Fig.1 : Change in fundamental frequencies f_0 with stiffness changes. (Black line indicates the change of convergent and divergent shapes before closure.)

3-10-3

3-10-3　日本語標準語話者の正中面における直音と拗音の声道形状の比較分析

A Comparison of the median plane vocal tract shapes of palatalized and non-palatalized consonants in standard Japanese

☆藤澤流以, 竹本浩典（千葉工大）, 前川喜久雄（国語研）

◆われわれは日本語話者 22 名のデータを収録した「リアルタイム MRI 調音運動データベース（rtMRIDB）」を公開した。そしてデータベースの動画から調音器官の輪郭を抽出する研究を行ってきた。

◆本研究では，日本語標準語話者１名の声道形状を分析し，カ行とサ行の対立する直音と拗音の調音運動を分析した。

◆その結果，全ての拗音で子音が硬口蓋化していることを確認した（Fig.1）。また，/ka/-/kja/，/ko/-/kjo/，/sa/-/sja/の対立の拗音では子音から母音への移行タイミングが早く，移行する速度が緩やかであった。これらの運動はわたり音の生成を示唆する可能性がある（Fig.2）。一方，/so/-/sjo/の対立では，子音から母音への移行タイミングと速度に差は見られなかった。また，/ku/-/kju/，/su/-/sju/の対立はいずれも動きが小さく，差が明確でなかった。

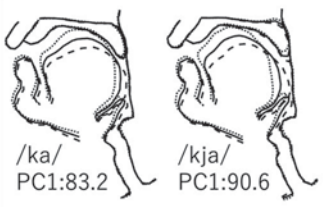


Fig.1 : /ka/と/kja/の PC1
実線：平均，破線：+1SD，点線：−1SD

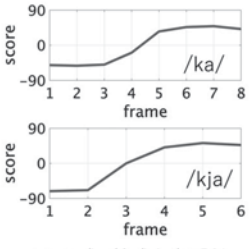


Fig.2 : /ka/と/kja/の PC1 スコアの時間変化

3-10-4

3-10-4　多話者におけるリアルタイム MRI 動画への歯列補填の精度分析

Accuracy of tooth superimposition on real-time MRI videos for multiple speakers

☆常盤朔也，竹本浩典（千葉工大），前川喜久雄（国語研）

◆前報でわれわれは 1 名の MRI 動画の各フレームに画像処理により歯列を補填する手法を開発して報告した。

◆本研究では 10 名の話者の MRI 動画の各フレーム（総フレーム数：289,304 枚）へ歯列を補填してその精度を検討した。

◆歯列を抽出したフレームで歯列を上下左右に 1 ピクセル平行移動，あるいは±1 度回転させて歯列がはみ出した割合（誤差面積率）の最大値を参考値とし，全フレームで誤差面積率が参考値を超えた割合を参考値超過率として精度を評価した。

◆その結果，参考値超過率が下顎では全ての話者で 5%未満であったが，上顎では 4 名の話者で 5%を超過した。

◆位置合わせの矩形領域を拡大し，下顎と同様に上顎でも動きに追従してフレームごとに矩形領域の位置を移動して補填した結果，全ての話者で参考値超過率が 5%未満に減少した。

Table1：上顎の歯列補填の精度

	s1	s4	s5	s7	s11
平均値	5.9	3.8	2.5	9.2	5.1
最大値	30.4	27.3	16.5	50.4	26.2
参考値	15.7	8.9	10.3	18.5	11.9
総フレーム数	28,160	27,140	30,727	30,720	29,188
参考値超過率	2.0	3.3	1.4	4.0	4.7
	s12	s17	s18	s26	s27
平均値	9.8	3.4	10.7	4.4	10.1
最大値	33.5	40.9	65.8	26.2	40.9
参考値	21.8	13.5	21.9	14.0	26.2
総フレーム数	29,701	28,676	30,208	27,648	27,136
参考値超過率	2.9	2.8	3.6	0.8	0.7

Table2：下顎の歯列補填の精度

	s1	s4	s5	s7	s11
平均値	21.4	6.8	2.5	20.6	11.3
最大値	50.0	21.1	17.9	42.5	31.3
参考値	39.6	24.7	28.8	35.1	30.2
総フレーム数	28,160	27,140	30,727	30,720	29,188
参考値超過率	0.4	0.0	0.0	0.1	0.0
	s12	s17	s18	s26	s27
平均値	11.9	9.9	12.4	9.2	7.5
最大値	59.9	33.2	45.0	26.5	27.8
参考値	47.7	37.5	36.5	35.5	45.1
総フレーム数	29,701	28,676	30,208	27,648	27,136
参考値超過率	0.1	0.0	0.2	0.0	0.0

3-10-5

3-10-5　性別による話速の違い —日本語日常会話コーパス（CEJC）を用いた分析—

Gender Differences in Speech Rate — An Analysis using the Corpus of Everyday Japanese Conversation (CEJC) —

○梁　辰，倉片憲治（早稲田大・人間科学）

◆話速は，音声の聞き取り易さに影響を及ぼす大きな要因の一つと考えられる。しかし，これまで日本語の日常会話の話速を大規模データに基づいて調査した例は知られていない。

◆本研究では，2022 年 3 月に公開された国立国語研究所「日本語日常会話コーパス」（CEJC）を使用して，日常会話の話速の計測と話速に及ぼす種々の要因の分析を行った。

◆分析結果の概要は，以下の通りである（Fig.1）。

① 話速は，発話区分の長さ（拍数）によって異なる。② 女性よりも男性の方が，全体的に話速は速い。ただし，20 代では男女差は小さい。③ 女性は年齢が高いと話速が遅い傾向があるのに対して，男性では年齢による話速の違いは小さい。

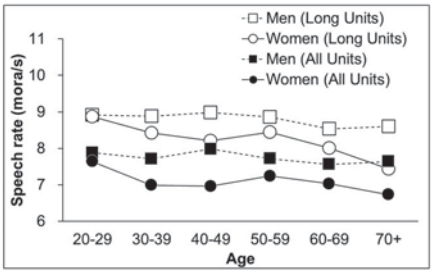


Fig.1: Speech Rate across Unit Length, Gender and Age.

3-10-6

3-10-6　WAVE による話速を変化させた時の舌運動の観測

Observation of tongue movements while changing speech rate by using the WAVE system

☆関純香，△下条未来，△武井俊輔，竹本浩典（千葉工大），能田由紀子，前川喜久雄（国語研）

◆本研究では，4 名の話者が自身にとって普通（normal），速い（fast），遅い（slow）話速で 2 つの朗読文「北風」（195 モーラ）と「宇宙（235 モーラ）を読み上げたときの舌運動を検討した。

◆舌の正中矢状面に約 1 cm の間隔をあけて前から T1，T2，T3 の 3 つの WAVE センサを設置し，100 Hz サンプリングで位置を記録した。

◆Table 1 で示す実発話時間と各センサの総移動距離を検討した結果，実発話時間に比べ総移動距離の増減率の方が小さかった。

◆すなわち，話速の変化は舌の移動範囲を縮小あるいは拡張するより，舌の移動速度を速くあるいは遅くすることで実現していた。

Table 1 実発話時間と各センサの総移動距離の変化量（%）
実発話時間：休止を除いた発話時間，**f-n**：普通より早く話すとき，**s-n**：普通より遅く話すとき

北風		変化量(%)				宇宙		変化量(%)			
		T1	T2	T3	実発話時間			T1	T2	T3	実発話時間
M1	f-n	23.4	20.0	25.3	−2.6	M1	f-n	1.2	7.6	3.2	−6.4
	s-n	21.9	29.1	28.4	8.6		s-n	−4.0	−7.9	0.5	10.4
M2	f-n	3.1	−4.8	−1.0	−8.0	M2	f-n	6.9	8.3	8.6	−9.5
	s-n	18.1	17.2	21.0	14.6		s-n	−5.0	−3.7	−1.0	8.5
F1	f-n	3.2	8.3	7.9	−4.4	F1	f-n	5.6	8.6	9.4	−10.4
	s-n	3.3	7.3	8.4	11.4		s-n	−1.9	6.9	6.8	4.6
F2	f-n	1.3	3.5	−2.7	−8.8	F2	f-n	15.5	17.3	4.9	−11.6
	s-n	3.1	11.7	6.7	8.9		s-n	15.2	31.8	26.5	19.6

3-10-7

3-10-7　声道立体形状の分析に基づく歌い手のフォルマントの生成メカニズムの検討

Examination of mechanism for generating singer's formant based on three-dimensional acoustic analysis of the vocal tract

☆志々目樹，△相葉まこ，戸田菜月，竹本浩典（千葉工大），高橋純（大阪芸大）

◆男性オペラ歌手の歌声のスペクトルには歌い手のフォルマント（SF）が見られるが，その生成メカニズムは明らかになっていない。

◆本研究ではバスバリトン 1 名の声道形状とその音響特性を比較・分析し，「SF は主声道の共鳴に由来する第 3 フォルマント，喉頭腔共鳴に由来する第 4 フォルマント（F4），梨状窩に由来する極が近接したものである」という仮説を検証した。

◆先行研究では喉頭腔を除去すると F4 のみが消失するが，本研究では F4 のレベルは約 14 dB 低下したが消失はしなかった（Fig. 1）。

◆歌唱発声時の声道モデルと喉頭腔を除去した声道モデル（主声道モデル）の F4 の瞬時音圧分布を確認したところ，共に左右の梨状窩が逆相であった。これにより，梨状窩による極零対の極に喉頭腔由来のフォルマントが完全に重畳していることが明らかになり，少なくともこの歌手では仮説を裏付ける結果が得られた。

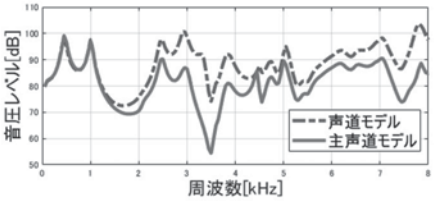


Fig.1：喉頭腔除去が伝達関数に与える影響

3-10-8

3-10-8　自己の録音音声の魅力と親近感の評価に個人特性が与える影響の分析

Influence of Personal Traits on Attractiveness and Familiarity of One's Own Voice

◎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研究ごとに異なる結果が示されている.

- ポジティブな反応：自己音声は他己音声より魅力的 [Peng+, 2020]
- ネガティブな反応：自己音声に違和感を覚える [Holzman+, 1966]
- 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いる[Peng+, 2020]が，先行研究では著者が定義した一つの個人特性のみを分析に用いており，どの個人特性がどの程度影響するか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（魅力，親近感）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.

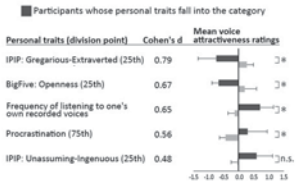


Fig.1：Top five effect sizes for attractiveness ratings

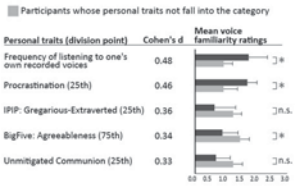


Fig.2：Top five effect sizes for familiarity ratings

3-11-1

3-11-1　変調フィルタバンクを用いた感情音声知覚に影響を与える変調周波数帯域の分析

Analysis of the Effect of Modulation Frequency Bands on Vocal Emotion Perception Using Modulation Filterbank

☆郭太陽，木谷俊介，鵜木祐史(北陸先端大)

◆背景：音声の振幅包絡線情報は感情知覚に重要な役割を果たしている. しかし，振幅包絡線のどの周波数帯域（変調周波数帯域）に，感情知覚に深くかかわる情報が含まれているか，明らかになっていない.

◆目的：感情知覚に寄与する変調周波数帯域を明らかにすること.

◆方法：変調フィルタバンク（MFB）を利用して変調周波数帯域を制限した雑音駆動音声の感情認識率を比較：(1) 重要な変調周波数帯域の上限と下限を検討，(2) [上限,下限]の帯域の重要性を検討，(3) 隣接する帯域の重要性も検討，(4) 重要な変調周波数帯域を絞り込み.

◆結果：(1) 重要な変調周波数帯域の上限が 8 Hz，下限が 2 Hz. (2) [2, 8] Hz の帯域だけでは不十分. (3) [0, 2] Hz と [8, 16] Hz の帯域も必要.

◆結論：(4) 感情知覚に重要な変調周波数帯域は [0, 16] Hz である.

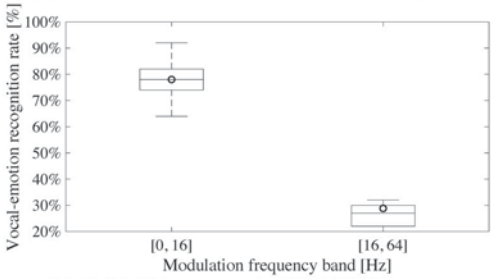


Fig.1：変調周波数帯域 [0, 16] Hzと変調周波数帯域 [16, 64] Hzに限定された振幅包絡線をもつ雑音駆動音声の感情知覚の結果

3-11-2

3-11-2　Interrupted mosaic speech revisited: Gain and loss of stretching on intelligibility

○ Kazuo Ueda, △Masashi Hashimoto (Kyushu Univ.), Hiroshige Takeichi (R-IH, RIKEN), and Kohei Wakamiya (Kyushu Univ.)

Periodicity and temporal fine structure provide a perceptual cue to connect speech fragments when the speech is interrupted. In mosaic speech stimuli, which are noise-vocoded with stepwise amplitude and spectral envelopes, interruption drastically reduces intelligibility because of lacking the perceptual cue. Here we report that the intelligibility of mosaic speech stimuli, consisting of 20 frequency bands and 20-ms segment duration, decreased from 95% to 78% and 33% when they were interrupted by 20- and 80-ms gaps. However, when the mosaic segments were stretched to fill the silent gaps, the intelligibility improved again to 92% and 54%, resulting in gains of 14% and 21% respectively. By contrast, intelligibility dropped by 7% with stretching and filling 160-ms gaps. The intelligibility minimums (9%-16%) were observed at the 160-ms original gap duration irrespective of stretching ratios. It is possible that stretching with short gaps effectively restored the original amplitude envelopes, whereas, with long gaps, the effect was detrimental. Furthermore, a probability summation model with a short and a long integration window, by which the intelligibility minimums were predicted for lowpass-filtered interrupted speech stimuli, seems to be highly relevant to the current results.

3-11-3

3-11-3　Selective Listening in Checkerboard and Interrupted Speech Stimuli with Two Talkers

☆Jun HASEGAWA, Kazuo UEDA (Kyushu Univ.), Hiroshige TAKEICHI (Riken), Gerard B. REMIJN, and Emi HASUO (Kyushu Univ.)

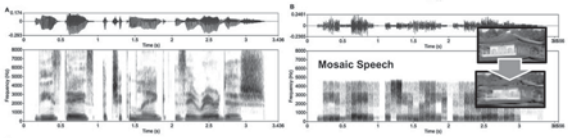
A two-talker situation is tougher than a single-talker situation when one has to selectively listen to a target speech. To simplify experimental conditions, we focused on the situation when two talkers speak different sentences alternating with each other, and introduced two types of alternations: only in time (two-talker staggered speech) and in time and frequency (two-talker checkerboard speech). To our knowledge, the intelligibility of two-talker checkerboard speech stimuli had never been investigated. Under three steps of segment duration (20, 80, and 320 ms) for both stimulus types and three steps of the number of frequency bands (4, 8, and 20) for checkerboard-type speech stimuli, native Japanese listeners (N = 22; age range: 21-34) successfully heard out a specific talker's sentence while ignoring the other for both stimulus types. The intelligibility curves of two-talker checkerboard speech stimuli exhibited characteristic U-shaped curves as previously reported with single-talker checkerboard speech stimuli (with 5-20% drop in intelligibility), suggesting a common auditory grouping mechanism works in both single-talker and two-talker cases. Or otherwise, the U-shaped curves may emerge from probability summation of two antagonistic processes in speech perception.

3-11-4

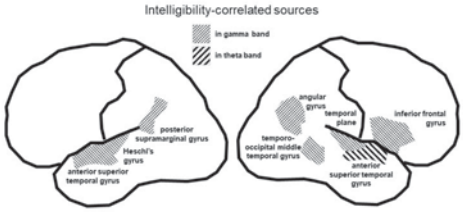
3-11-4　Magnetoencephalography elucidates mosaic speech processing

○Hiroshige Takeichi (R-IH, RIKEN), Kazuo Ueda (Kyushu Univ.), Takako Mitsudo (Hizen Psych. Center), △Alexandra Wolf (AIP, RIKEN), △Yoji Hirano (Univ. Miyazaki)

◆Magnetoencephalograms were acquired while 17 volunteers passively listened to filtered, noise-vocoded, and mosaicked speech stimuli with varying segment duration and the number and type of stimulus frequency bands, i.e., varying intelligibility.



◆The intelligibility-correlated brain activity was found in the theta and gamma but not in the alpha and beta MEG bands, in bilateral perisylvian regions.



◆Results support the hypothesized two temporal windows in speech processing.

◆This study is supported by KAKENHI JP19H00630.

3-11-5

3-11-5　言語情報の程度が音列の主観的速度に及ぼす影響

The effect of linguistic information on subjective speed of sound sequence

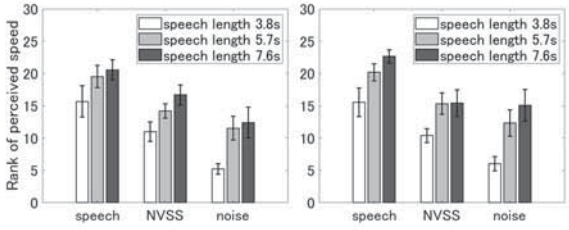
☆玉川真帆，坂本修一（東北大通研/院情科研）

◆本研究では，音声の聞き取りに関連する言語理解と話速の知覚の関係を検討するべく，了解性の異なる刺激３種（音声，劣化雑音音声，ノイズ列）に対する主観的速度を調査した．

◆聴取者は，刺激３種 × 音声長３段階 × 音声長におけるポーズ長の割合３段階の２７刺激に対し，知覚された速度の順位を回答した．

◆実験の結果，刺激の種類によって主観的速度に差が生じることが明らかとなり，刺激の了解性および内容理解が話速の知覚に影響を与えることが示唆された．また刺激に対する了解性が変化した場合にも，音声長は変わらず主観的速度に影響を及ぼすことが示された．

◆一方で，主観的速度に対するポーズ比の影響は見られなかった．これは，好まれるポーズ長が刺激ごとに異なったことが原因で，ポーズ比の影響が一貫して得られなかったためであると推測される．



(a) pause length ratio：0.211　(b) pause length ratio：0.368

Fig.1：Ranks of perceived speed

3-P-1

3-P-1　ゲームにおける近接空間の音場の再現に向けた再生システムの検討

Near-Field compensated sound field simulation system for games

○小島健二，△小島拓朗（カプコン）

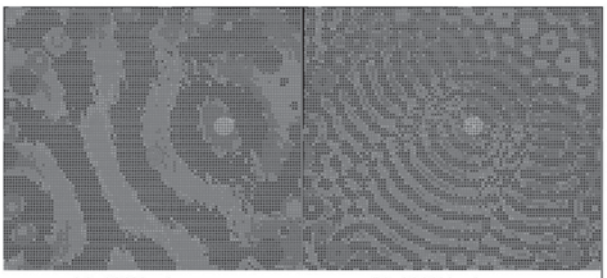
◆リアルタイム処理が求められるゲーム内の仮想空間で近距離音場を再現するために，NFC-HOA を利用して軽量に動作する再生システムを検討した．

◆NFC-HOA の方向成分の計算においてルジャンドル多項式を利用してエンコードデコードの計算処理を軽量化した．

◆動径成分の計算において，距離によらず決定できる変数を事前計算することを試みた．

◆ゲームエンジンで NFC-HOA による聴取者周辺の近距離音場を可視化した．

◆今後，低域が持ち上がったり，動径成分のフィルタが不安定となる課題を解決し，実際のゲームで運用できるかを検討する．



(a). 600Hz sine wave　(b). 2000Hz sine wave

Fig.1：Sine wave sound field around the listener using a game engine

3-P-2

3-P-2　Context-Aware なニューラルネットワークによる残響付き音声からの擬似 RIR 生成

Virtual Room Impulse Response Estimation from Reverberant Speech with Context-Aware Neural Network

☆宮田涼司，羽田陽一（電通大）

◆目標とする環境の空間特性を既に録音されている音信号から取得する方法として，室内インパルス応答 (RIR) の推定が挙げられる．

◆RIR のふるまいを直接推定することは難しいため，畳み込まれた音声が聴感上区別できないような表現として，擬似 RIR の生成を検討した．

◆統計的 RIR モデルとディープニューラルネットワーク（DNN）を組み合わせた学習的 RIR モデルによって，残響付き音声から，対応する擬似 RIR を生成することを検討した．

◆実験の結果，提案手法は従来手法に比べ残響時間と波形の誤差が少ない擬似 RIR を生成できることが確認できた．

◆また，統計的 RIR モデルと DNN をそれぞれ独立して利用する場合に比べ，それらを組み合わせた場合にも残響時間と波形の誤差が少ない擬似 RIR を生成できることが確認できた．

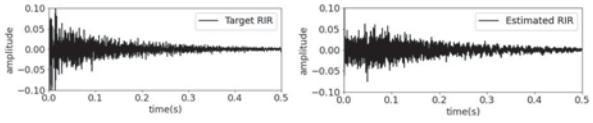


Fig 1: Target RIR waveform　Fig 2: Estimated RIR waveform

3-P-3

3-P-3　等価音源法に基づいた制御領域可変型 Pressure Matching 法による音場制御

Sound field reproduction with control-area-variable pressure matching based on equivalent source method

☆松本綾太，津國和泉，鈴木薫佳，池田雄介（東京電機大）

◆背景

- Pressure Matching 法 (PM 法) はスピーカから制御点までの伝達関数の測定が必要で，制御点やスピーカ数の増加に伴い，実装が困難

◆提案

- 等価音源法を用いた PM 法により，制御点を自由に配置し，周波数ごとに制御点と制御範囲を変更した音場再現手法の提案

◆実験

- 実測実験では，　SNR20dB 基準で制御半径を動的に設定
- 結果から，より広範囲の音場制御が可能となることがわかった。

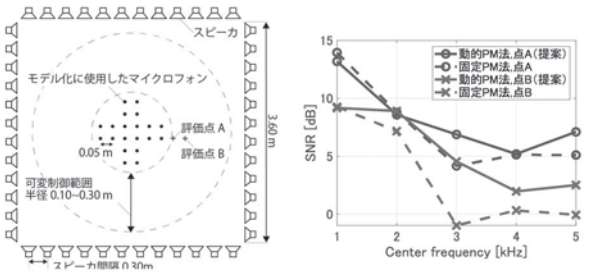


Fig.1：Experimental condition and comparison of SNRs for frequency bands

3-P-4

3-P-4　スパース等価音源法に基づく反射音を利用した三次元音場合成

Three-dimensional sound field synthesis using reflections based on sparse equivalent source method

☆小畑悠矢，鈴木薫佳，津國和泉，池田雄介（東京電機大）

◆三次元音場合成は多数のスピーカを三次元的に配置する必要

◆少数の測定情報を用いて直接音と壁面からの反射音をスパース等価音源法でモデル化し，半円状の３層スピーカアレイを用いた音場合成手法を提案

◆シミュレーション実験では，方位角 $0° \leq \phi \leq 90°$ からの平面波を合成し，その精度を反射音の有無とスピーカ数の違いで評価

◆実験の結果，$\phi = 0°$ から 20° までは反射音の有無によって SNR の差は小さいが，$\phi = 30°$ 以降では反射音の利用で改善され，スピーカが配置されていない位置からの音場を合成できることが示された

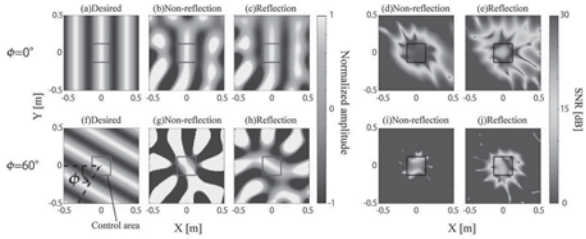


Fig.1：Comparison of sound fields and SNRs using 48-ch loudspeaker array (Frequency:1kHz)

3-P-5

3-P-5　スパース等価音源法に基づいた少数マイクロホンを用いた散乱音場の可視化

Visualization of scattered sound field using small number of microphones based on sparse equivalent source method

☆鬼澤諒介，津國和泉，池田雄介（東京電機大）

◆複雑な音の伝搬を理解するには音場の可視化が有効

◆少数の測定情報を用いて直接音と散乱音場を分解して音場を可視化することを目的とし，スパース等価音源法を用いた入射音場と散乱音場の可視化手法を提案

◆シミュレーション実験では、物体に対する入射音場と散乱音場を円形マイクロホンアレイの情報から推定し、その精度を評価

◆実験の結果，入射音場と散乱音場の両方で非常に高い推定精度を確認

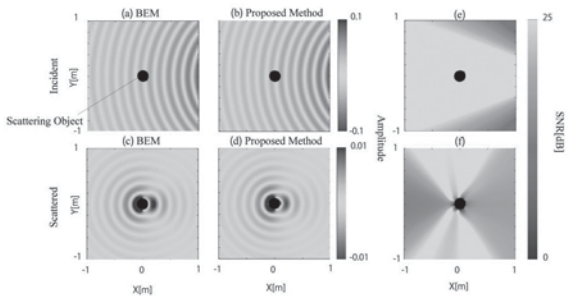


Fig.1：Comparison of Proposed method and BEM (Frequency:1kHz)

3-P-6

3-P-6　サイドローブ抑圧を適用したビームフォーミングフィルタの周波数特性に関する検討

Investigation on Frequency Response of Beamforming Filters with Sidelobe Suppression

☆大屋敷凌佑，渡邉貫治，西口正之，安倍幸治（秋田県立大），

◆音空間合成法を実現すべく，遅延和法によるビームフォーミング(BF)を基にした空間収音・再生システムを構築してきた。

◆前報[1]ではビームを向ける方向に合わせ，目標信号を変化させる「可変窓」の考案により，従来法と比較して約 10 dB 程度のサイドローブ抑圧を達成した。

◆一方，可変窓に基づき算出された BF フィルタは，低域ゲインが増大することが分かった (Fig. 1)。　[1] 大屋敷ら，音講論, pp. 489–492 (2022.9)

◆正則化処理を行った結果，大きな性能低下を起こすことなく，フィルタゲインを抑圧することに成功した (Fig. 2)。

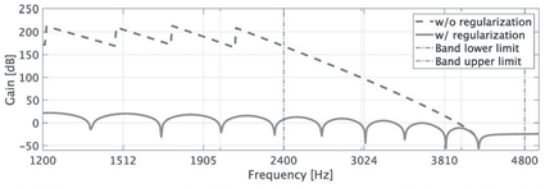


Fig.1：Frequency charactaristics of BF filter (beam = 0°, mic ch = 16)

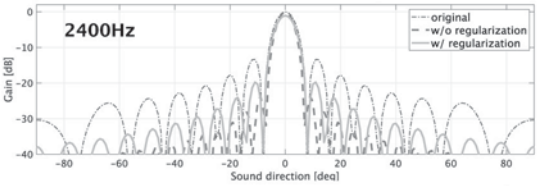


Fig.2：Beam pattern for ideal array observation signal (beam = 0°)

3-P-7

3-P-7　レイリー型 RBM を用いた独立低ランク行列分析に基づくブラインド音源分離

Blind Source Separation based on Independent Low-Rank Matrix Analysis using Rayleigh Restricted Boltzmann Machines.

☆古田翔太郎，中鹿亘（電通大）

◆ブラインド音源分離とは，音源位置やマイク位置，部屋の形状等を含む混合系が未知の条件で，観測信号のみから混合前の音源信号を推定する音響信号処理技術である。

◆既存手法の一つである独立低ランク行列分析は，空間モデルと音源モデルの観点からを考えられている。また，音源モデルには生成モデルである制限ボルツマンマシン(RBM)を用いることができる。

◆既存手法 GB-RBM における音源分離実験では，多くの場合で分離精度を向上することができたが，コスト関数の不一致やハイパーパラメータにおける不安定さが目立っていた。

◆上記を改善するために，生成モデルとして**原点対称複素ガウス分布**を仮定する。このとき，振幅を**レイリー分布**，位相を一様分布と見なせるので，音源モデルにRayleigh-RBM を用いた新たな手法を提案する。

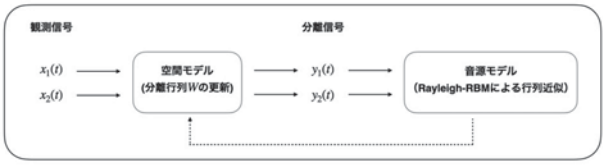


Fig.1 : Overview of the proposed method

3-P-8

3-P-8　１スピーカで近傍再生を実現する逆相波誘導エンクロージャの提案

Proposal for an Opposite-Phase Sound Wave Induction Enclosure for Near-Field Reproduction with a Single Loudspeaker.

○加古達也，千葉大将，野口賢一（NTT）

◆　スピーカの背圧を利用した逆相波誘導エンクロージャを利用することで，１スピーカで近傍再生の実現を目指す

- ➢ 音漏れを抑えるオープンイヤー向けエンクロージャを発展
- ➢ より大型のスピーカで近傍再生を目指す

◆　逆相波誘導エンクロージャ

- ➢ スピーカ背面の逆相波を近傍再生に利用するには，ヘルムホルツ共鳴の共振周波数を高く設計する必要
- ➢ エンクロージャの容積およびポート長を小さくし，逆相波を前面に誘導する距離差を小さくする設計を提案

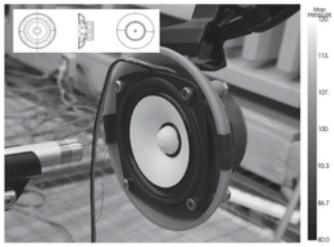


Fig.1 Proposed model of opposite phase sound induction enclosure.

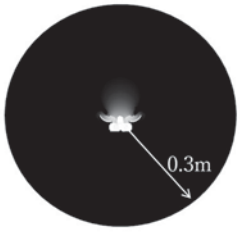


Fig.2 Sound pressure distribution in proposed model by FEM.

3-P-9

3-P-9　ボリュメトリック音声制作に向けた収音精度の検証

Study of sound recording for volumetric audio production

○木下光太郎，松井健太郎（NHK）

◆被写体を取り囲むように配置した複数台のカメラから同期取得した映像に3次元再構成技術を用いて3次元モデルを作り出す**ボリュメトリックキャプチャ**の研究・開発が盛んとなっている。

◆CG と異なり実写ベースの３次元モデルを作り出せるため，AR や VR などのコンテンツなどの臨場感・没入感の向上が期待できる。

◆NHK でもボリュメトリックキャプチャスタジオとして**メタスタジオ**の開発・整備を進めている(Fig.1)。

◆ボリュメトリックキャプチャを活用した任意視点映像に対応する音声として**ボリュメトリック音声制作**を提案する。収音・編集・再生の制作フローについて検討中である。

◆収音環境の構築に向けた予備実験として，無響室と残響可変室で放射特性の測定を行った。

◆残響可変室では外乱の影響による放射特性の変化が確認された。

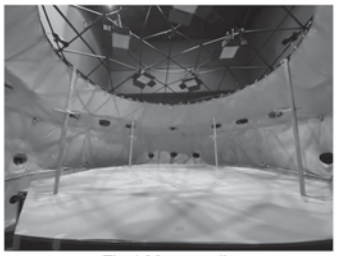


Fig.1 Meta studio

3-P-10

3-P-10　ボリュメトリック音声信号再生のためのバイノーラルレンダリング法の検討

A study on binaural rendering method for volumetric audio signal reproduction

○松井健太郎（NHK）

◆ NHK では，ボリュメトリックキャプチャスタジオ設備として，「メタスタジオ」の研究・開発を進めている（Fig. 1）。

◆ この映像のボリュメトリックキャプチャに対応する音声制作技術としてボリュメトリック音声制作を提案し，提案する制作フローのうち，ヘッドホンによるバイノーラルレンダリング法を概説する。

◆ メタスタジオには，ステージを取り囲むようにフレームが配置されている。このフレームにマイクロホンアレイをマウントし，音源オブジェクトより放射される音波を方向別に取得する。

◆ Figure 2 に示すように，この音波により形成される音場を外挿し，任意の聴取位置に入来する音波により形成される音場を推定する。

◆ 聴取位置近傍の音場から，頭部伝達関数を用いて左右耳に提示するバイノーラル信号を算出する。

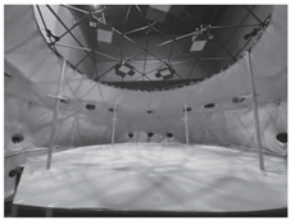


Fig. 1 : Interior view of the meta studio

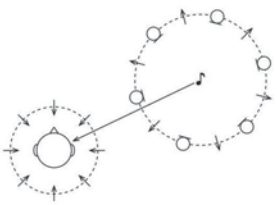


Fig. 2 : Binaural rendering

3-P-11

3-P-11　ボリュメトリック音声制作に向けたラベリアマイクロホン抗ノイズ化の検討

A study on the noise reduction of lavalier microphones for volumetric sound production

○小野一穂(NHK 財団)
松井健太郎, 中山靖茂, 杉本岳大, 木下光太郎(NHK)

◆ボリュメトリック音声制作において、映像へのマイクロホンの映り込みを避けて演者の歌唱などをクリアに収音するため、演者の服装等に仕込んだラベリアマイクロホンの使用を検討中である。
◆仕込んだラベリアマイクロホンは、演者の動きによる加振やこすれによる雑音が生ずるため、その低減に向けた基礎検討を行った。
◆ラベリアマイクロホンのケーブルに着目し、①外部の振動を伝えず表面の摩擦（こすれ）を減らす構造のケーブル、②いったんケーブルに乗った振動やこすれがカプセルに伝わるのを遮断する構造のケーブル、を有するマイクロホンを、それぞれ複数種試作した。
◆試作したマイクロホンのケーブルを加振する実験とこする実験を行い、生ずる雑音の特性を測定した結果、手法①はこすれに対して、手法②は加振・こすれの両方に対して雑音低減効果が見られた。

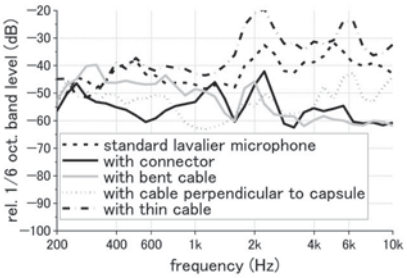


Fig. 1　Frequency response of noise due to rubbing

3-P-12

3-P-12　ボリュメトリック音声再生のための直線・球面アレイ協調型音場合成法

Linear and spherical arrays cooperative sound field synthesis for volumetric audio reproduction

○佐々木陽, 中山靖茂(NHK)

◆NHK では，これまでに直線アレイスピーカによる音場合成法の検討を行ってきた。
◆直線アレイスピーカでは，アレイの長さを超えた位置に仮想音源が配置された場合，正面の視聴エリアで音場を合成することが困難となる。
◆本稿では，直線アレイスピーカと球面アレイスピーカを分散配置し，それらを協調させた音場合成法について提案する。
◆直線アレイスピーカと四つの球面アレイスピーカにより，球面波音場と平面波音場を合成する２種類の数値シミュレーションを行った。
◆その結果，直線アレイスピーカのみでは合成が困難な音場についても，球面アレイスピーカを用いて合成音場を補完できることが確認できた。

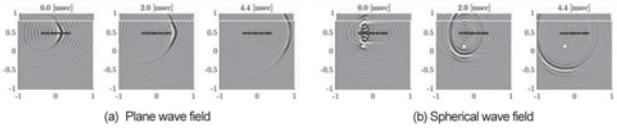


Fig.1 : Synthesized wave field with a linear loudspeaker array

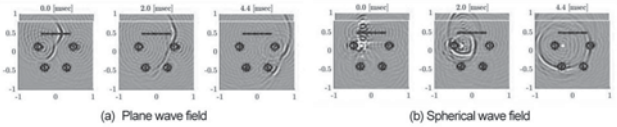


Fig.2 : Synthesized wave field with a linear loudspeaker array and multiple spherical loudspeaker arrays

3-P-17

音響差分説明文生成のための後半層表現の解きほぐし

High-level feature disentanglement for audio difference captioning.

◎竹内大起, 大石康智, 仁泉大輔, 原田登, 柏野邦夫 (NTT)

◆音響説明文生成：音響信号からその説明文を生成するタスク
- ➢ 特徴表現の類似部分と相違部分への解きほぐしが有効
- ➢ 仮説：音の違いの説明は比較対象に応じて変化
  - ✧ 特徴表現の類似部分と相違部分も比較対象に応じて変化

◆本研究では、比較対象を考慮した特徴表現解きほぐしを提案
- ➢ 音響埋め込み (前半層) ではなく音響差分エンコーダ (後半層) の出力に特徴表現解きほぐしを適用
  - ✧ 比較対象を参照した後に２つの入力音の特徴表現を類似部分と相違部分に解きほぐす学習

◆実験では、提案手法が説明文生成の精度を改善
- ➢ Attention weight の可視化を行い、比較対象に応じた特徴抽出を提案手法が誘導していることを確認

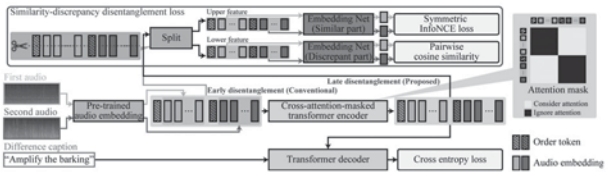


Fig.1 Illustration of proposed method

3-P-18

3-P-18　入力信号のモデル化を促す表現学習マスクモデリングデュオの提案

Masked Modeling Duo: Learning Representations by Encouraging Both Networks to Model the Input

○仁泉大輔, 原田登, 竹内大起, 大石康智, 柏野邦夫 (NTT)

◆自己教師あり学習 Masked Autoencoders (MAE)は画像だけでなく音響信号にも効果的であることを示した一方，再構成誤差に対して損失を計算するため，表現の学習は間接的であると考えられる。
◆本研究では，表現の予測誤差を利用して学習を行う自己教師あり学習 Masked Modeling Duo (M2D)を提案する。M2D は構成する２つのネットワークに対して入力のモデル化を促すことで，より効果的な表現学習を実現する。
◆実験では音響信号・画像入力を利用して，従来法との比較により提案法の有効性を確認した。

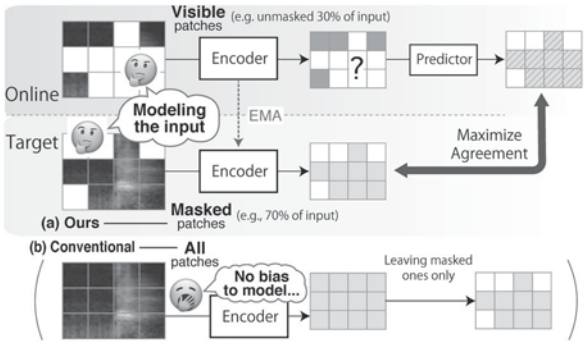


Fig. 1 MAE pre-training scenario.

3-P-19

3-P-19　DAGMM を用いた生活音の異常検知システムに関する検討

Using life sounds by DAGMM Anomaly detection system

○青石涼大(秋田県大院)，
安倍幸治，渡邉貫治，西口正之(秋田県大)

◆音を活用した見守りシステムにおいて，GMM を使用した異常検知システムを開発してきたが，使用する音響特徴量を手動で入力しなければならないという課題があった。これを解決する為のシステムとして DAGMM を使用し，入力音から自動で特徴量を選定し，検知するシステムを作成し，先行研究のオートエンコーダ(AE)を使用したシステムと比較し，評価実験を行う。

- 評価用日常音において DAGMM に基づく開発システムの方が AE より多くの音を異常として検知し，誤検知が多いシステムとなった
- 評価用異常音においても提案システム（DAGMM）の方が AE より異常検出性能が低くなった.

◆DAGMM のパラメータに問題がある可能性について引き続き検討する

	DAGMM	AE
評価用日常音 (84375フレーム)	1098	219
評価用異常音 (150フレーム)	99	109

Fig. 1 評価実験の結果

3-P-20

3-P-20　32 bit float 録音による打球音データセットの構築

Construction of a hitting sound dataset using 32-bit float recording

☆山田航輔，鈴木千文，西野隆典(名城大)

◆野球における打球の弾道，方向，飛距離及び打球音はボールとバットが接触する場所によって変化するため，打球音とこれらには関連があると考えられる。

◆打球音と打球の関係の調査のため，32 bit float 録音を利用し，打球音データセットの構築を行った。

◆計 523 の打球音を収録し分析を行った。

◆スイング速度と打球速度の散布図からは関係が見いだせなかったが，SVM を用いた打球(フライ，ライナー，ゴロ)の判別では 40 %程度の平均正答率であった。

◆今後の課題として，データ数が不足しているため，より多くの打球音の収録を行うことが挙げられる。

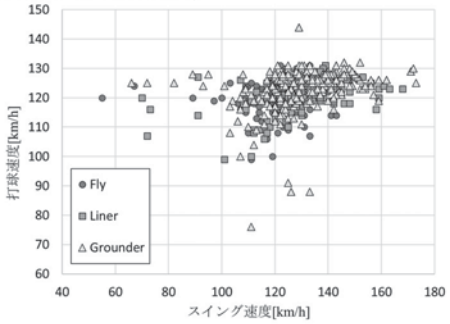


Fig. 1 Scatter plot based on swing and ball speeds.

3-P-21

3-P-21　MRI 騒音下における被験者音声検出のための VAE による生成データの合成法

Composite of generated data by VAE for subject voice detection in MRI noise

☆若松奏，武藤憲司(芝浦工大)，小林洋介(室蘭工大)

◆MRI 被験者の発話音声を検出して医療スタッフに伝えるシステムによって，コミュニケーションを可能にすることが目的である．これまでに，我々は様々な MRI 撮像音に対応するために VAE(Variational autoencoder)による生成データを用いた継続学習を検討した．しかし，MRI 撮像音を重畳した被験者音声を VAE に学習させると音声の特徴が欠落するという課題があった.

◆本報告では，VAE で生成した MRI 撮像音のスペクトログラム画像に被験者音声のスペクトログラム画像を合成したデータを用いてマルチコンディション学習を行った.

◆MRI 撮像音を重畳した被験者音声を VAE で直接生成する手法と比べて，提案手法は 0.214pt だけ F 値が高かった.

Table. 1 F-measure of each step by voice detection model

		Step1	Step2	Step3
Proposed method	EPI noise	0.973	0.971	0.958
	T1W noise		0.695	0.730
	T2W noise			0.873
	Average	0.973	0.833	0.854
Conventional method	EPI noise	0.860	0.782	0.778
	T1W noise		0.683	0.574
	T2W noise			0.569
	Average	0.860	0.733	0.640

3-P-22

3-P-22　異常音検知タスクにおけるモデルの軽量化

Model weight reduction in abnormal sound detection task

☆加藤洸人，立蔵洋介(静岡大)

◆研究目的

大規模な AI モデルの精度をできるだけ維持しつつ軽量化することで，エッジデバイスへ搭載してオフライン運用することによるシステムの安定化と推論における電力消費量の削減を目的とする.

◆提案手法

知識の蒸留により，モデルサイズの大きな教師モデル L-AE の性能に小さな生徒モデル S-AE の性能を近づける．S-AE の学習時に L-AE と S-AE のボトルネック層の MSE を損失に加えることで蒸留を行う.

◆実験

全 7 機種の正常な動作音のみを学習させ，動作音の正常・異常を識別させる．蒸留をしない小さなモデル S-AE-nd を加えた 3 つのモデルをソースドメインとターゲットドメインそれぞれに対する AUC (AUC-s, AUC-t)と pAUC での評価を比較して蒸留の効果を調査する.

◆結果

今回のモデルはモデルサイズを小さくしただけでも予測精度は同等であったため，蒸留による明確な効果は確認できなかった.

Table.1: Anormaly detection results.

Machine	Method	L-AE	S-AE	S-AE-nd
All	AUC-s	72.54%	70.78%	71.60%
	AUC-t	48.12%	47.92%	47.50%
	pAUC	54.81%	54.44%	54.46%

3-P-23

3-P-23　距離に基づく音源分離を用いたシングルチャンネル環境音分類

Single-channel Environmental Sound Classification Using Distance-Based Sound Separation

☆小倉稜也(都立大/JAIST)，井本桂右(同志社大)，
貴家仁志，塩田さやか(都立大)

◆環境音分類ではマルチチャンネル信号を用いることが一般的だが，チャンネルの一部が欠損することで分類性能が大幅に低下する．

◆本稿では，シングルチャンネル信号を用いて，環境音分類の前処理として距離に基づく音源分離を行うことを提案する．

◆Fig.1 は，マイクロホンからの距離が異なる二種類の環境音を重畳した音源から，提案法を用いて元の音源に分離した結果を表すスペクトログラムである．

◆提案法を用いて処理した環境音を入力として用いて環境音分類を行うことで，処理しない場合に比べて分類性能が向上した．

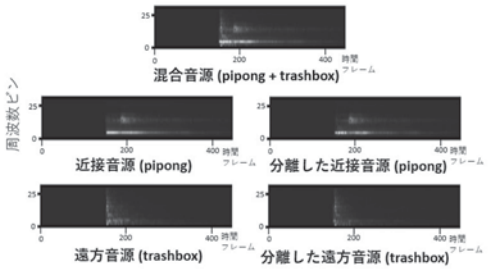


Fig.1 : Spectrograms of mixed and separated sound using pipong for near sound and trashbox for far sound.

3-P-24

3-P-24　光ファイバセンシングのための学習済み環境音認識モデルを用いたイベント分類

Example of an abstract sheet (16pt)

◎砺波 紀之，美島 咲子，近藤 玲史(NEC)，
井本 桂右(同志社大学)，△樋野 智之(NEC)

◆光ファイバセンシング技術のひとつである、振動や音響信号を検出可能な Distributed Acoustic Sensing (DAS) のためのイベント分類手法を提案する。

◆DASは耐電磁性やセンシング範囲の広域性、通信用既設ファイバの利用可能性から、これまで地震検知や構造物監視、侵入検知などに応用されている。

◆DASで取得された音データや事前学習済みモデルなどの不足によって、汎用的な環境音分析の研究は限られてきた。

◆本研究では、マイクロホンで収録された音響信号で構築された大規模音認識モデルを用いたDASのためのイベント分類手法を提案する。

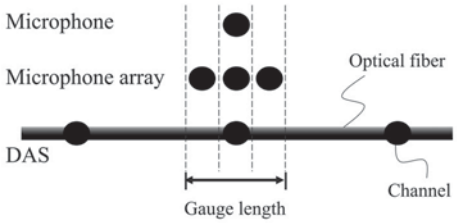


Fig. 1: Comparison of DAS and microphone

3-P-25

3-P-25　First Shot 異常音検知のためのおもちゃ稼働音データセットとベースライン実装

ToyADMOS2+ Dataset for the First-Shot Anomalous Sound Detection and a Baseline System Implementation

○原田登，仁泉大輔，大石康徳，竹内大起，安田昌弘(NTT)

◆本稿では，DCASE (Detection and Classification of Acoustic Scenes and Events) 2023 Challenge Task 2 で実施された First-shot 異常音検知タスクの概要と，新規に収録したデータセット ToyADMOS2+，およびベースライン実装を紹介する。

◆First Shot 異常検知のために，Fig.1 に示すように，おもちゃのドローン(ToyDrone)，おもちゃの電車(ToyNscale)，掃除機(Vacuum)，おもちゃの戦車(ToyTank)の 4 種類の機器の稼働音を新規収録した。

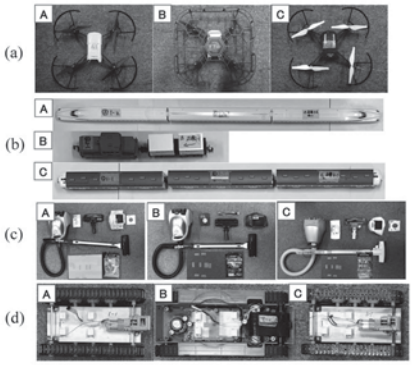


Fig.1 : Miniature-machine models.

3-P-26

3-P-26　Snore event detection using large-scale pre-trained audio neural network

○ Jiping Wu, Jianmin Wu, Xiaolong Lin, Gang Qiao (Baidu Inc.)

◆Snoring is a typical early symptom of Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS). Audio-based snore event detection can serve as the first step to OSAHS screening.

◆Convolutional neural networks (CNN) have been proven to be effective in sound event detection. Recently, models pre-trained on large-scale datasets have shown outstanding performance in recognition tasks.

◆In this work, a CNN system pre-trained on AudioSet is proposed. Fig. 1 shows the pipeline of snore event detection systems. The Mel spectrograms of sleep recording were fed to the model for learning to distinguish snoring and non-snoring sounds. Compared to the un-pretrained systems, the pre-trained systems can achieve better performance, with an average F1-score improvement of 1.9%

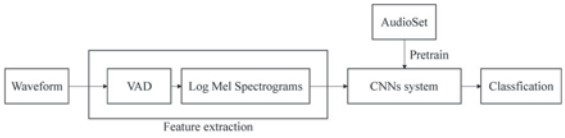


Fig.1 : Pipeline of snore event detection system

3-P-27

3-P-27　機種非依存異常音検知のための自己符号化器と分類器の条件付け

Conditional autoencoder and classifier for machine-independent anomalous sound detection.

○太刀岡 勇気（デンソーアイティーラボラトリ）

◆DCASE 2022 task2 では、機種を表す *machine type* と、同一機種内の部分集合である *section id* の２種類の手がかりが提供されている
- ➢ 通常、単一機種の正常音のみで機種依存モデルを複数学習する
- ➢ 機種非依存モデルが実現できれば、モデルの切り替えなしで複数の機種に適用可能であるため、有用

◆inlier modeling: AE/VAE を条件づける方法 (図 1)

◆outlier exposure: マルチタスク学習を単一の機種非依存分類器に適用 (図 2) → 機種ブラインドな機種非依存異常音検知

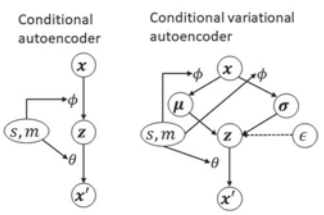


Fig.1: Conditional autoencoder and variational autoencoder conditioned with machine type *m* and section id *s* (inlier modeling)

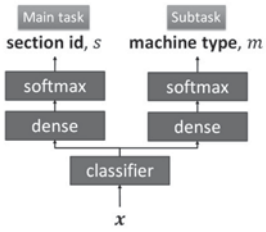


Fig.2: Multitask learning for conditioning classifier by machine type *m* and section id *s* (outlier exposure)

3-P-28

3-P-28　フレームシフトの調整による低速度域車両走行音のデータ拡張

Augmentation of Various Speed Data by Controlling Frameshift for Acoustic Traffic Monitoring

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

◆機械学習を用いた音響交通モニタリング(マイクロホンにより収録した車両走行音から交通状態を推定)において，十分に学習データがない場合においても推定精度を向上させるデータ拡張手法を提案した．

◆提案手法は時間周波数領域におけるフレームシフトを調整することで，学習データにおける低速度域データの不足を補うことができた．

◆交通速度の正解値と推定値の時間変化を Fig. 1 に示す．データ拡張を行わなかった上図に対し，提案手法によりデータ拡張を行った下図は学習データにない低速度域の交通速度に対する推定精度が向上した．

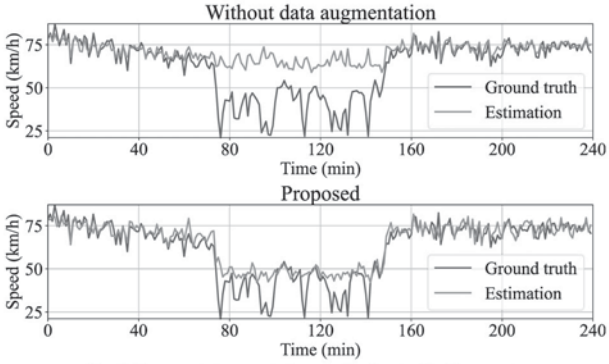


Fig. 1: Temporal changes in true and estimated traffic speeds with and without data augmentation

3-P-38

3-P-38　深層学習モデルを用いた楽曲と混合した雑音の抑圧性能の評価

Evaluation of suppression performance of noise mixed with music based on deep learning

☆岩本 晃周，西村 竜一（和歌山大）

◆録音済の楽曲に含まれてしまった雑音を抑圧することを目的に、深層学習モデル Conv-TasNet を用いた雑音抑圧性能を評価

◆4 通りの学習条件の Conv-TasNet モデルを比較
- ➢ 損失関数：MSE、SI-SNR
- ➢ 学習データ（楽曲・雑音の混合音）の SNR：20 dB、5 dB

◆主観評価（5 段階）によって雑音抑圧と楽曲劣化の結果を調査
- ➢ 評価項目：「雑音は楽曲に対して小さい」「雑音は気にならない程度である」「楽曲の音質劣化が小さい」
- ➢ 実験協力者：6 名
- ➢ 評価サンプル数：140（抑圧前後のものをランダム再生で提示）
  - ✧ 楽曲 2 種類、雑音 7 種類 × SNR 20 dB、5 dB

◆損失関数 SI-SNR、学習データの SNR を 5 dB の条件で学習した Conv-TasNet モデルの雑音抑圧性能が高い結果となった（Fig. 1）

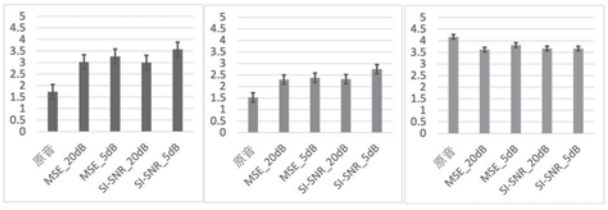


Fig. 1Evaluated rates ("Q1. Noise is small relative to the music", "Q2. Noise is not bothersome", "Q3. Sound quality degradation of the music is small")

3-P-39

3-P-39　深層両耳マルチフレーム MVDR フィルタに基づく拡散性雑音の残留量制御

Residual Control of Diffuse Noise Based on Deep Binaural Multi-frame MVDR Filter

◎佐藤僚，昼間信彦(リオン)

◆両耳音声強調に用いられる深層マルチフレーム MVDR フィルタへ雑音抑制量の目標値を補助的に入力することにより、拡散性雑音の残留量を制御する手法を検討し、音声強調性能や空間情報の保持へ与える影響について調査した。

◆提案手法では、先行研究と同様に残留量をポスト処理よりも優れた精度で制御可能であった。

◆雑音の残留量を増やすことにより以下の結果を確認した。
- ➢ 拡散性雑音の空間情報(MSC)の改善
- ➢ 音声成分の劣化が緩和
- ➢ 抑制量の目標値を-120 dB から-21 dB とした際、MBSTOI が改善

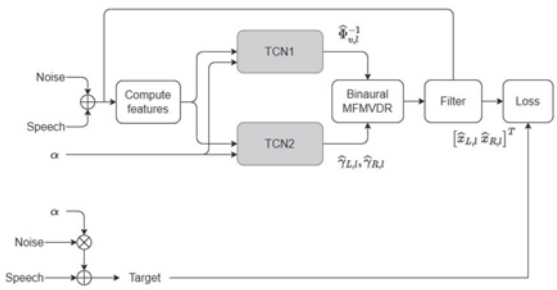


Fig.1: Block diagram of the proposed method.

3-P-40

3-P-40　Mask-based Neural Beamforming の移動音源拡張に関する検討

Extension of mask-based neural beamforming for moving sources

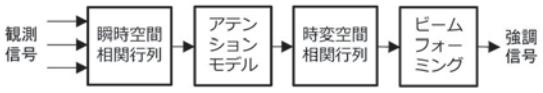
○落合翼，デルクロアマーク，中谷智広，荒木章子（NTT）

◆ Mask-based Beamforming における空間相関行列 (SCM) 推定問題の移動音源拡張

→ 移動音源の追従に最適な時変 SCM の推定方法に関する検討

◆ 提案手法

1. 従来の空間相関行列推定の枠組み（e.g., オンライン）を一般化
   → 瞬時 SCM とアテンション重みに基づいて時変 SCM を表現
2. 深層学習により音源の追従に最適なアテンション重みを推定



◆ 評価実験

1. 左表より，提案手法 att_mvdr ＞ 従来手法 onl_mvdr, blk_mvdr 移動音源に対し，より高い音声強調・音声認識性能の達成を確認
2. 右図（ビームパターン）より，提案手法によって推定された時変フィルタが移動音源を適切に追従していることを確認

Method	non-moving SDR ↑	non-moving WER ↓	moving SDR ↑	moving WER ↓
mixture	5.3	4.9	5.3	4.9
tiv_mvdr	15.1	**2.9**	11.4	3.8
onl_mvdr	13.5	3.4	10.2	4.1
blk_mvdr	13.0	3.1	11.4	3.8
att_mvdr	**17.8**	3.4	**16.7**	3.8
+ smoothing	15.4	3.0	13.9	**3.4**

黒線：音源方向

青線，赤線：1kHz, 2kHz

3-P-41

3-P-41　音声強調におけるドメイン適応のための Noise2Noise 学習

Noise2Noise learning for domain adaptation of speech enhancement

○李莉，関翔悟（サイバーエージェント）

◆深層学習を用いた音声強調において，ノイジー音声とクリーン音声のペアを学習データとする教師あり学習の有効性が示されている．一方で，実環境データに対しては学習時と推論時のデータの分布間に不一致が生じることから，強調性能が低下する傾向がある．

◆本稿では，音声強調におけるドメイン適応のために，RemixIT における再混合を二回行うことにより得られるノイジーペアに対して Noise2Noise 学習を適用することを提案し，その有効性を CHiME-7 Task2 で確認する．

Table 1　LibriCHiME-5 評価セットにおける SI-SDR [dB].

Methods	1 spk	2 spk	3 spk	Avg.
Sudo rm-rf	7.94	7.66	5.97	7.80
RemixIT	11.05	10.76	9.86	10.94
+N2N loss (prop.)	11.43	11.61	10.89	11.46
only N2N loss (prop.)	**11.48**	**12.11**	**11.81**	**11.65**

Table 2　CHiME-5 評価セットにおける DNS-MOS.

Methods	OVR	BAK	SIG
Sudo rm-rf	**2.88**	3.59	3.33
RemixIT	2.84	**3.63**	3.29
+N2N loss (prop.)	2.84	3.42	**3.37**
only N2N loss (prop.)	2.84	3.57	3.31

3-P-42

3-P-42　音響による接近車両検出における長時間データを効率的に利用可能な CNN モデルの検討

A Study on CNN model with Long-Time Input Data for Approach Vehicle Detection Using Sound

○伊藤隆佑，神谷珠緒，旭健作，坂野秀樹（名城大院）

◆出会い頭衝突事故の防止を目的に，道路上で収録した音を用いた接近車両検出を検討している．従来と同じ構造の CNN モデルに 2.5 倍の長時間データ（2.5 秒分）を入力すると，全体的な検出精度は向上するが，他車が自車に最接近する時刻などで検出精度が低下していた．

◆本研究では，より高い検出精度を目指し，長時間データを効率的に利用可能な新たな CNN モデルとして，データを中央（1.25 秒）で分割して入力するモデルを提案する．分割された入力データは，それぞれ従来研究と同じ構造のネットワークを通過させて 2 つの特徴量に変換される．その後，連結層で連結し，全結合層へ入力され判定される．

◆提案モデルを評価した結果，最接近時刻（0.0 秒）において従来モデルより正答率が**約 0.4 ポイント向上**した．また，接近する車の検出において全体的に正答率が**0.02 ポイント以上向上**した．

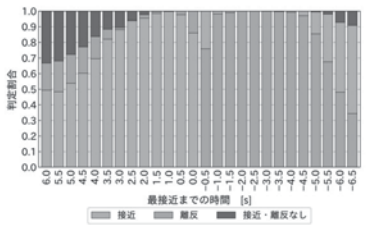


Fig.1: The classified probability each time in approaching and moving away situations. (New model)

3-P-43

3-P-43　複数マイクロホンによるスペクトログラムを用いた CNN による接近車両検出に関する検討

A Study of an Approaching Vehicle Detection by CNN Using Spectrograms with Multiple Microphones

☆神谷珠緒，伊藤隆佑，旭健作，坂野秀樹（名城大院）

◆CNN を用いて、車両走行音から接近車両をリアルタイムに検出する検討を行っている。しかし、これまで使用してきたデータセットは、リアルタイム性保持のために 1 入力データが(67×33×1)と、比較的小さいデータサイズであることが検出精度向上の妨げとなっていた。

◆そこで音響データ収録時に使用したマイクロホンアレイの他チャネルデータを利用する。本研究では、1 入力データを 1 チャネルから 4 チャネルへ拡張することで、先行研究からの検出精度向上を図る。

◆作成した 4 チャネルデータを用いて学習を行った結果、**検出精度は約 78%**となり、1 チャネルデータを学習させたモデルから約 5.2pt 精度向上した。特に接近データの検出精度は約 30pt 精度向上しており、接近データの判定精度向上に有効であることが確認した。

Table.1 : Confusion Matrix of the proposed model. Values in this table show the predicted percentage for each correct label data.

(%)		predicted label: approach	leave	nocar
correct label	approach	83.14	3.50	13.36
	leave	17.28	62.96	19.76
	nocar	9.89	2.32	87.79

3-P-44

3-P-44　複数音高推定における楽器依存性の少ないモデルによる推定精度向上に関する検討

Investigation for improving estimation accuracy by a model with less instrument dependence for multi-pitch estimation

☆川凌司，坂野秀樹，旭健作（名城大）

◆音高情報と音色情報に分離する音高推定モデルを提案する.

◆提案手法

- 音高情報と音色情報スペクトルの推論には U-Net を用い，それらの畳み込みによって入力スペクトルを再構成する.
- 音高推定は分類器の前段に U-Net を置く構成を用いる.

◆結果

- 提案手法の音高推定結果は Average Precision で 78.09%であり，従来手法の約 1.2%向上した.
- 再構成した音高情報と音色情報のスペクトルは期待したスペクトル形状とはならなかった.

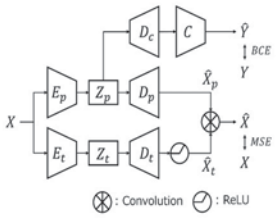


Fig. 1 The proposed architecture.

Table 1 Multi-pitch estimation results.

Model	Average Precision (%)
U-Net+Res34	76.83
Proposed	78.09

3-P-45

3-P-45　wav2vec 2.0 を用いた音素認識による日本語単語了解度推定

Japanese word intelligibility estimation by phoneme recognition using wav2vec 2.0

☆服部真稀，近藤和弘（山形大院・理工学研）

◆音声了解度の客観評価手法として，音素認識を用いたノンリファレンス式の推定手法を検討

◆了解度の主観評価手法 JDRT (Japanese Diagnostic Rhyme Test)を音素認識で模擬することで了解度を推定する

◆音素認識モデルは wav2vec で構築し，試験用音声や雑音などでファインチューニングさせた

◆先行研究(GMM-HMM マルチコンディション)の適応学習の際に課題であった原音付近の認識率の低下を克服 (Fig. 2)

◆主観評価に対しては相関係数 0.95，RMSE 15.42．誤差が依然として大きいのが課題であるが，モデルの学習条件については，まだ改善の余地があり、更なる精度向上が見込める

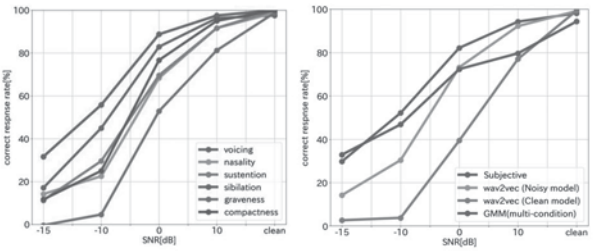


Fig.1 : Correct response rate with wav2vec model for each attribute vs. SNR (Left)
Fig. 2: Correct response rate with subjective & objective vs. SNR (Right)

3-P-46

機械学習を用いた家畜感染症の早期発見の検討

Examination of early detection of livestock infectious diseases using machine learning

○岡田なつみ，石光俊介，△湯本誠司，的場悠基（広島市立大院），
△三上修，井上寛暁（農研機構），田中大介（新居浜光線高専）

◆家畜毎に収録した体内伝導音から罹患個体を早期発見する手法を検討している。本研究では、肺炎感染豚４頭から健常と罹患の二値分類モデルを作成し、学習データとは異なる収録期間のデータを評価した。

◆学習モデルは、訓練精度 98.6 %、検証精度 98.6 %、テスト精度 96.3 %となった。作成した学習モデルで Table 1 の３頭を評価した。罹患正答率と健常正答率をそれぞれ Fig.1 と Fig.2 に示す。罹患正答率は異なる収録期間のデータでも向上したが、健常正答率は低下した。今後は健常正答率を向上させるパラメータを検討したい。

Table 1 Evaluation data

収録期間	頭数	データ数	
		健常	罹患
2021/9~10	3	1877	977

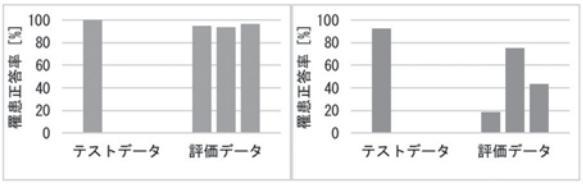


Fig.1 : Positive Accuracy　　Fig.2 : Negative Accuracy

3-P-47

3-P-47　体内伝導音を用いた豚の罹患判定モデルの構築

Construction of a model for disease detection using body-conducted sound

○的場悠基,石光俊介,△岡田なつみ,△湯本誠司（広島市立大院),△三上修,
△小川洋介,△井上寛暁,△石田三佳（農研機構),△田中大介(新居浜高専)

◆農林水産省の畜産統計によると，1962 年をピークに豚の飼養戸数が減少する一方，１戸あたりの飼養頭数は増加している．このような状況下で家畜の感染症が蔓延した場合，大きな被害が出る恐れがある．よって，家畜を効率的に管理し，罹患を早期に発見するシステムの開発が求められている.

◆本研究では，豚の体内伝導音のスペクトログラム，ケプストラム，Δケプストラムの３つをそれぞれ入力として，ニューラルネットワークの１種である LSTM(Long short-term memory)を用いて罹患前後の２クラス分類を行うモデルの構築を行った.

◆罹患豚８頭を用いて罹患判定モデルを構築し，そのモデルを用いて学習に未使用な豚 6 頭を分類した．その結果，健常時はΔケプストラムモデルの分類精度が全体的に高いが，罹患時はスペクトログラムやΔケプストラムの方が分類精度は高いことがわかった.

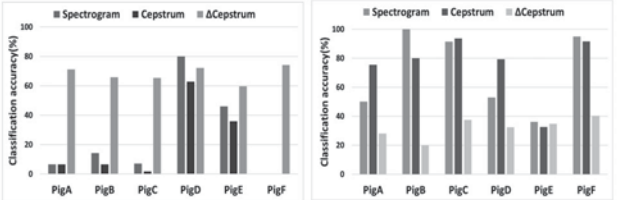


Fig.1(a) 健常時の分類精度　　Fig.1 (b) 罹患時の分類精度

3-P-48

3-P-48　振幅・位相スペクトルおよび時間包絡が等しくリサージュ図形のみが異なる両耳雑音で想起される心象印象の比較

Comparison of impressions evoked by binaural noises with equal frequency and phase spectra and time envelopes but different Lissajous figures

☆沖田和久，梨子木快晴(NBU)，近藤善隆((株)ジェイテック)，
河納隼一((株)アーネット)，福島学(NBU)，△松本光雄()，柳川博文(arsl)

◆目標：両耳知覚で想起する空間印象と物理量の対応を明らかにする
◆前報：l(t)=sin(ωt)とr(t)=sin(ωt+θ)から作成した白色化信号θを変えることで「広さ」感が変わることを明らかにした
◆不足：白色雑音との比較が未実施
◆本稿：白色雑音と白色化信号で次の２項目を調査
　１）区別できるのか，２）想起する印象の違い
◆刺激：白色雑音）１）同相（In-phase），２）逆位相（Anti-phase），３）無相関（Out of phase）
　白色化信号（Randomly Permuted Signal Sin and Cos signals）
◆結果：１）区別された，２）想起された印象がわかった

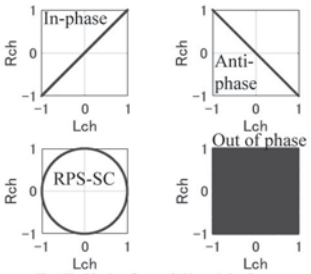


Fig. 1 The Lissajous figures of 4 binaural stimuli.

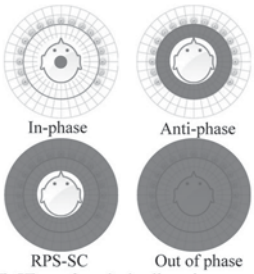


Fig. 7 The most frequently selected impression patterns

3-P-49

3-P-49　映像シーンと音響情報の両耳間相関関数による整合性の検討

Investigation of Coherence Using Inter-channel Correlation Functions between Visual Scenes and Auditory Information

☆梨子木快晴，沖田和久(NBU)，河納隼一(㈱アーネット)，
近藤善隆(㈱ジェイテック)，福島学(NBU)，
風間道子(log a)，柳川博文(arsl)

◆　課題：音で想起する情景と映像で想起する情景の整合性
◆　目的：整合性を何で評価することが人の印象と対応するのかを明らかにする．
◆　本稿：音のシーンチェンジ点（ASCP: Acoustic Scene Change Point）と映像のシーンチェンジ点（SSCP: Sight Scene Change Point）で評価
✧　結果：入選　ASCP と SSCP　一致点が多い（Fig. 1 左）
✧　　落選　ASCP と SSCP　少ない（Fig. 1 右）

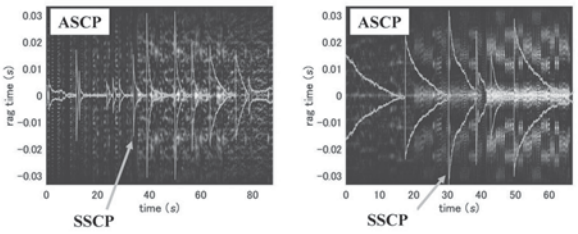


Fig.1 Correspondence of Temporal Variation between Acoustic Scene Change Point (ASCP) and Sight Scene Change Point (SSCP)
(Left：Prize-winning artworks, Right：Rejected artwork)

3-Q-1

3-Q-1　End-to-end モデルを用いたオーバーラップに頑健な音声認識モデルの評価

Evaluation of Robust Speech Recognition Models for Overlap Using End-to-end Models

☆三河多聞(豊橋技科大)，△山岡大志(Poetics)，北岡教英(豊橋技科大)

◆日常的な会話において複数の音声が重なり合う「オーバーラップ」は，相槌などの短い音声により頻繁に発生し，音声の波形が類似する複数の音源が重なり合うことで，音声認識精度の低下につながる。
◆本研究では，まず日本語話し言葉コーパスから，音声認識の対象となる音声と，相槌などを想定した短い音声を取り出し，オーバーラップを含む音声データセットを人工的に作成する。このデータを利用し，音声認識モデルに対して fine-tuning により耐オーバーラップ性を付与し，fine-tuning の前後での CER の変化を比較した。
◆fine-tuning の実施により，オーバーラップの強度（SNR）に関わらず CER が改善し，音声認識精度を維持したまま，モデルが耐オーバーラップ性を獲得することが確認できた。また，モデル構造が異なる音声認識モデルにおいても精度が改善することを確認した。

Table 1　CER of each models for test set.

Model	non-overlap	Overlap Test Set (SNR) 10	5	0	-5	-10	random
Conformer-Transducer							
Pre-train	**4.17**	10.10	12.48	14.08	14.90	15.48	13.52
SNR 10 Fine-tuned	4.40	**5.41**	6.45	7.99	9.73	11.36	8.10
SNR 5 Fine-tuned	4.49	5.55	**6.01**	6.84	8.25	9.81	7.19
SNR 0 Fine-tuned	4.50	6.13	6.18	**6.71**	7.89	9.09	**7.03**
SNR -5 Fine-tuned	4.37	6.91	6.61	6.92	**7.84**	**8.94**	7.34
SNR -10 Fine-tuned	4.47	7.54	7.41	7.56	8.14	9.11	7.92
random Fine-tuned	4.48	5.71	6.24	6.92	8.03	9.15	7.10

3-Q-2

3-Q-2　認識困難な単語に注目した文章生成と音声合成による音声認識のためのデータ拡張

Data augmentation for speech recognition by sentence generation and speech synthesis focusing on hard-to-recognize words.

☆北條圭悟(豊橋技科大)，河東宗祐(Poetics Inc.)，北岡教英(豊橋技科大)

◆End-to-End 音声認識は，非常に高精度な認識が可能であるが，実際の認識結果を見ると特定の単語が上手く認識できない場合がある。
◆この対策として，その単語の発話が含まれる学習データの拡張が挙げられるが，音声認識の学習には音声データとそれに対応した書き起こしが必要で，収集するには時間的・金銭的に膨大なコストを要する。
◆本研究では，大規模言語モデルと音声合成モデルを用いた音声認識モデルが認識困難とする単語に注目したデータ拡張手法を提案する。
◆既存の音声認識モデルが認識困難とする単語を抽出し，GPT-2 と Tacotron2 及び Parallel WaveGAN を用いて，その単語を用いた文章生成と対応する音声を生成し，音声認識用データの拡張を行う。
◆実験の結果，提案手法は音声認識モデルが認識困難とした単語の認識精度を改善させることが判明した。

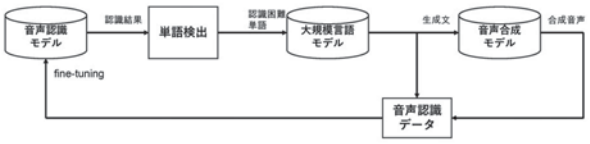


Fig.1：Overview of the proposed method.

3-Q-3

3-Q-3　End-to-End 複数言語音声認識モデル訓練における距離学習の効果

Effectiveness of Metric Learning for End-to-End Multi Language Speech Recognition Model Training

☆土橋晃弘(山梨大･院), レオ チー シャン(山梨大･院), 西崎博光(山梨大･院)

◆ wav2vec2.0 をベースにした複数言語の音声認識システムの認識性能向上のために，距離学習を用いた訓練方法を提案

- ➢ 図１に示すような強制アライメント処理と距離学習方法を，CTC 損失関数をベースにした音声認識システムに実装し，効果を検証
- ➢ 異なる言語に対しても，同じ文字が同じ音響特徴ベクトルを生成するように学習

◆ 複数言語の音声認識精度(CER)が全言語において 0.7 %向上. また，より適切な特徴（エンベッド）ベクトルを抽出できた

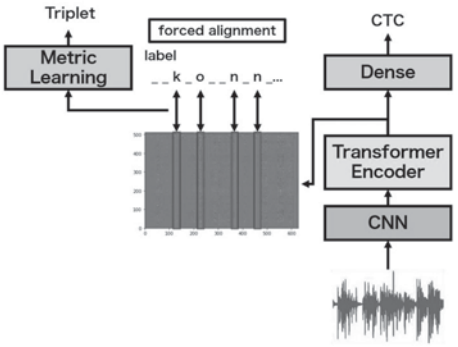


Fig. 1 ASR model with forced alignment and metric learning

3-Q-4

3-Q-4　非流暢現象ラベル付き発音形認識モデルとテキスト変換モデルを組み合わせた音声認識システム

Automatic speech recognition system integrating a disfluency-labeled speech recognition model and a text conversion model

〇藤江真也(千葉工大/早稲田大), 小林哲則(早稲田大)

◆非流暢現象（フィラー，言い淀み）を扱う音声認識システムを検討

◆日本語の発音形（カナ文字系列）を基本的な出力とし，さらに非流暢現象ラベル（フィラーは"+F"，言い淀みは"+D"）を組み合わせたトークンを出力とする発音形音声認識モデルを構築

◆さらにその出力を入力とし，平文（かな漢字混じり文）に変換する機械翻訳モデルを構築して組み合わせ，全体のシステムを構築（Fig.1）

◆発音形認識モデルを日本語話し言葉コーパスで学習・評価した結果，文字誤り率，非流暢現象検出精度とも従来研究と同等の結果が得られた．実際の対話システムへの応用に向けてストリーミング音声認識での実験でも同等かそれ以上の性能が得られた（Table 1）

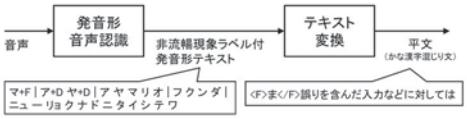


Fig.1 : System Overview

Table 1 : The experimental results

	CER[%]			Fillers			Disfluencies		
	eval1	eval2	eval3	Prec.	Rec.	F1	Prec.	Rec.	F1
Conf./TF+CTC	4.4	3.2	4.2	0.941	0.925	0.933	0.835	0.528	0.647
CBS/TD	4.7	3.3	4.2	0.933	0.927	0.930	0.747	0.577	0.651

3-Q-5

3-Q-5　Cross-lingual Mapping for Hierarchical Softmax in Low-Resource Multilingual Speech Recognition

☆ Zhengdong Yang, △ Qianying Liu (Kyoto Univ.), Sheng Li (NICT),
△ Chenhui Chu, △ Fei Cheng and △ Sadao Kurohashi (Kyoto Univ.)

◆ We presents a novel approach to enhance Automatic Speech Recognition (ASR) performance for low-resource languages, addressing the limitations of current systems that cover only a small fraction of spoken languages worldwide.

◆ We propose a method based on cross-lingual embedding clustering to construct a hierarchical decoder for multilingual ASR systems, enabling the system to capture token similarity and transfer knowledge from resource-rich languages to those with limited resources.

◆ The method involves creating binary representations of cross-lingual embeddings with hierarchical clustering and constructing the hierarchical softmax decoder with these representations for training and decoding.

◆ Experiments on the Turkic language group show that the proposed method outperforms previous baselines, with pre-trained cross-lingual embeddings performing better than a mapped embedding method, demonstrating the potential of pre-trained cross-lingual embeddings.

Method	CER%					
	ba	ky	tr	tt	uz	AVG
Softmax	17.1	18.4	16.4	11.1	18.5	16.3
H-Softmax (Huffman)	16.8	17.8	14.9	11.1	16.5	15.4
H-Softmax (Mapping)	15.4	17.5	12.0	9.7	19.0	14.7
H-Softmax (LABSE)	**13.4**	**14.8**	**11.5**	**8.3**	**14.7**	**12.5**
H-Softmax (XLM)	14.5	16.6	12.1	9.1	15.0	13.5

3-Q-6

3-Q-6　CTC デコーダを用いたストリーミング可能な音声認識における Density Ratio Approach に基づくドメイン適応

Streaming ASR with CTC decoder based on Density Ratio Approach for Domain Adaptation.

☆高城巽成 1, 若林佑幸 1, 小川厚徳 2, 北岡教英 1, (1 豊橋技科大, 2 NTT)

＜ 研究目的 ＞

- ストリーミング音声認識において、訓練データとテストデータのドメインの違いによる精度低下が課題としてある。
- 言語モデルのみを用いた Density Ratio Approach(DRA)に基づくドメイン適応によりストリーミング音声認識の精度向上は可能か？

＜ 提案手法 ＞

- CTC のフレーム単位の出力に対して逐次的に DRA を適用するためにデコーディングアルゴリズムを改良する。(図１)
- CTC の条件付き独立性の仮定を考慮し、減算用言語モデルとして 1-gram が適していると考え、2-gram, 3gram も用いた様々な組み合わせで実験し、検証する。

＜ 実験結果 ＞

結果としては、提案手法により、ストリーミング音声認識のドメインが異なるデータに対して精度向上することができた。
しかし、減算に 1-gram を用いることが必ずしも良い結果とはならないこともわかった。

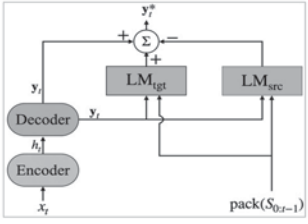


Fig 1 Extension of DRA to apply to frame-by-frame output.

3-Q-7

3-Q-7　A Multimodal Model for Personality Recognition through Speech

○Nathania Nah (Tokyo Tech), △Yuri Tsuchiya (Tokyo Tech),
Takafumi Yoshinaga (YCU), Koichi Shinoda (Tokyo Tech)

Exploring the field of affective computing is important for understanding how humans think and interact with each other. Personality computing focuses on methods of performing the automatic detection of human traits which compose their personality. Using the Five Factor Model of Personality as a measure to describe a subject's personality, temperament, and psyche, this work employs a multimodal model to perform automatic personality recognition on speech. We employ the use of speaker and phone disentanglement in speech representation learning, a technique known to be effective in emotion recognition, to predict scores for personality traits trained on the UDIVA dataset and outperform current methods that use visual features.

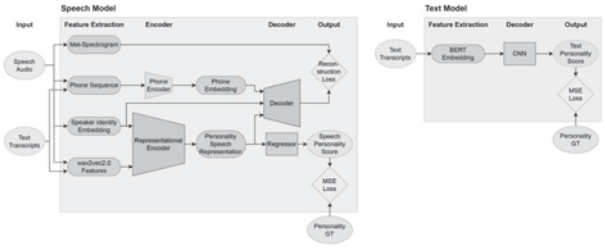


Fig.1：Overview of the architecture of our multimodal personality recognition model.

3-Q-8

3-Q-8　吃音者の音声認識における連発ラベル導入による連発箇所の検出

Detection of consecutive parts of speech by inserting consecutive labels in speech recognition for people who stutters.

☆松坂勇樹, 高島遼一(神戸大),
△安井美鈴(大阪人間科学大), 滝口哲也(神戸大)

◆吃音とは言葉が滑らかに話せない発話障害の一つであり，音声特徴自体は健常者に近いが，発話の際に吃音症状が生じるため滑らかな発話が困難であり，円滑なコミュニケーションをとること難しくなる.

◆効率的な吃音緩和の訓練のために，発話内容における吃音の発生箇所を音声認識で自動検出することが期待される.

◆吃音の大きな特徴として、「連発」と「伸発」と「難発」があるが，本研究では初期検討として，Fig.1(上)のように同じ言葉を繰り返してしまう症状である「連発」を対象に，連発箇所の自動検出を目指す.

◆本研究では連発箇所を検出するために，吃音者の読み上げ発話に対して，Fig.1 のように連発箇所に連発ラベル「#」を付与し，音声認識モデルの学習を行うことで，音声認識による自動検出を試みる.

◆また，連発箇所の検出精度を高めるために，連発ラベルを含めた健常者事前学習を行う手法も検証した.

◆吃音者の収録音声による実験の結果、提案手法の有効性を確認した.

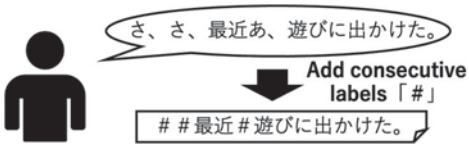


Fig.1：Characteristics of consecutiveness in stuttering symptoms and the assignment of consecutive labels.

3-Q-9

3-Q-9　話し言葉を考慮した言語特徴による音声感情認識

Speech emotion recognition by linguistic features for spontaneous speech.

☆岸恵汰(山形大), 佐藤清秀(山形大･ピットアート), 小坂哲夫(山形大)

◆従来，BERT を利用した言語特徴による音声感情認識では，書き言葉で学習した事前学習モデルを用いるのが一般的であった.

◆一方，音声感情認識は話し言葉を分類の対象とするため，**<u>口語調の感情認識に適した BERT モデルを使用する必要性</u>**がある。

◆本研究では，事前学習に使用した学習データの違いが音声感情認識にどのような影響を与えるかを検証した.

◆事前学習モデルとして３種類を比較した．また感情抽出する対象のテキストとして，書き起こしおよび音声認識結果２種類（感情適応の有無）を利用した.

◆従来法よりも，**<u>SNS のテキストで事前学習した BERT モデルが最も高い認識率</u>**となった.

Table.1：Recognition results for each model [%].

対象テキスト / 事前学習モデル	書き起こし	ASR(適応有)	ASR(適応無)
話し言葉	67.50	62.75	53.25
書き言葉	67.50	64.00	57.25
SNS	**80.00**	**76.75**	**59.75**

3-Q-10

3-Q-10　単語埋め込み表現を用いた感情音声からの字幕画像生成手法の検討

Subtitle generation from emotional speech using word embeddings

☆中村史也(神戸大), 相原龍(三菱電機), 高島遼一, 滝口哲也(神戸大),
△伊谷祐介 (三菱電機)

◆ 人間は発話内容を理解する際，言語情報以外にも話者の表情や声のトーンなどから推測される感情情報を利用している。

◆ 本研究は，音声認識 (ASR) で得られる言語情報と感情音声認識 (SER) による感情情報を字幕のフォントとして表現することを目的とする。

◆ Web サイトから収集したフォントのタグ情報と，発話テキストから抽出した重要フレーズおよび感情語との単語埋め込みの類似度スコアを算出することで感情音声に合ったフォントを検索する手法を提案する。

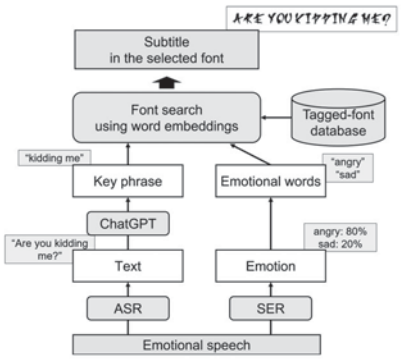


Fig.1：Procedure for generating subtitle images from emotional speech.

3-Q-11

3-Q-11　天井カメラの音声・映像情報に基づく対話活性度の推定

Estimation of conversational activity
based on audio and video information from a ceiling camera

○金宜鉉，△野田玲子，須藤隆（東芝）
△黄静君，△西村信孝（東芝インフラシステムズ）

◆2名の者が会話または意見を交換しあう時の盛り上がりの度合を対話活性度と総称する。天井カメラ映像のように明確ではない人物の音声情報や画像情報に基づいて、対話活性度を推定する深層学習モデルを提案する。

◆音声情報と画像情報とのマルチモーダル学習及び、発話区間推定・笑い声区間推定・感情認識とのマルチタスク学習を適用することで、高精度化を図った。

◆テストデータセットを用いた定量評価では、提案モデルからの対話活性度推定値と対話活性度正解ラベルで算出した平均平方二乗誤差と平均平方二乗誤差率が、各実験モデルと比べて一番低い０．６７と１３．４４%となり、提案モデルの有効性を確認することができた。

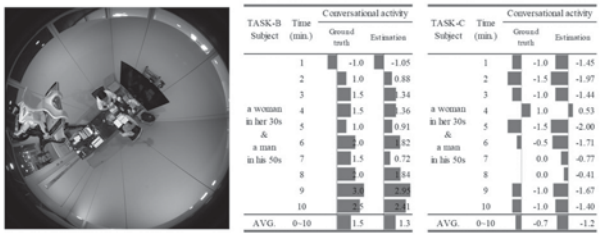


Fig.1 : Example of conversational activity estimation by the proposed model
(-2: discrete, -1: stray, 0: stagnation, 1: somewhat active, 2: active, 3: very active).

3-Q-12

3-Q-12　パーキンソン病の重症度に対する位相特徴量の分析

Analysis for advanced stages of Parkinson's disease
using phase characteristics

◎細田侑也（豊橋技科大）

目的

◆構音障害診断では音声の振幅スペクトルから特徴量を抽出.

◆副産物である位相スペクトルにも有効な特徴量が含まれるはず.

手法

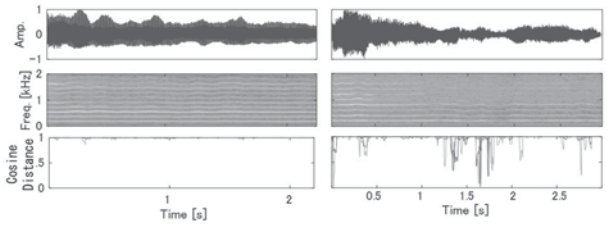
◆時間方向および周波数方向における位相の関係性を表す高調波間の位相差が次数に依存せずに一定であることを導出.

◆高調波間の位相差における観測値と理論値との一致度を cos 類似度で評価して，調波構造の欠如や不規則な変動を検出.

結果

◆高調波間の位相差から調波構造の状態の可視化を達成

◆軽度以上の構音障害を患う患者群で一致度の低下を確認.



(a) Healthy Control.　(b) Parkinson's disease patient.

Fig 1 : Waveform (Top), spectrogram (Middle panel), and degree of coincidence (lower) of sustained vowel /a/.

3-Q-13

3-Q-13　形態素単位での分析によるリアルタイム感情識別システム

Real-time emotion discrimination system analysed on morpheme unit

☆福吉力（広島市大院），中山仁史（広島市大院）

◆本研究では，単語単位に分割した音声を対象に音響特徴量を抽出し，形態素ごとの性能に注目したリアルタイム感情識別システムを提案する.

◆形態素解析したテキストデータ，音素アライメントとモーラ数を併用することで音声の単語分割を行う.

◆音声信号から音響特徴量への変換には openSMILE を用いた.

◆Table 1に形態素単位での感情識別実験におけるデータセット全体及びテストデータに対する各形態素の識別精度を示す.

◆実験結果から，発話時間の長い名詞や動詞よりも，それらを強調する連体詞，形容詞や副詞などの言葉に感情情報が含まれている可能性が示唆された.

◆発話文単位の推定結果から，悲しみの識別で精度が下がっており，中立状態に誤分類される傾向にある.

Table 1. Estimated number of correct labels per morpheme.

	N.	V.	Int.	Prt.	Aux.	Adj.	Attri.	Adv.	Pref.	Conj.
Total number in closed dataset	1792	761	87	1596	456	84	76	100	16	24
Correct for estimated labels	1625	689	80	1441	422	79	72	94	15	18
Recognition accuracy	0.91	0.91	0.92	0.90	0.93	**0.94**	**0.95**	**0.94**	**0.94**	0.75
Total number in open dataset	434	170	19	369	106	17	16	13	3	8
Correct for estimated labels	267	98	12	214	73	12	12	7	2	2
Recognition accuracy	0.62	0.58	0.63	0.58	0.69	**0.71**	**0.75**	**0.54**	**0.67**	0.25

3-Q-14

3-Q-14　肥大した仮声帯物理モデルが声帯振動に及ぼす影響について

Influence of enlarged ventricle fold physical model on vocal-ventricular fold co-oscillations

☆韓聰（立命館大），足立恭平（立命館大），顔聡（立命館大），
徳田功（立命館大）

◆声帯および仮声帯モデルを用いて、仮声帯肥大を模擬した喉頭吹鳴実験を行った。

◆声帯モデルのみが振動している場合に比べ、声帯と仮声帯が同時に振動する場合には、基本周波数が低下した。

◆仮声帯が通常の大きさのときは、仮声帯は振動せず、声帯のみが振動した。

◆仮声帯が肥大したとき、仮声帯は声帯と同期振動した。

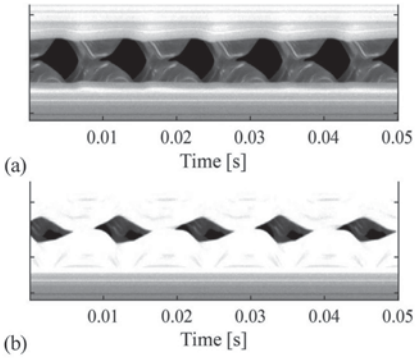


Fig.1 Kymograms observed when the ventricular folds size was (a) normal, (b) 1.3 times.

3-Q-15

3-Q-15　MRI 型声帯物理モデルを用いた声道狭窄模擬実験

Effect of vocal tract constrictions on physical model of the vocal folds

☆足立恭平(立命館大), 韓聰(立命館大), △廣瀬巧朗(立命館大), 顔聡(立命館大), 徳田功(立命館大)

◆声門上部の狭窄と口唇部の狭窄、及びそれらの組み合わせが発声に及ぼす影響について調べた。

◆声帯物理モデルには、MRI 型と EPI 型を用いて、発声閾値声門下圧を計測した。

◆EPI モデルでは声門上部に狭窄が有る場合に発声閾値圧が低下した。MRI モデルでは個体ごとのばらつきが大きく、一定の傾向はみられなかった。

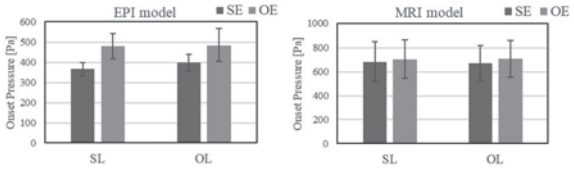


Fig.1 Comparison of onset pressure for EPI model (left) and MRI model (right). Comparison is made between vocal tracts with (SE) and without (OE) constriction at epilarynx. Effect of straw is considered by adding a narrow lip (SL) and normal lip (OL) to the vocal tract.

3-Q-16

3-Q-16　非負値行列因子分解に基づく音響的変化の可視化手法の基礎検討

Basic study of visualization for acoustic variation based on nonnegative matrix factorization

☆大村美結(広島市大), 大島風雅, 中山仁史(広島市大院)

◆非負値行列因子分解（NMF）での音響解析について，基底ベクトルの混合比や生起時間を得ようとした場合，基底間のスケールにおいて任意性の問題が生じる.

◆本稿では NMF を用いて音声信号を解析し，音響的変化を基底ごとに正規化を行いグラフ化することで可視化を行う手法を提案する.

◆実験結果より各アクティベーション同士の強度比較が可能であることが示され，単音節内の文脈依存音素を対象とした音声分離・音響分析の可能性を示した.

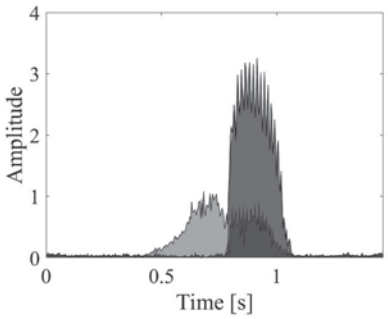


Fig. 1 Visualization results of stacked plot for the monosyllabic word /shi/.

3-Q-17

3-Q-17　基底数推定を用いた非負値行列因子分解の基礎検討

Basic study of nonnegative matrix factorization with estimated number of bases.

☆栗原和真, 大島風雅, 中山仁史(広島市大院)

- 音響解析における非負値行列因子分解（NMF）は任意数の周波数スペクトル及びその発音エネルギーに分解することができる.
- 既存の NMF では，信号中の潜在的パターンを示す基底数はユーザが設定する必要がある. 任意音源に対して適切な基底数を設定することは困難である上に，出力結果に大きな影響を与える.
- 我々は音声や楽音を対象に基底数を自動推定した上で NMF を行う手法を提案する.

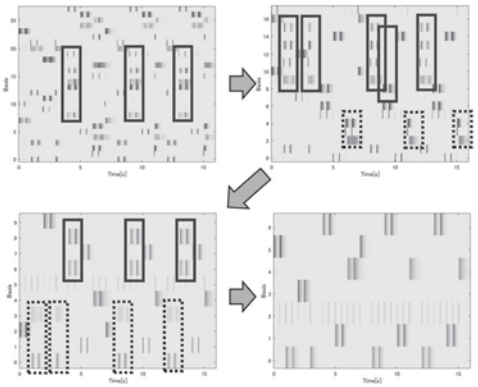


Fig. 1: Process of the merging bases with high similarity.

3-Q-18

3-Q-18　U-Net を用いた光骨伝導音の音質改善検討

Improvement on sound quality of optical-body-conducted speech using U-Net.

☆川本大貴, 中山仁史(広島市大院)

◆光骨伝導音は非磁性体で構成された光マイクロホンによって得られた骨伝導音である. 骨伝導音と同様に気導雑音に頑健であるとともに高周波数成分が減衰しにくいが，気導音と比較して周波数特性が異なり，声質の変化が生じる. この特性を活かして MRI のような強磁場・雑音下におけるコミュニケーション可能化を目標にする.

◆本研究では，光骨伝導音からより明瞭かつ自然な音声を得るために U-net を用いた音質改善を行う. 静粛環境にて気導音と光骨伝導音を同時収録し，SNR=10, 5, 0 dB になるように MRI 疑似雑音を光骨伝導音に付加した. U-Net によって雑音付き光骨伝導音をクリーンな気導音となるような音声変換を行い，学習済み WaveGlow によって音声変換後の音声生成を行った.

◆その結果, MCD による評価で約9.22 dB もの音質改善が確認できた.

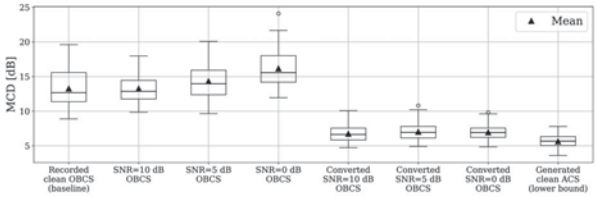


Fig.1 : MCD of speech signals before and after conversion.

3-Q-19

3-Q-19　円滑な対話における交替潜時の予測のための対話テンポの推定

Estimating Dialogue Tempo for Predicting Alternation Latency in Smooth Dialogue

☆岡永佳子，藤江真也（千葉工大）

◆適切な交替潜時は対話テンポに依存するという仮説を，日常会話コーパスの二者会話の観察により確認

◆重回帰分析とランダムフォレスト回帰モデルを用いて主観的な対話テンポを推定，及び Forward Selection を用いて有用な特徴量を調査

◆ランダムフォレスト回帰モデルの精度が高く，「発話の頻度」「交替潜時」「話速」が推定に有用な特徴量である可能性を確認（Fig. 1，Table 1）

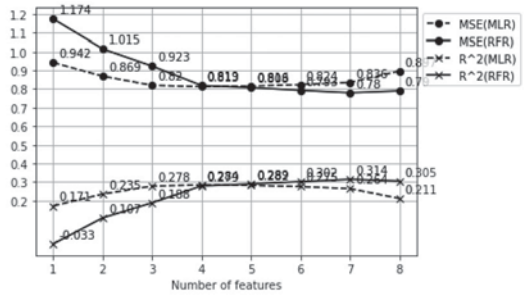


Fig.1 : Results from Forward Selection.

Table1 : The ranks of features based on importance for each estimation method

model＼feature	utterance frequency	Alternation Latency	Speech rate
MLR	1	2	3
RFR	1	5	3

3-Q-20

3-Q-20　Investigating Deep Prompting Approach in Spoken Lanugage Acquisition Agent

○ YANG Haoyuan (TokyoTech), OKAMOTO Takuma (NICT), SHINOZAKI Takahiro (TokyoTech)

◆This research leverages pre-trained spoken language models that aims at capturing intricate linguistic structures in spoken utterances to enhance the agent's performance and adaptability.

◆Learned representations of images will be concatenated as a prefix of pseudo speech units, working as a deep visual prompt for the pre-trained spoken language model for speech generation.

◆The results shows that the proposed method outperformed the baseline model in both observation stage and dialogue stage, reaching higher accuracy with only 10% tunable parameters to adapt to new datasets.

◆In future work, we will fix current limitation and adapt the sentence embedding mechanism for more generalized scenarios.

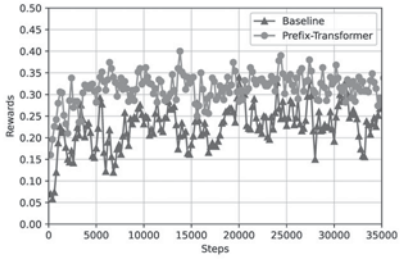


Fig.1 : Average Rewards in Dialogue Stage

3-Q-21

3-Q-21　発話タイミング推定における時間心理尺度の考慮

Consideration of psychological time scale in response timing estimation for conversation system

☆谷口友紀*，藤江真也*†，小坂直敏*　小林哲則*（*早大，†千葉工大）

◆ 目的：対話システムの発話タイミング推定器の学習方式とその評価指標に，タイミングのずれに関する心理時間尺度（感覚時間）を導入する．

◆ 貢献：

- 知覚実験によりタイミングのずれに関する弁別閾を求め，物理時間と感覚時間との関係式を導いた．(Fig. 1)
- 感覚時間を考慮した発話タイミング推定の評価尺度を提案し，発話タイミング推定結果を再評価した．従来指標に比べ，構文完了性の効果をより明確に示すことができた．(Table. 1, 2)
- 感覚時間を考慮したタイミング推定器の学習方式を提案し，発話タイミング推定精度を向上させることができた．(Table. 3)

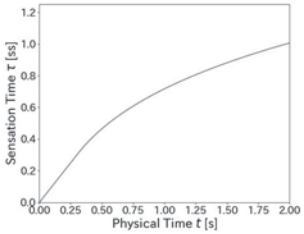


Fig. 1: Relationship between physical time and sensation time.

Table. 1: F1 evaluation using physical/psychological time scale.

Scale	Feature	$F_{0.25}$	$F_{0.5}$	$F_{1.0}$
Physical	Baseline	**0.390**	0.557	0.730
Physical	+ESC	0.386	**0.575**	**0.745**
Psychological	Baseline	0.442	0.615	0.739
Psychological	+ESC	**0.450**	**0.635**	**0.761**

Table. 2: F1 evaluation with transition interval of labelled training data during training.

Λ[mss]	$F_{0.25}$	$F_{0.50}$	$F_{1.00}$
0	0.442	0.615	0.739
30	**0.461**	0.639	0.761
60	0.446	**0.652**	**0.774**
120	0.425	0.635	0.763

3-Q-22

3-Q-22 音声合成モデルを用いた音声編集の検討

Investigation of speech editing using a speech synthesis model.

大西弘太郎，園部良介（ディー・エヌ・エー），
松永裕太（ディー・エヌ・エー/東大院）

★音声編集は与えられた音声の一部の言語情報を編集し、新たな音声を生成する

★深層学習に基づく音声編集手法では、音声編集用にモデルを作成する必要があったり、音声合成モデルが生成する音声を対象とした手法であったりする

★学習済みの音声合成モデルが生成する音声に対象とした手法に基づいた、実際の音声を編集することのできる手法を提案

★主観評価実験の結果、提案手法はメルスペクトログラムの結合による手法と比べて自然性で好ましいことを確認

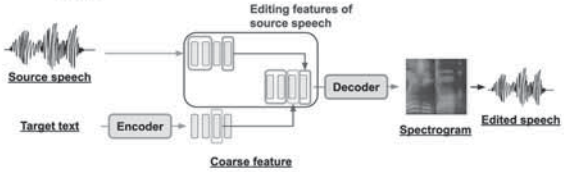


Fig. 1　Overview of the proposed method.

Table 1　Preference score (%) of edited speech naturalness evaluation.

Method	Preference score
Baseline (concat)	0.484
Proposed	0.516

3-Q-23

3-Q-23　ウェアラブルデバイスを用いた曖昧な入力からの会話支援システムの検討

Conversation Support System for Ambiguous Input Using Wearable Devices

☆岩崎茉理，原直，阿部匡伸(岡山大院・HS 統合科学研)

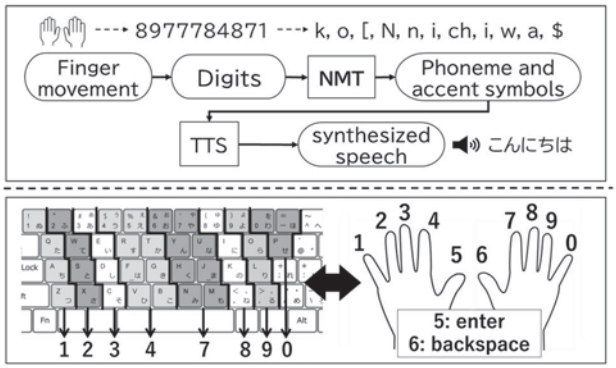
◆本報告では，テキストからの音声合成を用いた発話困難者向けの会話支援システムとして以下を提案する．

- ウェアラブルデバイスの為のタッチタイピングに基づく入力方式
- ニューラル機械翻訳（NMT）を用いた文章推定

タッチタイピングで定められる指とキーの対応関係を利用し，タイピング動作から入力文を推定する．これにより，入力の際に目視が不要となる．推定に NMT を用いることで，入力の曖昧性を解消する．

◆客観評価実験より，NMT の推定精度は 91.48%であることが示された．

◆主観評価実験より，提案システムを用いて会話を行うと，67.9%の文が相手に正しく伝わることが示された．



3-Q-24

3-Q-24　雑音環境下におけるテキスト音声合成システムの明瞭性向上に関する検討

A study on improvement of intelligibility for text-to-speech system in noisy environments

◎多湖崇宏，坂野秀樹，旭健作（名城大院・理工）

◆駅の構内放送や災害時の緊急放送では，背景雑音の有無に関わらず，聞き取りやすい明瞭な音声が必要である．

◆本研究では，そのような様々な雑音環境下への応用を想定し，雑音環境下であっても人間に聞き取りやすい音声を合成可能なテキスト音声合成（Text-to-Speech; TTS）システムについて検討する．

◆提案手法を実装した結果，合成音声の品質はそれほど良くないものの，雑音環境下における聞き取りやすさを意図的に制御し，聞き取りやすい明瞭な音声を生成することができた．

◆提案手法によって推測された雑音環境下における聞き取りやすさと，Ground truth の評価値との損失を，Fig. 1 に示す．

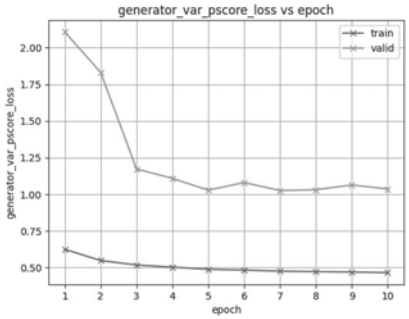


Fig. 1: Learning progress of Loss between the degree of popout inferred by proposed method and the ground-truth popout score.

3-Q-25

3-Q-25　SiFiSinger: SiFi-GAN を内包した歌唱音声合成

SiFiSinger: Singing Voice Synthesis including SiFi-GAN

☆芦田 裕飛，中鹿 亘（電通大）

◆ 歌唱音声合成においては基本周波数(f0)の制御と品質の双方が達成できることが望ましい

◆ テキスト音声合成においては End-to-End モデルが成功を収めているが，あくまでテキスト音声合成のためで，f0 の制御性は不足

◆既存手法の XiaoiceSing は WORLD ボコーダを用いることから，f0 の制御性は高いが，品質に限界あり

◆歌唱音声合成を End-to-End 化できれば品質が向上しそう

◆品質改善と f0 の制御性を両立できるようにするには

- 学習可能で f0 制御性もあるボコーダとして SiFi-GAN を活用

◆SiFi-GAN を内包した End-to-End 歌唱音声合成：SiFiSinger を提案

◆品質向上と f0 の制御性をおおよそ両立可能であることを実証

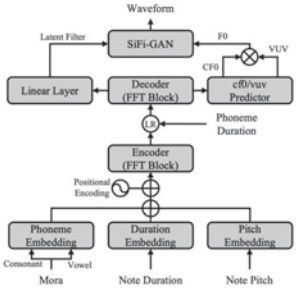


Fig.1：SiFiSinger architecture

3-Q-26

3-Q-26　カナに対する自動韻律付与における音声情報活用の検討

Investigation of Acoustic Information Utilization for Automatic Prosody Annotation in Kana.

○木内貴浩(豊橋技科大)，△鷲崎海(株式会社 Fusic)

背景

近年，音声の韻律情報を考慮した音声合成は自然な音声合成が可能．韻律記号のアノテーションは経験と知識が必要で難しく，コストがかかるため，自動韻律付与モデルが必要である．

提案手法

日本語の音声とテキストからマルチモーダルに韻律記号を予測するモデルを提案する．テキストと音声は事前学習済みモデルによって特徴量に変換され，Attention 機構を持つ Multi-Modal Fusion Decoder 用いてアライメントを取り，最終的に Decoder の出力から各トークンに対する韻律記号を予測する．

結果

テキストから韻律記号を予測するモデルと比較して，提案手法が一番良い精度であった．

更に，複数話者や様々な発話スタイルにおいても，高い精度で韻律記号を予測することができた．

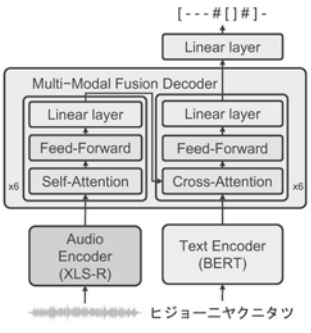


Fig 1 Proposed Method

3-Q-27

3-Q-27 日本語嫌悪感情音声の音響特徴量解析

Acoustic feature analysis of Japanese voices that express disgust

☆俣野文義，森勢将雅(明治大学)

◆現在，著者らは平静音声のみを学習したモデルから嫌悪感情合成音声を生成する手法について研究を行っている．

◆嫌悪感情の発露が確認されている音声の各母音に対して特徴量解析

- 音声：男性話者１名（平静・嫌悪 各1000文）
- 対象：基本周波数，フォルマント周波数，発話長，スペクトル重心
- 結果：基本周波数の下降・スペクトル重心の上昇

	n	Neu [Hz]	Dis [Hz]
/a/	6369	117.46	107.86
/i/	4110	118.59	109.46
/u/	2718	122.59	112.54
/e/	3182	120.33	111.25
/o/	5690	120.42	110.24

Fundamental frequency

	n	Neu [Hz]	Dis [Hz]
/a/	6201	1196.53	1267.94
/i/	3870	1287.22	1391.89
/u/	2551	1010.49	1096.80
/e/	3154	1434.48	1520.55
/o/	5393	816.71	858.59

Spectral centroid

3-Q-28

3-Q-28 音声特徴表現語に基づく音声の主観評価予測

Learning to assess speech based on subjective voice descriptors

◎近藤祐斗，亀岡弘和，田中宏，金子卓弘，原田登

（日本電信電話株式会社 NTT コミュニケーション科学基礎研究所）

◆本研究では音声特徴表現語（subjective voice descriptor: SVD）に基づく音声評価予測を扱い，ACR 形式の主観評価データを学習データとする MOSNet に似たニューラルモデル（pointwise approach）と，CCR 形式の主観評価データを学習データとするRankNetに基づくニューラルモデル（pairwise approach）を比較する．

◆比較実験を行うにあたり大規模な主観評価データセットを作成した．

◆実験を通じて提案法が SVD に基づく音声評価予測に有効であることと，CCR 形式のデータを用いて学習したモデルがACR 形式のデータを用いて学習したモデルより高い性能を達成することを確認した．

Tab.1 : The maximum value over all epochs of averaged pprefs across all trials and the standard deviation of the pprefs at the epoch that maximize, where the ppref (precision of preferences) means the accuracy rate for CCR format questions.

Speech assessment method	Youthful -sounding female voice	Youthful -sounding male voice	Resonant male voice
F_0 comparison	0.595	N/A	N/A
Pointwise approach	0.701 ±0.008	0.596 ±0.017	0.754 ±0.012
Pairwise approach ($w = 0.2$)	**0.703 ±0.010**	**0.641 ±0.014**	**0.759 ±0.009**
Pairwise approach ($w = 0.3$)	0.702 ±0.011	0.635 ±0.013	**0.759 ±0.008**

3-Q-29

3-Q-29 基本周波数およびスペクトルの微細変動を含むビブラート音声におけるスペクトル変動の有無が自然性に与える影響の調査

A study on effect on naturalness of vibrato by existence of spectral fluctuation in singing voice including fine fluctuation of spectrum and fundamental frequency.

◎鈴木千文，坂野秀樹，旭 健作(名城大)

◆本研究ではビブラートの変動を除去したビブラート合成音を作成

- スペクトルの変動を含む場合と含まない場合(スペクトル変動消失音声)でビブラートの変動を除去したビブラート合成音

◆ ビブラート合成音の作成方法

- スペクトルの各ビンの時系列と基本周波数の時系列に含まれる大局的な変動(カットオフ周波数 3Hz のローパスフィルタを使用)と微細変動(カットオフ周波数15Hzのハイパスフィルタを使用)を抽出

◆ スペクトル変動消失音声の作成

- ビブラートが開始するフレーム以降のスペクトルの値をビブラート区間の基本周波数の各フレームのスペクトルの値で置き換える

◆ 作成したビブラート合成音の印象

- 基本周波数のビブラートの変動が除去されている合成音の方が，スペクトルのビブラートの変動が除去されている合成音に比べてより不自然さが増す印象である
- ビブラート音声の知覚において基本周波数のビブラートの変動の影響の方が大きい可能性がある

3-Q-30

3-Q-30 PLS に基づく時変複素 AR 音声分析の検討

On a study of TV-CAR speech analysis based on PLS

○舟木慶一(琉球大学)

解析信号に対する時変複素AR(Time-Varying Complex AR:TV-CAR)音声分析の提案ならびに音声処理への適用の研究を行なっている。TV-CAR音声分析として、MMSE[1]、GLS/ELS[2]のみならず、ℓ_1ノルム正則化であるLASSO方式[3]、ℓ_2ノルム正則化としてRLP基準方式[4]、TRLP基準方式ならびにRLPとTRLPのハイブリッド方式[5]を提案し、音声のF_0推定を用いて評価を行った。さらに、前処理として骨導フィルタ[6]や適応的プリエンファシス[7]、ならびに解析信号の算出にフレーム毎のHilbert変換の導入の検討[8]も行った。今回、多重共線性により頑健とされる部分的最小二乗法(PLS)[9][10]に基づくTV-CAR分析を構築し、評価を行ったので報告する。

3-Q-31

3-Q-31 speech-laugh の発生機序解明に向けた冒頭母音の音響分析

Acoustical analysis on initial vowel of speech-laugh

☆瀬戸口 遼, 有本 泰子(千葉工大)

◆背景：発話と笑いが同時に発生する speech-laugh の発生機序が不明

◆目的：音響分析によって speech-laugh 冒頭と発話の母音生成の違いを検証する

◆手法:speech-laugh 冒頭母音と発話に含まれる母音全体からフォルマント周波数や声門開放率，スペクトル傾斜を抽出し，発声表現(発話または speech-laugh 冒頭)と母音の種類(五母音)を要因とした二元配置分散分析を実施

◆結果：各特徴量では男女ともに発声表現に有意な差がみられ，母音生成に違いがあることが示唆された

◆考察：女性では/a/や/o/で差がみられたことから，speech-laugh は口腔内の空間を狭めた時に発生している可能性がある

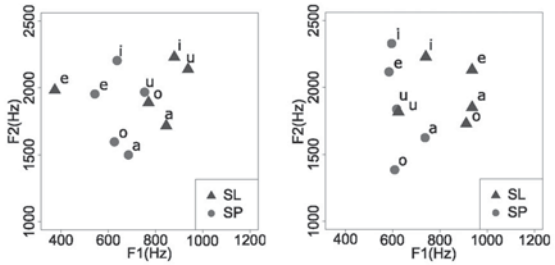


Fig.1:Vowel space for Speech(SP) and Speech-laugh(SL)
(Left:men, Right:women)

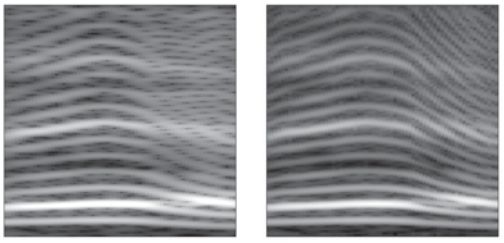
3-Q-32

3-Q-32 チャープレット変換と再割り当て法を用いた基本周波数推定

Pitch estimation using chirplet transform and reassignment method

☆中谷優太，赤石夏輝，山田宏樹，矢田部浩平（農工大）

- **チャープレット変換**
  信号の時間・周波数・チャープレートの３次元表現が得られる
  ※ チャープレート … 周波数の時間変化量
- チャープレートの情報を用いて**調波構造を強調**したスペクトログラムを作る (Fig.1)
- 短時間フーリエ変換によるスペクトログラムを用いたときより**基本周波数推定の精度が向上した**



(a)　(b)

Fig.1: Spectrograms obtained by (a) the short time Fourier transform and (b) the proposed method

3-Q-33

3-Q-33 声帯膜振動から膜-声帯振動への遷移

Transition from vocal membrane oscillation to vocal membrane – vocal fold co-oscillation

☆小畑大樹，吉谷友紀,△廣野凌(立命館大),
△西村剛(京都大),徳田功(立命館大)

◆声帯膜物理モデルに内転を加えることで声帯膜振動を再現した。

◆物理モデルと摘出喉頭を用いて吹鳴実験を行い、声門下圧、音声、高速度動画を計測した。

◆流量を増加させることで、声帯膜のみが振動する状態から、膜と声帯が同時に振動する状態に遷移する様相が確認できた。

◆物理モデルと摘出喉頭の両方で、共に声帯膜振動より声帯振動の方が、発声効率が高い傾向が確認された。

◆物理モデルと摘出喉頭の両方で、共に遷移直前の声帯膜振動において、発声効率が低くなる区間が確認された。

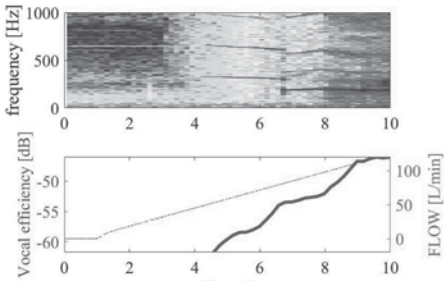


Fig. 1 Spectrogram and vocal efficiency of physical model.

3-Q-34

3-Q-34 音声のスペクトル包絡からの基本周波数予測

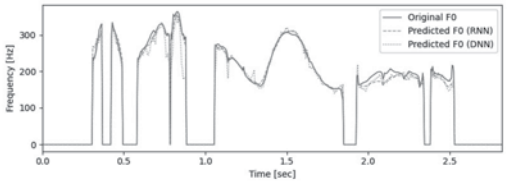
Prediction of Fundamental Frequency Using Spectral Envelope of Speech.

〇今野 英明, △寺崎 巧真(北海道教育大)

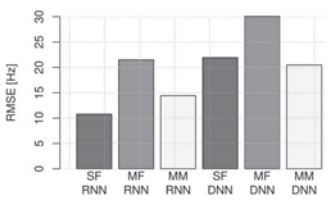
◆音声の基本周波数 (F0) は声道特性（スペクトル包絡）と相関がある

◆音声のスペクトル包絡から基本周波数を予測して精度を調査
- LSTM を用いたリカレントニューラルネットワーク (RNN)
- 順伝播型の深層ニューラルネットワーク (DNN)

◆予測結果例：女声「あと 30 分の猶予が与えられた」



◆予測誤差 (RMSE)



SF: 単一話者女声 (JSUT コーパス)，MF : 51 名女声 (JVS コーパス)，MM : 49 名男声 (JVS コーパス)

3-Q-35

3-Q-35　声門幅を狭めた声帯物理モデルにおける声門内圧測定:正弦波非定常流の流量振幅が与える影響

Measurement of intra-glottal pressure with narrowed glottal width: effect of flow amplitude on unsteadiness of the glottal pressure

☆中村和史，△井上晴貴，徳田功(立命館大)

◆アクリル製 M5 型声帯物理モデル（Fig.1）を作成した。
◆声門幅を 0.20 mm に固定した。
◆定常流を流す実験と流量が正弦波状に変動する非定常流実験を行い，声門内圧，声門下圧を測定した。
◆非定常流の声門下圧振幅を 300 Pa と 900 Pa と設定した。
◆定常流実験と非定常流実験における声門内圧力分布の比較を行った。
◆声門下圧振幅 900 Pa の場合に定常流と非定常流で声門内圧に有意差が見られたが，300 Pa の場合には有意差が見られなかった（Fig.2）。

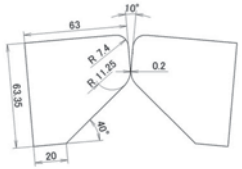


Fig.1 : Configuration of M5 Model.

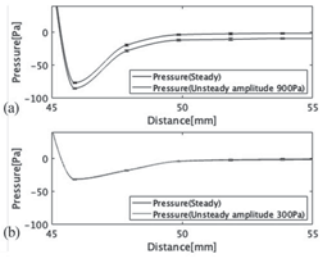


Fig.2 : Comparison of the intraglottal pressure distributions between steady and unsteady flow experiments. Pressure amplitude was 900 Pa in (a), 300 Pa in (b).

3-Q-36

3-Q-36　超音波画像診断法による口蓋形状の造影について

Visualization of the palate by ultrasound imaging

○北村達也(甲南大)，孫静，林良子(神戸大)，
能田由紀子，前川喜久雄(国語研)

◆超音波画像診断法では超音波ビームが舌表面で反射するため，口蓋が造影されない．**口蓋の形状を知りたい！**
◆水やゼリーを口腔内に保持すると口蓋が造影されることが知られているが，より適した媒質はないか？
◆水，炭酸水，ゼリー，プリン，こんにゃく，チーズなどの飲料や食品をテスト．
◆クラッシュタイプこんにゃくゼリー（マンナンライフ ララクラッシュ）を用いた超音波画像を Fig. 1 に示す．
◆媒質による口蓋造影効果には個人差があるらしい．
◆ポスター発表では，MR 画像との比較結果も示す．

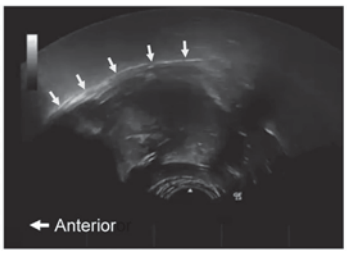


Fig.1 : Ultrasound image acquired while a participant held crush-type jelly made from *Konnyaku* in the mouth. Allows indicate the palate.

3-Q-37

3-Q-37　声帯結節が発声に及ぼす影響を調べるための物理モデル実験

A physical model experiment to investigate the effect of vocal nodules on vocalization

☆上田可奈子，徳田功(立命館大)

◆幾何学的数値を参考に声帯結節を模擬した物理モデル（EPI 結節モデル）と，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。
◆結節は，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。
◆結節が大きくなるにつれて，オンセット圧は徐々に上昇し，病変の位置で振動が弱くなった。

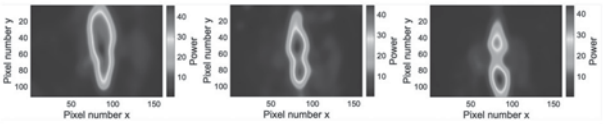


Fig. 1 Laryngo-topographies of vocal fold models with a nodule size of (a) 1.5, (b) 2.0, (c) 3.0 mm.

3-Q-38

3-Q-38　１次元畳み込みニューラルネットワークを用いた IPA からの調音運動データ生成

Generation of Articulatory Movement Data from IPA using 1DCNN

☆丸山虹樹，澤田隼，大村英史，桂田浩一（東京理科大）

本研究では，1 次元畳み込みニューラルネットワーク(1DCNN)を用いて，国際音声記号(IPA)から調音運動データの一種である EMA データを生成する手法を提案する(Fig. 1)．従来手法では，ルールベースで生成された言語特徴量から BLSTM-LSTM を用いて，EMA データを生成している．この手法では概ね良好な EMA データが生成されているが，言語特徴量は汎用性が低く，調音情報を十分に含んでいるとはいえない．また，BLSTM-LSTM は時系列上で長期的な特徴を捉えることを得意とするが，調音運動データの生成には短期的な文脈の方が重要であると考えられる．そこで本研究では，汎用性が高く，調音情報を多く含んでいる IPA を入力として，1DCNN を用いて EMA データを生成する．また，更なる高精度化のために，IPA 間もしくは EMA データの測定点間の相互関係をモデル化することができる SEBlock を用いる方法を検討する．

実験では，IPA と EMA データを含んだ日本語のデータセットを用いた．従来手法と比較して高精度かつ高速に EMA データを生成することができた．

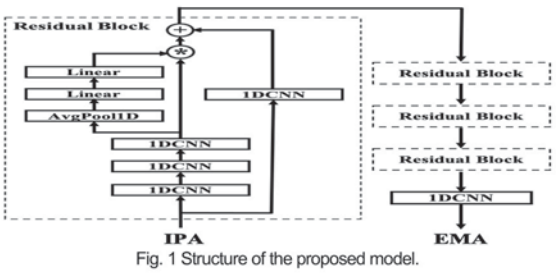


Fig. 1 Structure of the proposed model.

3-Q-39

3-Q-39　ｅスポーツに適したマウスガードの検討 ～話せるマウスガードの作成～

Examining of Mouthguards Suitable for eSports
- Creation of a speakable mouthguard -

○葭田敏之（葭田歯科医院）

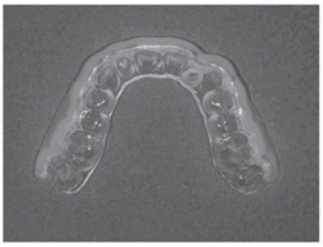
◆ ｅスポーツをプレー中に食いしばり、歯や顎の痛みを感じる人がいる。

◆ マウスガードはｅスポーツにも必要である。

◆ プレーだけでなく配信を行うため発話の必要性は従来のスポーツよりも高い。

◆ 従来のスポーツ用のマウスガードは話しづらい。

◆ プロを含む 3 名の上位プレーヤーに色々なマウスガードを作り、使いやすいマウスガードを検討しました。



3-Q-40

3-Q-40　Perceptual Assimilation of English High and Mid Vowels by Japanese Listeners: The Role of L1 Lexical Knowledge

日本語話者による英語の高舌および中舌母音の知覚同化について：L1 語彙知識の役割

○Kanako Tomaru（Tokyo University of Science）

◆The current research investigates the effects of the lexical knowledge of the native language (L1) on the perception of non-native sounds using the Perceptual Assimilation Model (PAM) framework.

◆The current research focuses on Japanese speakers' perception of the English high and mid vowels, i.e., /ɪ/ and /ɛ/.

◆The comparison between the baseline (*NoAct* conditions) and conditions that activate lexical knowledge (*Act e* and *Act a* conditions) shows no significant effect. The summary is shown in Fig. 1.

◆The results of the perceptual experiment (1) suggest no L1 lexical effects on perceptual assimilation of the target English vowels, i.e., /ɪ/ and /ɛ/, and (2) imply that the process of perceptual assimilation may be independent of the process of lexical decision.

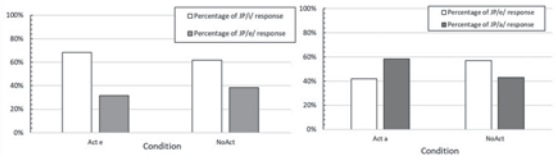


Fig.1 : The results of the English /ɪ/ vowel (light) the English /ɛ/ vowel (right)

3-Q-41

3-Q-41　Pause position analysis in spontaneous speech for L2 English fluency assessment

Sylvain COULANGE[1], Tsuneo KATO[2]
[1]LIDILEM/LIG, Université Grenoble Alpes; [1,2]SLPL, Doshisha University

When learning a foreign language, fluency is often considered a key characteristic for being easily understood by listeners. To assess fluency in a spontaneous speech situation, automatic tools typically deal with the frequency and mean duration of silent and filled pauses, but pauses are is not necessarily a problem, rather, they should make speech more comprehensible. It can, however, be detrimental when they occur in unexpected positions. We present a fully automated pipeline for identifying pause positions at both the phrase (syntactic structure) and lexical levels (part of speech of surrounding words). Analysis of 177 L2 English learners has shown that pause position varies greatly from one student to another, and should be better taken into account when assessing speaking.

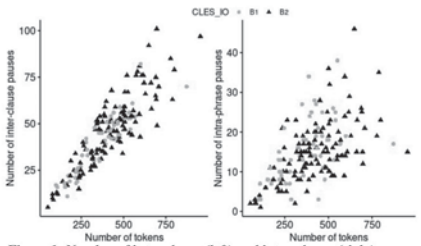


Figure 1: Number of inter-clause (left) and intra-phrase (right) pauses per speaker.

3-Q-42

3-Q-42　授業内音声イベント検知モデルと授業活性度推定モデルの統合

Integration of a machine learning model for detecting acoustic events in class with a model for estimating class activity

☆高瀬壮彦（滋賀大・院），市川治（滋賀大）

◆本研究では「授業活性度」指標を導入し、主観的な授業評価を定量的に行うことを目標としている．ここで「授業活性度が高い」とは，教師が生徒とインタラクティブな授業を行うスタイルのことである．

◆データには教員に装着したピンマイクを用いて中学校で収録した 21 コマの授業の音声を使用した．

◆前報では，授業中の各時点において何が起きていたかを示した音イベントについて正確な情報を手作業で作成し，正解ラベルとして授業活性度の評価に用いて実験を行った．

◆本報告では「授業内音声イベント検知モデル」の出力である予測ラベルを授業活性度推定の入力特徴量とすることで，授業の音声から機械学習による授業活性度の推定がどの程度可能であるか検証を行う．

◆授業活性度推定モデルの学習にはLightGBMの回帰モデルを使用しており，21fold のクロスバリデーションを行った．正解ラベル使用時についても同じ条件で実験を行った．

◆表１は授業活性度を分類問題とした場合の混同行列である．精度は等しく，QWK では予測ラベル使用時が高い結果を示した．しかし，MSE では予測ラベル使用時が大きい値であったため，推定ラベルを使用すると正解との誤差は大きくなる結果となった．

Table 1 Experimental results

予測ラベルを入力とした場合

Actual \ Predicted	0	1	2
0	2	2	0
1	2	7	2
2	0	2	4

Accuracy : 0.619　QWK: 0.592

正解ラベルを入力とした場合

Actual \ Predicted	0	1	2
0	2	2	0
1	3	7	1
2	1	1	4

Accuracy : 0.619　QWK: 0.481

3-Q-43

3-Q-43　ドイツ語弱音節ミニマル・ペアの知覚が習熟度に与える影響

Impacts of German language proficiency on the perception of unstressed vowels

◎粕谷麻里乃（東邦音大・音楽），荒井隆行（上智大・理工）

◆**背景** ドイツ語には音韻表記上類似する音素がある（例 Feta [feːˈtaː] / Väter [feːˈtɐ]）．類似した音素は，通常発話においてはTargetlessにより知覚・生成上の曖昧さも指摘されるが，[e] はr音化することがあり [a] との弁別を助けることがある．

◆**目的** 母語の音素数も言語リズムも異なる日本語母語話者を対象に，語末が弱音節開母音 [a] / [e] となるペアの知覚実験を行う．統制群の母語話者も含め，知覚の実態を明らかにする．習熟度別被験者を集め，習得過程上の特徴や生成との関わりを示す．

◆**結果** 発話者により [a] / [e] の知覚の正答率に差を確認した．

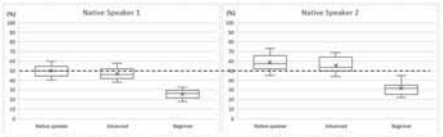


Fig.1：習熟度別知覚の正答率 (左：母語話者，中央：上級者，右：初級者)

◆**考察** 正答率の差は，[a] と [e] のフォルマント周波数の違い，フォルマント周波数の軌跡に観察された F2，F3 の動きである．上級者，母語話者では，この差を知覚，生成できた話者が一致していた．初級者は，生成ではフォルマント周波数上に違いはほぼなかったが，知覚においては一部被験者が両者の違いに感度を示した．このことは，知覚の早期習得の可能性も考えられ，被験者数を増やし継続的に調査を進める．

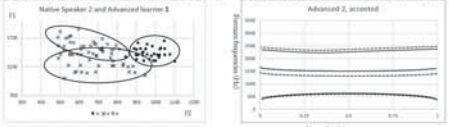


Fig.2：(左) フォルマント周波数 (F1, F2). (一部母語話者と上級者)
Fig.3：(右) フォルマント周波数の軌跡 (F1, F2, F3). (一部上級者)

3-Q-44

3-Q-44　家庭教師による１対１オンライン授業の授業品質評価法の提案

Estimation of online lecture quality taught by private tutors

☆高田朋幸（大阪工大院），脇田由実（大阪工大院）

我々は，オンライン塾の授業評価を自動化するため，受講生の応答発声における F0 の時間変化に注目し，発声毎の F0 の標準偏差を分析した．授業終盤の 20 %の時間帯の応答発声の F0 標準偏差は，授業の良し悪しで差があり，良い授業の方が大きくなる傾向があった．また，上記時間帯のF0標準偏差を用いた授業良し悪しの識別率は正解率は74 %であった．23%の誤りの主な要因は F0 抽出誤りである．学生応答発声の F0 標準偏差にて授業良し悪しの判断の可能性を示唆したが，課題は F0 抽出誤りである．下図は，良い授業と悪い授業における上記 F0 標準偏差の比較図である．（横軸は正規化した授業の経過時間を示す）

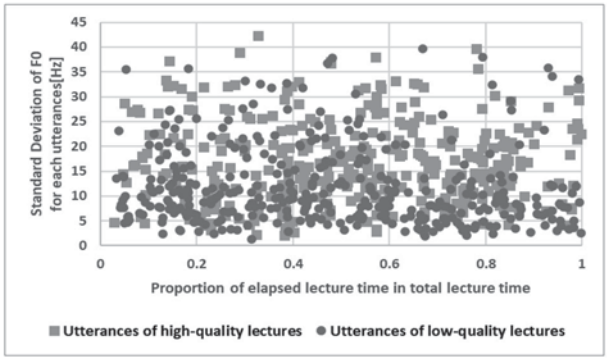


Fig.1：Relationship between proportion of elapsed lecture time and SD-F0 values

3-Q-45

3-Q-45

外国語コミュニケーションにおける聴取力と発声力をシームレスに強化するインタラクティブ性及びゲーム性の高い音声教材の開発

Development of English training drills to improve speech perception and production skills seamlessly with interactive gamification

峯松信明*，中西のりこ**，高英翔*，崔ジェヒョン*，孫海童*
*東京大学，**神戸学院大学

音声分析 ＋ 音声認識 ＋ 音声合成 ＋ 大規模言語モデル

を統合的に活用して，外国語音声コミュニケーションにおける

聴取力 ＋ 発声力 ＋ 会話力

を効果的に強化する

インタラクティブ性 ＋ ゲーム性

の高い音声教材を開発し，東大工学部にて

体育会系特訓型授業

をこの夏休みに実施しています。

3-Q-46

3-Q-46　小学校での音読課題の実施形態調査に基づく支援アプリの検討

Preliminary examination of an interface for the development of oral reading support application

○白勢彩子（東京学芸大），宮澤幸希，△佐藤可直（フェアリーデバイセズ）

◆小学校において，その教育効果への期待から音読が宿題として課されることが多い。しかしながら，教員にとっては自宅での実施内容が不明瞭，家庭では児童の動機づけなどの課題がある。

◆我々は，小学生の音読の宿題に特化したアプリによる課題解決を考え，児童向けの音読アプリの開発を目指して研究を進めており，本稿では，前回の調査を踏まえ，教育現場でのアプリの使用方法を複数提案し，それに応じたアプリケーションのインターフェイスをワイヤフレームにより実現し，小学校教員を対象にインタビューを行った。

◆音読の宿題について，場面１：教室で確認，場面２：端末経由の提出，場面３：オンライン提出の３つの場面での活用と手続を呈示したところ，場面３が最適との意見であった。しかしながら，場面３は児童が録音音声をオンラインストレージ等へアップロードし，教員がアップロードされた音声にアクセスする必要があり，アプリで処理を完結できないという問題がある。場面２についてはセキュリティの問題から不適切との指摘があり，これも併せ，近接でのファイル送信機能を用いて運用することを，今後，検討したい。

Fig.1：Three types of assignment submission methods and evaluation procedures.

3-Q-47

3-Q-47　読み上げ音声との比較に基づく強調の定義と抽出

Definition and extraction of emphasis based on comparison with read speech

○中嶋秀治（日本電信電話（株），NTT コミュニケーション科学基礎研究所）

◆背　景：　分かりやすく伝えるためには「強調」が有用だが，どのことばを強調するのか（強調位置）に関する教示がほぼない

◆課　題：　①強調位置の聞き取りはコスト高，②強調位置演出とそれに合わせた演技音声収録で，演出位置と音声との位置ズレが生じてないか？　合うまでの収録は自然か低コストか？

⇒　音声からの抽出が望ましくないか？

◆本研究：　音声からの強調位置抽出実験を通して，音声中のどこのなにを聞き取って「強調」と判定しているか？　を検討

◆本　稿：　韻律特徴の変動方向と比率が，強調抽出対象音声と基準音声との間で大きく異なる箇所を強調位置として抽出．これを通して強調の定義と抽出法を考察

対象音声：高齢者にとって聞いて分かりやすいと考えられる音声

基準音声：ニュース読み上げスタイルの音声

話者：　現役アナウンサー

◆これができたら：

a) 自然な状況で収録された模範的な音声から強調位置を収集し，話し方の教示に

b) 音声合成用学習データ内や合成音声内での強調表出の適切性評価に

c) 語彙情報との関連を分析する際の適切な対象に

3-Q-48

3-Q-48　幼児の言語能力と社会認知能力の発達に運動が与える効果

A study on the effects of sports coordination exercises on the development of language abilities and social abilities in preschoolers

○須藤路子（順天堂大），△東根明人（コーチングバリュー協），伊藤憲治（国立感覚器セ），△桐野衛二（順天堂大）

本研究の目的は、コーディネーション運動と体育指導による運動が、幼児の言語能力と社会認知能力の発達に与える効果を測定することであった。

本研究の実験被験者は、保育園 2 ヵ所の年中クラス 4 歳～5 歳児の 32 名 (男児: 18 名、女児: 14 名)であった。

言語能力は、語彙流暢性テストにより、言語生成能力を測定した。言語理解能力の測定には、遠城寺式乳幼児分析的発達検査法を使用した。社会認知能力は、「心の理論」の共感性を測定した。運動能力は、立ち幅跳び、ボール投げ、および捕球回数を測定した。実験手順は、1) 保育所において、コーディネーション運動を実施する運動グループと体育指導による運動の比較グループを用意した。2) 実験期間は週 1 回 30 分、3 か月とした。3) 実験期間の前後に、上記の３能力の測定を実施した。

コーディネーション運動と体育指導による運動の効果に関し、言語生成能力の発達と、運動能力の中で、立ち幅跳びにおいて、コーディネーション運動が有意に効果的であることが示された。これは、幼児と指導者の双方向のコミュニケーションを重視するコーディネーション運動と指導者主導型の体育指導との重要な差違によるものである。コーディネーション運動は、微細運動に先立って発達する全身運動である立ち幅跳びの上達を促進したと考えられる。

3-Q-49

3-Q-49　雑音環境下におけるヒトの話者識別の誤り傾向

Error tendency of human speaker identification in noisy environments

☆鈴木良平(上智大)，網野加苗(科警研)，荒井隆行(上智大)

◆法科学で扱う音声資料には，様々な雑音が混入する。

◆聴取検査による音声鑑定など，ヒトの話者認識の捜査応用を考えるにあたっては，犯人の取り逃がしや冤罪を防ぐため，ヒトの話者認識にはどのような誤り傾向があるのか調べる必要がある。

◆犯罪捜査においては，雑音を含む音声も証拠となり得るが，雑音環境下の音声の個人性やヒトの話者認識について調べた研究は少ない。

◆本研究では聴取実験を行い，雑音の程度や種類，用いる単語を変化させることで識別正答率，他者受容率(FAR)，本人拒否率(FRR)がどう変化するか統計的に分析を行った。

◆この結果，SN 比の低下と識別正答率低下の関連やヒトの話者識別について FAR > FRR となる可能性を確かめられた．加えて聴取による話者識別の FAR と FRR には複雑な傾向がある可能性が判明した。

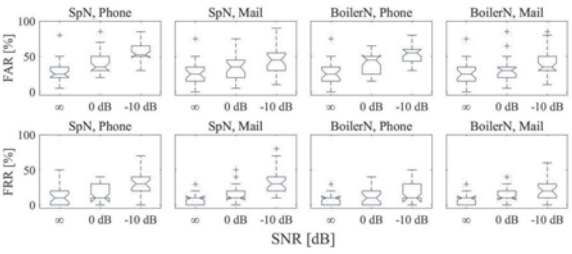


Fig.1 : SNR effect on FAR & FRR

3-Q-50

ネコの声の言語表現について －日本語話者へのアンケート調査から－

Onomatopoetic representations for cat sounds: A questionnaire survey of Japanese speakers

○横山安紀子（日本大学）

ネコの声については様々な種類があると言われるが，ノドを鳴らす声について見ると，実際の動物の声と人間が表現するオノマトペとは音響音声学的に見てもかなり乖離している。本研究では実際の鳴き声から人間の言語表現に至る過程を解明するための1つの手がかりとして，日本語話者を対象としたアンケート調査を行った。

調査参加者：「ネコを飼ったことのある人」および「ネコを飼ったことのない人」（各群 20 代～60 代の5つの年代別男女 1000 人）

質問1：ネコがうれしいときや満足しているときにノドを鳴らす声はどんな声に聞こえますか？　質問2：なぜそのように思いましたか？

結果は Fig. 1～Fig. 4 に示した。質問1および質問2については x^2 検定により有意差が認められた。年代別，男女別での有意差は認められなかった。

全体のアンケート調査から，動物に接する経験の有無にかかわらず，日本語話者は辞書項目にある日本語オノマトペ表現を使用する傾向が高いことがわかった。

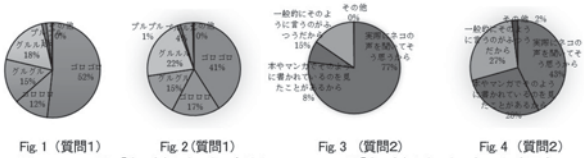


Fig. 1（質問1）　Fig. 2（質問1）　Fig. 3（質問2）　Fig. 4（質問2）

※Fig. 1, Fig. 3 は「ネコを飼ったことのある人」，Fig. 2, Fig.4 は「ネコを飼ったことのない人」を示す